

In The Name Of God

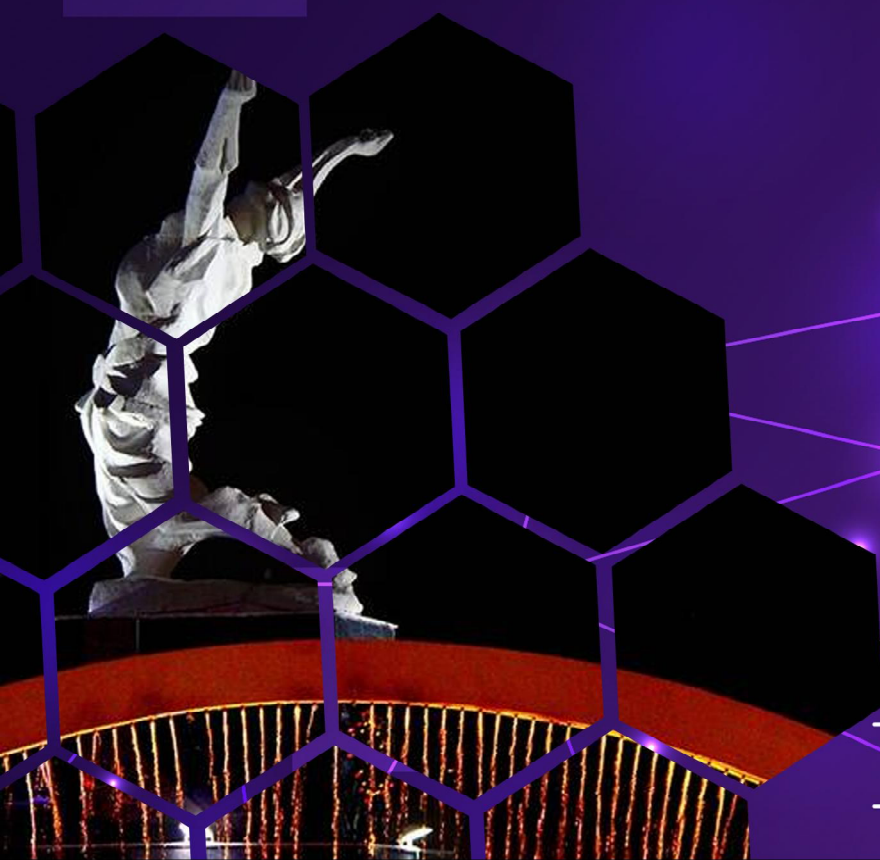


PROCEEDINGS

OF THE 15 SEMINAR

ON PROBABILITY & STOCHASTIC PROCESSES

- August 30-31, 2025
- University Of Kurdistan



**Proceedings of the 15th Seminar on Probability and Stochastic
Processes**

30-31 August 2025

University of Kurdistan, Sanandaj, Iran

Identification

Title : Proceedings of the 15th Seminar on Probability and Stochastic Processes

Subjects: Statistics, Probability, Stochastic Processes

Editors: Dr. Sirus Fathi Manesh, Dr. Hadi Emami, Dr. Hadi Ahmadi

Date: 30-31 August 2025

Location: University of Kurdistan, Sanandaj

Preface

It is with great pleasure that we present the proceedings of the Fifteenth Seminar on Probability and Stochastic Processes, held at the University of Kurdistan in collaboration with the Iranian Statistical Society. This seminar served as a vital platform for the national scientific community, bringing together leading experts, researchers, and students to foster dialogue and disseminate cutting-edge research.

This volume contains a collection of the accepted English articles from the seminar. We are confident that the diverse and innovative work presented here will provide valuable insights and inspire further research, particularly for graduate students embarking on their academic careers.

The seminar received an enthusiastic response, with 98 submissions in both Persian and English. Following a rigorous peer-review process, 75 of these articles were accepted for presentation and publication, either in full or as extended abstracts. The review process was greatly supported by esteemed colleagues from universities across the country, to whom we are deeply indebted. The most prominent theme among the accepted papers was "Time Series and Applications," with 15 contributions.

The scientific program was enriched by seven invited keynote lectures and six specialized sessions covering a wide spectrum of topics, including probability theory, stochastic processes, time series, distribution theory, stochastic analysis, financial mathematics, statistical inference for stochastic processes, and the growing applications of stochastic methods in artificial intelligence and data science. A dedicated workshop on financial mathematics further enhanced the event. The articles in this volume are organized according to these central themes and are indexed in the Islamic World Science Citation Center (ISC) and the Civilica database. We hope this collection serves as a meaningful contribution to the advancement of knowledge in its field.

Organizing an event of this scale would not have been possible without the collective effort of many individuals and institutions. We extend our sincere gratitude to the numerous reviewers from universities nationwide for their diligent work in evaluating submissions. We are also profoundly thankful for the unwavering support of the President of the University of Kurdistan, the Dean of the Faculty of Basic Sciences, and the President of the Iranian Statistical Society. The dedicated efforts of the University of Kurdistan's IT unit, administrative staff, and student volunteers were instrumental in the seamless execution of the seminar.

Finally, we gratefully acknowledge the financial support provided by the Planning and Budget Organization of Kurdistan Province.

The Scientific and Executive Committees

Contents

Organizers and Sponsors	1
Executive Committee	2
Scientific Committee	3
Articles	4

Organizers and Sponsors:



Executive Committee:

Hadi Ahmadi

Hossein Bevrani

Kourosh Dadkhah

Hadi Emami (Secretary of the Executive Committee)

Eghbal Ghaderi

Sommayeh Jaffar-Ramashti

Sirous Fathi Manesh

Sommayeh Zand

Shaho Zarei

Scientific Committee:

Mohammad Amini

Ferdowsi University of Mashhad

Hossein Bevrani

University of Kurdistan

Sirous Fathi Manesh
(Secretary of the Scientific committee)

University of Kurdistan

Omid Kharazmi

Vali-e-Asr University of Rafsanjan

Ali Reza Nematollahi

Shiraz University

Zahra Rezaei Ghahroodi

University of Tehran

Ahmad Reza Soltani

Shiraz University

Shaho Zarei

University of Kurdistan

Articles

Rethinking Uncertainty Measures: Replacing Entropy with Extropy in Learning Algorithms, Abolhosseini, S. and Khorashadizadeh, M. and Chahkandi, M	11
Asymptotic Analysis of Numerically Estimated Likelihood Estimators within the Context of Error-Prone Regression Models, Ahangari, M., Golalizadeh, M., Rezaei Ghahroodi, Z.	18
The Law of Iterated Logarithm for the Smooth Estimator of Distribution Function in Length-Biased Data, Ajami, M. and Akbari, M.	26
Parameters Estimation for Two Lomax Exponential Distributions Under Joint Type-II Censored Data, Akbari, S. and Bevrani, H. and Maroufi, M.	31
From Probability Uncertainty to Nonlinear PDEs, Akhtari, B.	39
Review on Complete Convergence for m-NOD Random Sequences, Amini, M. and Naderi, H.	45

A Review of Quantum-Based Distance Metrics and Their Applications in Modern Finance, Arian, H.	46
A Novel Bayesian Two-Stage Approach for Dynamic Prediction Using Joint Modeling of Multiple Longitudinal Markers and Time-to-Event Data, Baghfalaki, T.	47
Allocated Optimal Strategies to the Wealth Process of a Diffusion Risk Model, Baziyari, A.	48
Case deletion diagnostic and Generalized Cook's distance Using Liu Estimator in Linear Mixed Measurement Error Models, Borhani, Sh. and Ghapani, F.	57
Some Inequalities for Absolute Moments of Feasible Acceptable Random Variables, Eghbal , N.	62
An interactive R Shiny Application for Learning Times Series Analysis, Frances, M. C. and Salehi, M. and Bekker, A. and Arashi, M.	69
Milstein scheme for Numerical solution of first-order uncertain stochastic differential equations, Ghasemifard, A.	77

Estimation of the Intensity Function of Point Processes Using Soap Film Bases, Ghorbani, M.	78
Kolmogorov-Arnold Networks (KAN): Revolutionizing Prediction in Machine Learning, Hagbin, H.	79
Goodness of Fit Test Based on Entropy and Dirichlet Process, Haghsheenas, S. and Saberi, Z.	84
The PRGM Prediction for Finite Population Parameters Using a Parametric Model Under SE and RN Loss Functions, Jafaraghaie, R.	92
Statistical Inference Based on Composite Likelihoods for Spatio-temporal Point Processes, Jalilian, A.	99
Complete convergence for randomly weighted sums of WOD sequences, Javanmardi, R. and Nezakati, A.	100
Reliability Measures in Semi-Markov Repairable Systems, Khorshidian, K. and Fathizadeh, M.	109

On the Double-Log Skew-Normal Distribution and Its Application, Mahdavi, A. and Negarestani, H.	117
Modeling Electricity Demand: A Comparative Study of Univariate, Multivariate, and Functional S.S.A., Mahmoudvand, R.	124
Improved Generalized Additive Models in Time Series, Malekzadeh, A. and Rajabi Naraki, F.	130
Stochastic Modeling with Normal Mean–Variance Mixtures: A Robust Linear Experts Framework for Censored Data in Machine Learning, Manouchehri, T. and Nematollahi, A. R.	136
A Comparison of Maximum Likelihood Estimation and Particle Swarm Optimization for Parameter, Estimation in Logistic and Probit Regression Models, Mohammadi, S. M. and Etesami, R. and Madadi, M.	144
Smooth Transition Markov Switching R-vine copulas, 26 Mohammadi Bastini, F. and Jafaraghaie, R.	153
Bayesian Neural Networks using Variational Bayes, Monemi, M. and Amini, M. and Taheri, S.M. and Fahim, A.	161

Inhomogeneous Mark Correlation Functions for Point Processes on General State Spaces, Moradi, M.	167
Optimizing Generative Adversarial Networks for Financial Time Series: A Stochastic Approach with Novel, Loss Metrics, Moradi, P. and Kazemi, S. M. M.	168
Determination of Minimal Weighted Risk Sample Sizes in Sampling Inspections with Stochastic defect, rate, Naghizadeh Qomi. M.	176
On the Stability and Convergence of a Finite Difference Scheme for Stochastic Partial Differential Quations, Namjoo, M. and Aminian, M. and Negarestani, H. and Salemi, H.	184
Unsupervised and Supervised Time Series Grouping in Data Science, Nematollahi, A. R.	191
Statistical Inference for Spatio-Temporal Gibbs Point Processes, Raeisi, M.	205

Estimating the Effective Reproduction Number of Tuberculosis in a Branching Process, Rastegar, M. and Fakoor, V. and Taghi Shakeri, M.	206
European Option Pricing under Market Uncertainty and Liquidity Risk: A Heston-Based Stochastic Volatility Model, Rezaei Bahrmand, M.	217
Markovian Queueing Systems and Application in Real Data, Shams, M. and Mirzaie, A.	225
SDE-Based Models for Trajectories on Linear Networks, Sharifi, A.	233
Application of Functional Linear Regression with Impact Points to Real Data: A Case Study, Shirvani, A.	234
Application of Statistical Models in Designing a System to Save AnimalLives, Taheri Saif Abad, M.	238
Lookback options in stochastic intensity models, Tahmasebi. M.	244

Bayesian Estimation for the Parameters of the Lindley-Burr XII Distribution, Tarvirdizade, B. and Babayi, S.	245
Some Estimators for the Residual Varentropy Function of Length-biased Data, Zamani, R.	253
Pathwise Grid Valuation of Fixed-Income Portfolios with Applications to Risk Management, Zamani, Sh. and Chaghazardi, A. and Arian, H.	261
Modified Maximum Likelihood-Based Test for Nonlinear Time Series Model Selection, Zamani Mehryan, S.	262
Estimation of count parameters in zero-inflated Poisson regression model based on Stein-Liu estimators, Zandi, Z. and Bevrani, H.	270
Value at Risk Estimation Using the ARMA–GARCH Framework, Mohebbi, S. and Mohammadian Mosammam, A.	277



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Rethinking Uncertainty Measures: Replacing Entropy with Extropy in Learning Algorithms

S. Abolhosseini^{*1}, M. Khorashadizadeh², M. Chahkandi³

^{1, 2, 3} Dept. of Statistics, University of Birjand

Abstract

This paper investigates replacing entropy with extropy as the impurity measure in decision tree algorithms. Decision trees are widely used and interpretable machine learning models that recursively partition data based on features to create hierarchical decision rules. Impurity measures such as entropy and the Gini index play a key role in selecting the best feature and split point at each node. Extropy, a newer measure focusing on the probability of error rather than uncertainty, has been proposed as an alternative to entropy. Using the Iris dataset, the performance of decision trees constructed with entropy, extropy, and the Gini index was compared. Results showed that the entropy-based tree achieved the highest accuracy of approximately 97.78%, while extropy- and Gini-based trees both attained similar accuracies around 93.33%. The performance difference between entropy and Gini is generally small, and due to its computational efficiency, the Gini index is often preferred in practical applications. Extropy, with its focus on error probability, demonstrated competitive performance close to Gini and can serve as a suitable alternative in scenarios where error-focused impurity assessment is desirable.

Keywords: Decision tree, Extropy, Entropy, Gini Index.

Mathematics Subject Classification (2020): xxAxx, xxBxx, xxCxx.

1 Introduction

In recent years, machine learning algorithms have attracted significant attention due to their ability to analyze complex data and extract meaningful patterns. Among these algorithms, decision trees stand out as one of the most widely used and interpretable models. A decision tree is a hierarchical structure that recursively partitions data into subsets based on feature values, resulting in simple and understandable decision rules. Due to their interpretability, computational efficiency, and broad applicability in classification and regression tasks, decision trees have been extensively employed in various fields such as business analytics, medicine, and social sciences. Impurity measures play a crucial role in the construction of decision trees, as they guide the selection of the best feature and split point at each node. The most commonly used impurity measures include **Entropy** and the **Gini Index**, both of which quantify the

^{*}Corresponding author, abolhosseinis@birjand.ac.ir

degree of disorder or uncertainty in the data. Information Gain, based on entropy, has been the primary splitting criterion in classical algorithms like ID3 [Quinlan \(1986\)](#). However, a relatively less explored impurity measure called **Extropy** [Lad and *et al.* \(2015\)](#) has recently been proposed. Extropy focuses on the probability of error and can serve as an alternative splitting criterion. Given the structural similarities between extropy and the Gini index, as well as the differences with entropy, a thorough comparison of these three measures can provide deeper insights and potentially improve decision tree performance. The objective of this paper is to replace the traditional information gain with extropy gain in decision tree algorithms and to conduct a comprehensive comparison among extropy, entropy, and the Gini index. Using the Iris dataset, we evaluate the impact of these impurity measures on tree construction, classification accuracy, and other performance metrics.

2 Literature Review and Theoretical Background

Decision trees are one of the most popular supervised machine learning algorithms used for classification and regression tasks. The model is structured as a hierarchical tree consisting of a root node, internal nodes (decision nodes), branches, and leaf nodes (terminal nodes). Each internal node represents a test on a feature, branches correspond to the outcome of the test, and leaf nodes represent class labels or predicted values. A key objective in decision tree construction is to create subsets of data that are as homogeneous as possible with respect to the target variable. This homogeneity ensures clear decision boundaries and improves classification accuracy. To achieve this, impurity measures are used to evaluate how well a feature splits the data at each node.

2.1 Impurity Measures

The most commonly used impurity measures include:

- **Entropy:** Measures the level of uncertainty or disorder in the data. It is defined as

$$\text{Entropy} = - \sum_{i=1}^n P_i \log P_i,$$

where P_i is the proportion of samples belonging to class i . Lower entropy indicates more homogeneous subsets.

- **Gini Index:** Measures the probability of incorrectly classifying a randomly chosen element if it was labeled according to the class distribution in the subset. It is calculated as

$$\text{Gini} = 1 - \sum_j P_j^2.$$

Values closer to 0 indicate purer nodes [Quinlan \(1996\)](#).

- **Extropy:** A less explored impurity measure defined as

$$\text{Extropy} = - \sum_{i=1}^n (1 - P_i) \log(1 - P_i),$$

which focuses on the probability of error and offers an alternative perspective on impurity.

2.2 Feature Selection and Splitting Criteria

At each node, the decision tree algorithm selects the feature and split point that maximize the reduction in impurity. For entropy, this reduction is quantified by Information Gain, which measures the decrease in entropy after the split. Similarly, for Gini and Extropy, analogous gain measures can be defined to guide the splitting process.

2.3 Related Work

Classical algorithms such as ID3 and C4.5 [Lad and *et al.* \(2015\)](#) primarily use entropy-based information gain for splitting, while CART uses the Gini index.

3 Methodology

In this study, we propose a novel approach to decision tree construction by replacing the traditional information gain based on entropy with a new impurity measure called extropy. The goal is to evaluate whether extropy can serve as a more effective splitting criterion compared to entropy and the widely-used Gini index.

3.1 Extropy as an Impurity Measure

Extropy is defined as follows for a node containing samples from C classes, where P_i denotes the proportion of samples belonging to class i :

$$\text{Extropy} = - \sum_{i=1}^C (1 - P_i) \log(1 - P_i).$$

Unlike entropy, which measures uncertainty, extropy focuses on the probability of error, providing an alternative perspective on impurity.

3.2 Extropy Gain

Similar to information gain in entropy-based trees, we define extropy gain as the reduction in extropy after splitting a node based on a feature and a threshold. Given a dataset D , splitting it into subsets D_1 and D_2 , the extropy gain is computed as:

$$\text{Gain}_{\text{extropy}} = \text{Extropy}(D) - \frac{|D_1|}{|D|} \text{Extropy}(D_1) - \frac{|D_2|}{|D|} \text{Extropy}(D_2).$$

The feature and split point that maximize $\text{Gain}_{\text{extropy}}$ are selected at each node.

3.3 Decision Tree Construction

The decision tree is constructed recursively as follows:

- 1 At each node, calculate the extropy gain for all candidate features and their possible split points.
- 2 Select the feature and split point with the highest extropy gain.
- 3 Partition the data into left and right subsets based on the selected split.
- 4 Repeat the process recursively for each child node until a stopping criterion is met, such as:
 - Maximum tree depth
 - Minimum number of samples in a node
 - No further gain in extropy
- 5 Assign the majority class label to leaf nodes.

3.4 Dataset and Experimental Setup

We use the well-known Iris dataset [Unwin and *et al.* \(2021\)](#) for evaluation. The dataset contains 150 samples with 4 features each, classified into 3 classes. The dataset is randomly split into training (70%) and testing (30%) subsets. The decision tree models are trained separately using extropy, entropy (information gain), and Gini index as splitting criteria.

3.5 Performance Metrics

To evaluate and compare the models, the following metrics are used:

- Accuracy: The proportion of correctly classified samples.
- Precision: The ratio of true positives to the sum of true and false positives.
- Recall: The ratio of true positives to the sum of true positives and false negatives.
- F1-score: The harmonic mean of precision and recall.

Additionally, confusion matrices and graphical representations of the decision trees are generated for qualitative analysis [Han and *et al.* \(2012\)](#).

4 Experimental Results

In this section, we present the experimental evaluation and performance comparison of decision tree classifiers constructed using three different impurity measures: Extropy, Entropy, and the Gini Index. The experiments were conducted on the Iris dataset, a widely used benchmark in machine learning, and the models were evaluated using four key metrics: Accuracy, Precision, Recall, and F1-score.

4.1 Performance Metrics Comparison

Metric	Extropy-Based Tree	Entropy-Based Tree	Gini-Based Tree
Accuracy	0.9333	0.9778	0.9333
Precision	0.9184	0.9794	0.9444
Recall	0.9167	0.9778	0.9333
F1-score	0.9165	0.9777	0.9327

Table 1: Comparison of performance metrics in three types of decision trees

The entropy-based decision tree achieved the highest performance across all metrics, with an accuracy of approximately 97.78%. Both extropy- and gini-based trees showed similar results, with accuracy around 93.33%. Precision and recall values followed the same trend, indicating that entropy slightly outperforms the other two criteria in this particular dataset.

4.2 Discussion

These results align with findings in the literature where entropy often provides marginally better classification performance due to its comprehensive measure of uncertainty and information gain. However, the difference between entropy and gini is generally small, and gini is preferred in many practical applications due to its computational efficiency. Extropy, as a newer impurity measure focusing on the probability of error rather than uncertainty, demonstrated competitive performance close to gini. This suggests that extropy can be a viable alternative impurity measure in decision tree learning, especially in scenarios where error-focused impurity assessment is desirable.

4.3 Confusion Matrices

To further analyze the classification performance of the decision tree models based on Extropy, Entropy, and Gini Index, confusion matrices were generated for each model. The confusion matrix is a comprehensive tool that summarizes the number of correct and incorrect predictions broken down by each class. It provides insights into which classes are being confused by the classifier.

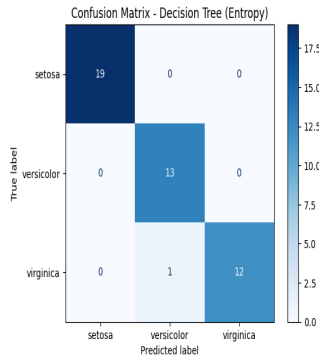


Figure 1: *
Entropy

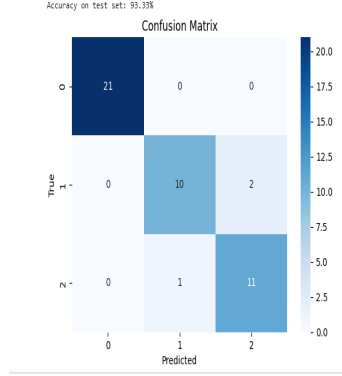


Figure 2: *
Extropy

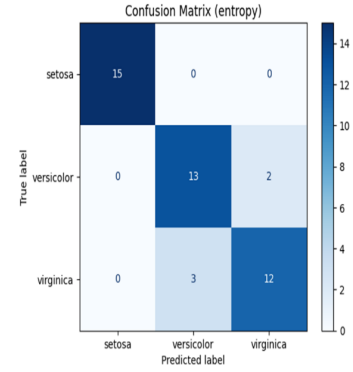


Figure 3: *
Gini Index

4.4 Decision Tree Visualizations

Graphical representations of the decision trees constructed using Extropy, Entropy, and Gini impurity measures are presented in Figures A, B, and C respectively. These visualizations illustrate the hierarchical decision rules and feature splits used by each model. The trees predominantly split on petal length and petal width features, consistent with domain knowledge about the Iris dataset. The structures of the trees are similar, reflecting the comparable performance of the impurity measures, but subtle differences in split points and tree depth can be observed.

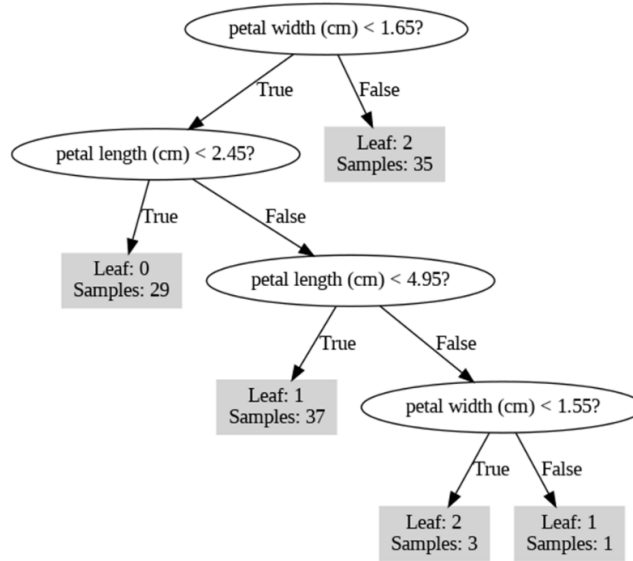


Figure 4: Extropy.

5 Conclusion

This study demonstrated that extropy, as a new impurity measure in decision tree algorithms, offers competitive performance close to the Gini index and can serve as a suitable alternative to entropy in scenarios where focusing on the probability of error is important. Although the entropy-based tree achieved higher accuracy in this research, the performance difference

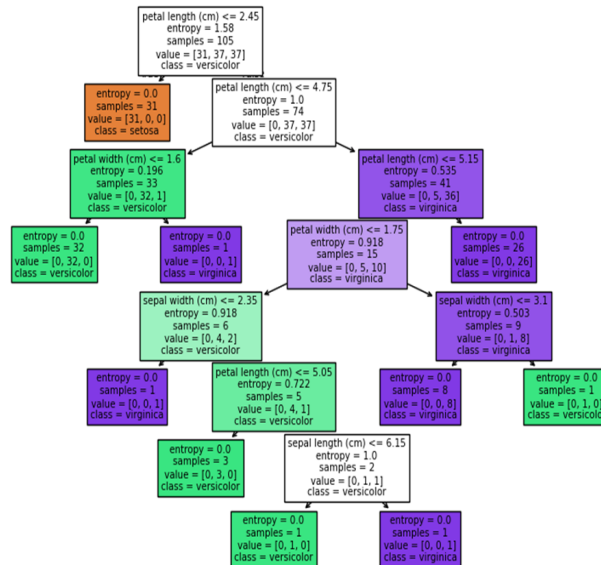


Figure 5: Entropy.

with extropy and Gini was negligible, and extropy may provide computational advantages. Therefore, the choice of impurity measure should be based on the specific needs of the problem and the trade-off between accuracy and computational efficiency. It is also recommended that further extensive experiments be conducted on more diverse datasets to better evaluate the generalizability and benefits of extropy.

References

- Han, J., Kamber, M., and Pei, J. (2012), *Data mining: Concepts and. Techniques*, Waltham: Morgan Kaufmann Publishers.
- Lad, F., Sanfilippo, G., and Agro, G. (2015), *Extropy: Complementary dual of entropy.*.
- Quinlan, J. R. (1986), *Induction of decision trees*, Machine Learning, **1**, 81-106. [4] Quinlan, J. R. (1996). Improved use of continuous attributes in C4.5. Journal of Artificial Intelligence Research, **4**, 77-90.
- Quinlan, J. R. (1996), *Improved use of continuous attributes in C4.5*, Journal of Artificial Intelligence Research, **4**, 77-90.
- Unwin, A., and Kleinman, K. (2021)., *The iris data set: In search of the source of virginica.*, Significance, **18(6)**, 26-29



Figure 6: Gini Index.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Asymptotic Analysis of Numerically Estimated Likelihood Estimators within the Context of Error-Prone Regression Models

Maryam Ahangari^{*1}, Mousa Golalizadeh¹, Zahra Rezaei Ghahroodi²

¹Department of Statistics, Tarbiat Modares University

²School of Mathematics, Statistics and Computer Science, University of Tehran

Abstract

Generalized linear mixed models (GLMMs) represent prevalent methodologies for the examination of multilevel datasets, employing the maximum likelihood framework for the estimation of model parameters. In the context of nonlinear regression models pertaining to the outcome variable, it is imperative to address classification errors associated with responses as well as measurement errors present in covariates, in order to ensure the precision of inferences, which consequently renders the likelihood function analytically intractable. In such scenarios, a traditional yet adaptable methodology to tackle numerical integration challenges, is the Monte Carlo simulation; nonetheless, simulation analyses indicate that the utilization of alternative approximation techniques, such as the multivariate Gauss-Hermite quadrature method, demonstrates significantly superior performance in practical applications. This manuscript delineates the general conditions requisite for the Gauss-Hermite approximation technique, aimed at guaranteeing both consistency and asymptotic normality. Furthermore, a simulation study is included to exemplify this point.

Keywords: Multivariate Gauss-Hermite quadrature, Measurement error, Misclassification, Consistency, Asymptotic normality.

Mathematics Subject Classification (2020): 60F99.

1 Introduction

The application of multilevel models is to form a pattern for the relation between the response variable and a set of covariates by incorporating the correlation between repeated measurements. When the response variable is non-normally distributed, generalized linear mixed models (GLMMs) are extensively applied, considering a nonlinear link between the mean of the response variable with a set of predictors. In many situations, some of the predictors of GLMMs may not be measured precisely, so measurement error in continuous covariates is

^{*}Corresponding author, maryam.ahangari@gmail.com

a challenging problem in the analysis of multilevel-structured data with discrete responses. General strategies of handling covariate mismeasurement effects in estimating the parameters corresponding to correlated data is the Monte Carlo scheme. Recently, [Xie and *et al.* \(2017\)](#) have shown that the Monte Carlo Newton-Raphson (MCNR) method gives accurate estimates in the logistic regression model for analyzing longitudinal biomarker data, accounting for left-censoring and covariate measurement error.

As a distinct source of error, the misclassification of categorical response variables within the context of regression analysis represents a significant consideration in evaluating the validity of parameter estimation as well as in facilitating subsequent statistical inferences. Consequently, the rectification of misclassification has emerged as a focal point for researchers aiming to achieve precise estimates alongside valid confidence intervals. For further clarification, the research conducted by [Buonaccorsi \(2010\)](#) is recommended. In the context of response misclassification, [Neuhaus \(2002\)](#) assessed the extent of asymptotic bias inherent in the estimation of logistic regression parameters when the response is subject to classification error within a longitudinal binary data framework. [Paulino and *et al.* \(2003\)](#) have additionally introduced a Bayesian paradigm for analyzing binary responses that are affected by classification error, employing a random effects logistic regression model. Furthermore, [Tang and *et al.* \(2015\)](#) concentrated on longitudinal validation data-driven studies that involve repeatedly measured responses characterized by measurement error. They assumed a differential misclassification mechanism and employed the maximum likelihood approach to correct for misclassification in the binary response.

Measurement error in covariates together with misclassification in the outcomes occur frequently in epidemiological studies, and failure to account for these two major sources of error, results in potentially biased and inaccurate estimates. [Roy \(2012\)](#) has conducted a comprehensive examination of these dual sources of errors within the framework of longitudinal data, aiming to rectify both response classification inaccuracies and covariate measurement discrepancies, through the application of the Monte Carlo Markov Chain (MCMC) computational methodology. [Yi \(2017\)](#) has investigated methodologies for concurrently addressing errors in both response variables and covariates, pertaining to univariate as well as correlated datasets. The structure of this article is as follows: In Section 2, we describe generalized linear mixed models involving binary correlated response components in conjunction with non-differential misclassified responses and covariate measurement error when covariates are subject to measurement error. We also discuss the likelihood function and the application of the MGHQ method in scenarios where precise measurements of certain covariates are unavailable. In Section 3, we focus on the conditions required to ensure the consistency and asymptotic normality of estimators based on such approximations. Finally, in Section 4 we apply the theory to a simulation study based on a random effects logistic regression model subject to response misclassification and covariate measurement error.

2 The Generalized Linear Mixed Misclassification and Measurement error Model, Basics and Notation

Suppose $Y_i = (Y_{i1}, \dots, Y_{in_i})^T$ denotes the observed outcomes for the i th subject, $1 \leq i \leq m$, where m is the total number of independent individuals, and n_i is the number of observations for individual i . Let Y_i be a misclassification-prone response variable that follows a generalized linear mixed model with random intercept for each individual. The vector of model covariates $X_i = (X_{i1}^T, \dots, X_{in_i}^T)^T$ and $Z_i = (Z_{i1}^T, \dots, Z_{in_i}^T)^T$ are assumed to be time-varying. Let Y_{ij} be the response variable for subject i at the j th occasion measurement ($j = 1, \dots, n_i$), X_{ij} as the vector of true covariates which are error-prone and unobserved and hence latent, and Z_{ij} be the vector of error-free covariate. Furthermore, let X_{ij}^* and Y_{ij}^* be surrogate measurements of X_{ij} and Y_{ij} , respectively, that are observed in the absence of X_{ij} and Y_{ij} . Note that the

notations used here are similar with those utilized in the literature. See, e.g., [Ahangari and *et al.* \(2019\)](#)

2.1 The Outcome Process, Measurement Error Models and Classification Error Probabilities

Let Y_{ij} be a binary response variable for subject i at occasion j . Assume that the binary outcome is associated with the collection of covariates via the random effects logistic regression framework as

$$\text{logit}[P(Y_{ij} = 1|x_{ij}, z_{ij}, \tau_i)] = \beta_0 + \beta_x x_{ij} + \beta_z z_{ij} + \tau_i, \quad i = 1, \dots, m, \quad j = 1, \dots, n_i. \quad (1)$$

In (1), $\beta = (\beta_0, \beta_x, \beta_z)^T$ is the vector of fixed parameters. We assume that the random effects τ_i , ($i = 1, \dots, m$) are independent from each other, and also independent from the error-prone covariates X_{ij} . In this article, it is assumed that $\tau_i \sim N(0, \sigma_\tau^2)$, where σ_τ^2 is unknown and assumed to be constant for all subjects.

In (1), the covariate X_{ij} is not directly observable, and X_{ij}^* is the only available measurement in the absence of X_{ij} , so X_{ij}^* can be regarded as the proxy variable for X_{ij} . In the following, it is assumed that X_{ij}^* is a combination of X_{ij} plus some error, specifically classical measurement error, in which the true but latent variable is measured with additive error, usually assumed to have constant variance, i.e.

$$X_{ij}^*|x_{ij} = x_{ij} + e_{ij}. \quad (2)$$

where e_{ij} is the error term such that $E(e_{ij}|x_{ij}) = 0$ and $\text{Var}(e_{ij}|x_{ij}) = \sigma_e^2$, when X_{ij}^* and X_{ij} are scalar. It is concluded that $E(X_{ij}^*|x_{ij}) = x_{ij}$, so X_{ij}^* is unbiased for the unobserved x_{ij} . We also consider that the measurement error in X_{ij}^* is non-differential. This means that Y_{ij} is conditionally independent of X_{ij}^* , given x_{ij} and if X_{ij} is observed, the proxy variable X_{ij}^* adds nothing to the predictor of Y_{ij} , i.e.;

$$P(Y_{ij} = 1|x_{ij}, x_{ij}^*, z_{ij}, \tau_i) = P(Y_{ij} = 1|x_{ij}, z_{ij}, \tau_i). \quad (3)$$

In the following, it is assumed that we have homoscedastic measurement error, i.e. the variance of X_{ij} given x_{ij} is constant.

With random covariate X_i subject to error, we assume distributional assumption on error-prone covariate X_i , i.e. structural approach in which covariate X_i has a normal distribution with mean μ_x and variance Σ_x , and is independent of τ_i . We choose a fully structural case for the mismeasured covariate with X_i assumed i.i.d with

$$X_i = (X_{i1}^T, \dots, X_{in_i}^T)^T \sim N(\mu_x, \Sigma_x) \quad i = 1, \dots, m,$$

where

$$\Sigma_x = \begin{pmatrix} \sigma_{x1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{x2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{xn_i}^2 \end{pmatrix}.$$

The measurement error in the multivariate case, also has a normal distribution with the following structure: $e_i \sim N(\mathbf{0}_{n_i}, \sigma_e^2 \mathbf{I}_{n_i})$. Suppose that Y_{ij} is also subject to misclassification, and Y_{ij}^* is an observed value of Y_{ij} . Particularly, let

$$P(Y_{ij}^* = 1|Y_{ij} = 0, x_{ij}, z_{ij}, \tau_i) = P(Y_{ij}^* = 1|Y_{ij} = 0) = \gamma_{01} \quad (4)$$

and

$$P(Y_{ij}^* = 0|Y_{ij} = 1, x_{ij}, z_{ij}, \tau_i) = P(Y_{ij}^* = 0|Y_{ij} = 1) = \gamma_{10}, \quad (5)$$

denote the simple case of misclassification probabilities. The probability $1 - \gamma_{10}$ is often called the sensitivity of the measurement Y_{ij}^* , and $1 - \gamma_{01}$ is called the specificity. In (4) and (5), γ_{01} and γ_{10} are treated as unrecognized, nonnegative classification error probabilities that are constant and independent of other covariates within the model. This declaration is generally known as ‘non-differential misclassification’. According to the notations given so far, the conditional probability for the observed response given the true values of covariates and the random component is provided as follows:

$$\begin{aligned} P_{ij}^* &= P(Y_{ij}^* = 1 | x_{ij}, z_{ij}, \tau_i) \\ &= P(Y_{ij}^* = 1 | Y_{ij} = 1, x_{ij}, z_{ij}, \tau_i) P(Y_{ij} = 1 | x_{ij}, z_{ij}, \tau_i) \\ &\quad + P(Y_{ij}^* = 1 | Y_{ij} = 0, x_{ij}, z_{ij}, \tau_i) P(Y_{ij} = 0 | x_{ij}, z_{ij}, \tau_i) \\ &= (1 - \gamma_{10}) P_{ij} + \gamma_{01} (1 - P_{ij}) = \gamma_{01} + (1 - \gamma_{10} - \gamma_{01}) P_{ij}, \end{aligned} \quad (6)$$

where the quantity P_{ij} is described as

$$P_{ij} = P(Y_{ij} = 1 | x_{ij}, z_{ij}, \tau_i) = \frac{\exp(\beta_0 + \beta_x x_{ij} + \beta_z z_{ij} + \tau_i)}{1 + \exp(\beta_0 + \beta_x x_{ij} + \beta_z z_{ij} + \tau_i)}. \quad (7)$$

Besides, let us consider X_{ij}^* as a surrogate for X_{ij} with classical additive measurement error as (2), where $e_{ij} \sim N(0, \sigma_e^2)$. Then, the conditional distribution of $X_{ij}^* | x_{ij}$ can be written as $X_{ij}^* | x_{ij} \sim N(x_{ij}, \sigma_e^2)$.

2.2 Likelihood Function Arrangements

In this Section, we provide a parametric likelihood procedure that allows for classification error in the binary responses and measurement error in the covariates in a random effects logistic regression model. According to (1) and (2), if $\theta = (\beta, \sigma_\tau^2, \sigma_e^2)^T$ is a vector of associated parameters along with the vector of misclassification rates $\gamma = (\gamma_{01}, \gamma_{10})^T$, for the i th subject at occasion j , given the observed data $(y_{ij}^*, x_{ij}^*, z_{ij})$ for $i = 1, \dots, m$ and $j = 1, \dots, n_i$, the log-likelihood function inspired by the observed data encompassing the measurement error distribution and the classification errors can be written as follows:

$$\ell(\theta; y^*, x^*) = \sum_{i=1}^m \log \int_R \int_{R^{n_i}} \prod_{j=1}^{n_i} f(y_{ij}^* | x_{ij}, z_{ij}, \tau_i; \beta) f(x_{ij}^* | x_{ij}; \sigma_e^2) f(x_i; \mu_x, \sigma_{x_j}^2) f(\tau_i; \sigma_\tau^2) dx_{ij} d\tau_i. \quad (8)$$

For the approximation of the integral (8), multivariate forms of the Gauss-Hermite quadrature approximation method is needed to handle both unobserved covariate x_{ij} at n_i occasions and the random effects τ_i . Due to the fact that both x_{ij} and τ_i are independent from each other, the product of their density functions can be defined as a multivariate normal distribution.

The multivariate Gauss-Hermite quadrature (MGHQ) method extends the univariate Gauss-Hermite quadrature rule by applying it to each coordinate of the integrated variables. To approximate the integrals in (8) which are $(n_i + 1)$ dimensional, a matrix of GHQ nodes are generated for unobserved variables X_{ij} at occasion j and also for the random effects τ_i . This is done based on the mean vector and variance-covariance matrix of the multivariate normal distribution $f(x_i, \tau_i; \sigma_{x_j}^2, \sigma_\tau^2)$, with respect to which the integral is calculated. Additionally, a vector of weights corresponding to this multivariate normal distribution is computed to complete the approximation. The problem is that the integrals in (8) are multidimensional, with their values depending upon both the random effects as well as the underlying latent covariate X_{ij} across n_i occasions. The estimated parameter vector $\hat{\theta}_{ML} \in \Theta \subset R^p$ is acquired by maximizing $L(\theta; y^*, x^*)$, with p as the number of parameters in the regression model. As previously noted, the integrals in (8) cannot be evaluated in a closed form. To manage unobserved covariates and random effects, multivariate Gauss-Hermite quadrature (MGHQ) approximation approach is used, by choosing sets of GHQ nodes $\hat{x}_{kj} = (\hat{x}_{k1j}^{(1)}, \hat{x}_{k2j}^{(2)}, \dots, \hat{x}_{k(n_i+1)j}^{(n_i+1)})'$, $\hat{\tau}_k = (\hat{\tau}_{k1}^{(1)}, \hat{\tau}_{k2}^{(2)}, \dots, \hat{\tau}_{k(n_i+1)}^{(n_i+1)})'$ in addition with a vector of weights $w_k = (w_{k1}^{(1)}, w_{k2}^{(2)}, \dots, w_{k(n_i+1)}^{(n_i+1)})'$. In this setting, the parameter k refers to the indices $k_1, k_2, \dots, k_{(n_i+1)}$, each index shows different sets of quadrature points assigned, in order

to approximate each of $(n_i + 1)$ integrals. Consequently, according to [Heiss and Winschel \(2008\)](#), the MGHQ approximation to the log-likelihood function of (8) will result the following structure:

$$\begin{aligned}\tilde{\ell}(\boldsymbol{\theta}; y^*, x^*, z) &= \sum_{i=1}^m \log \left[\sum_{k_1=1}^{q_1} \sum_{k_2=1}^{q_2} \cdots \sum_{k_{(n_i+1)}=1}^{q_{(n_i+1)}} w_{k_1}^{(1)} w_{k_2}^{(2)} \cdots w_{k_{(n_i+1)}}^{(n_i+1)} \right. \\ &\quad \left. \times \prod_{j=1}^{n_i} f(y_{ij}^* | \hat{x}_{kj}, z_{ij}, \tau_k; \beta) f(x_{ij}^* | \hat{x}_{kj}; \sigma_e^2) \right] \\ &= \sum_{i=1}^m \log \left[\sum_{k_1=1}^{q_1} \sum_{k_2=1}^{q_2} \cdots \sum_{k_{(n_i+1)}=1}^{q_{(n_i+1)}} w_{k_1}^{(1)} w_{k_2}^{(2)} \cdots w_{k_{(n_i+1)}}^{(n_i+1)} \prod_{j=1}^{n_i} \phi(y_{ij}^*, \hat{x}_{kj}, z_{ij}, \tau_k) \right].\end{aligned}\quad (9)$$

Note that

$$\{\hat{x}_{kj} = (x_{k_1j}^{(1)}, x_{k_2j}^{(2)}, \dots, x_{k_{(n_i+1)}j}^{(n_i+1)})' : 1 \leq k_1 \leq q_1; 1 \leq k_2 \leq q_2; \dots; 1 \leq k_{(n_i+1)} \leq q_{(n_i+1)}\}.$$

On the whole, this technique results the maximum approximated likelihood estimator (MALE), i.e.

$$\hat{\theta}_{MALE} = \arg \max_{\theta \in \Theta} \tilde{\ell}(\boldsymbol{\theta}; y^*, x^*, z).$$

The selection of appropriate quadrature points and weights is crucial to achieving the desired level of accuracy in numerical integration. While increasing the parameter k enhances the approximation accuracy, it also incurs a corresponding increase in computational cost. Consequently, it is essential to determine the necessary level of accuracy that guarantees the resulting estimator maintains both consistency and asymptotic normality. To address this, we introduce a link function $K : \mathbb{N} \rightarrow \mathbb{N}$ that relates the number of quadrature points k to the sample size n , in other words $k = K(n)$. In the context of Monte Carlo techniques, the likelihood function is approximated through simulation, and consistency is ensured under standard regularity conditions provided that the number of simulation draws k increases with the sample size, i.e. $k = K(n) \geq n$.

3 Asymptotic Theory Corresponding to Approximated Likelihood Function

As previously noted, when the likelihood contribution $\ell(\boldsymbol{\theta}; y^*, x^*, z)$ cannot be derived in closed form due to its analytical intractability, it becomes necessary to employ an approximation method. The approximated likelihood $\tilde{\ell}(\boldsymbol{\theta}; y^*, x^*, z)$ depends on an accuracy parameter k . This parameter shows the degree of approximation error as well as the interrelated computational costs. Accordingly, we define the quantity

$$\epsilon(k) = \sup_{\theta \in \Theta} |\tilde{\ell}(\boldsymbol{\theta}; y^*, x^*, z) - \ell(\boldsymbol{\theta}; y^*, x^*, z)| + \sup_{\theta \in \Theta} \|\nabla_{\theta} \tilde{\ell}(\boldsymbol{\theta}; y^*, x^*, z) - \nabla_{\theta} \ell(\boldsymbol{\theta}; y^*, x^*, z)\|, \quad (10)$$

which measures the worst-case approximation error of the function ℓ and its gradient $\tilde{\ell}$.

Beyond establishing the consistency of the maximum likelihood estimator, the only additional requisite for the consistency of $\hat{\theta}_{MALE}$ is the uniform convergence in probability of the surrogate log-likelihood $\tilde{\ell}(\boldsymbol{\theta})$ to the true, but analytically intractable $\ell(\boldsymbol{\theta})$. To rigorously pursue this aim, it is expedient to adopt the following notation for the partial derivative of a sufficiently smooth function g :

$$D_{\alpha}^{(\theta)} g(\theta) = \frac{\partial^{|\alpha|}}{\prod_{j=1}^p \theta_j^{\alpha_j}} g(\theta),$$

where $\alpha = (\alpha_1, \dots, \alpha_p)$ is a multi-index indicating the order of partial derivatives for different parameters $\theta_1, \dots, \theta_p$. In the subsequent analysis, for the purpose of establishing an upper bound on the magnitude of $\sup_{\theta \in \Theta} |\tilde{\ell}(\boldsymbol{\theta}; y^*, x^*, z) - \ell(\boldsymbol{\theta}; y^*, x^*, z)|$ together with $\sup_{\theta \in \Theta} \|\nabla_{\theta} \tilde{\ell}(\boldsymbol{\theta}; y^*, x^*, z) - \nabla_{\theta} \ell(\boldsymbol{\theta}; y^*, x^*, z)\|$, we bring into action

$$\bar{\epsilon}_r(k) = \max_{|\alpha|_1 \leq r, \theta \in \Theta} \sup |D_{\alpha}^{(\theta)} \ell(\boldsymbol{\theta}; y^*, x^*, z) - D_{\alpha}^{(\theta)} \tilde{\ell}_k(\boldsymbol{\theta}; y^*, x^*, z)| \quad (11)$$

for $r = 0, 1, 2$; indicating all partial derivatives with respect to θ up to order r , ([Griebel and et al., 2019](#)).

3.1 Consistency and Asymptotic Normality

In the ensuing Section, we shall formulate comprehensive conditions on the log-likelihood contribution (8) and the approximated function (9) to guarantee the desirable asymptotic properties of the MAL estimator. We begin by revisiting the assumptions underpinning the consistency of the maximum likelihood estimator as established by Newey and McFadden (1994).

Lemma 3.1. *Assume that (i) For all $\theta \neq \theta_0$, we have $f(y; \theta) \neq f(y; \theta_0)$; (ii) $\theta_0 \in \Theta$ and Θ is compact; (iii) $\log f(y; \theta)$ is continuous at each $\theta \in \Theta$ with probability one; (iv) $\mathbb{E}_y[\sup_{\theta \in \Theta} |\log f(y; \theta)|] < \infty$. Then, $\text{plim}_{n \rightarrow \infty} \hat{\theta}_{ML} = \theta_0$ (consistency).*

We next invoke the asymptotic normality of the ML estimator.

Lemma 3.2. *Assume that (i) the assumptions of Lemma 3.1 hold and that (ii) $\theta_0 \in \Theta$; (iii) $f(y; \theta)$ is twice continuously differentiable and $f(y; \theta) > 0$ in a neighborhood \mathcal{N} of θ_0 ; (iv) $\int \sup_{\theta \in \mathcal{N}} \|\nabla_{\theta} f(y; \theta)\| dy < \infty$ and $\int \sup_{\theta \in \mathcal{N}} \|\nabla_{\theta\theta} f(y; \theta)\| dy < \infty$; (v) The function defined by $\mathcal{L} = \mathbb{E}_y[\nabla_{\theta} f(y; \theta_0)(\nabla_{\theta} f(y; \theta_0))^T]$ exists and is nonsingular; (vi) The last assumption is $\mathbb{E}_y[\sup_{\theta \in \mathcal{N}} \|\nabla_{\theta} \log f(y; \theta_0)\|] < \infty$. Then $\sqrt{n}(\hat{\theta}_{ML} - \theta_0) \xrightarrow{d} N(0, \mathcal{L}^{-1})$ (asymptotic normality).*

According to Lemma 3.1 for an assumed function $f(y; \theta)$, in order to handle with consistency of MALE, we use the specific setting where the log-likelihood has the form (8) and the MGHQ-approximated function follows the (9) contribution. For more details, see Griebel and et al. (2019).

Corollary 3.3. *Assume that (i) there exists $\bar{\delta} > 0$ such that for all $\theta \in \Theta$ it holds that $l(\theta; y^*, x^*, z) > \bar{\delta} > 0$; (ii) the function $\phi(y_{ij}^*, x_{kj}, z_{ij}, \tau_k)$ is continuous in $\theta \in \Theta$; (iii) With accuracy parameter k , $\text{plim}_{k \rightarrow \infty} \bar{\epsilon}_0(k) = 0$; (iv) $K(n)$ is monotonically increasing in n . Then, $\hat{\theta}_{MAL}$ is a consistent estimator of θ_0 , i.e.*

$$\hat{\theta}_{MAL} \xrightarrow{P} \theta_0. \quad (12)$$

To characterize the asymptotic distribution of the MALE, it is essential that the worst-case approximation error (11) not only converges to zero but does so at a sufficiently fast rate.

Corollary 3.4. *Assume that the assumptions of Lemma 3.2 and Corollary 3.3 hold. Also assume that (i) for all (y^*, x^*, z) it holds that $\phi(y_{ij}^*, x_{kj}, z_{ij}, \tau_k)$ is twice continuously differentiable; (ii) for all (y^*, x^*, z) and $\theta \in \Theta$ all terms in (8) remain bounded. (iv) $\text{plim}_{k \rightarrow \infty} \bar{\epsilon}_2(k) = 0$; (v) $R(n)$ rises fast enough with n to ensure $\sqrt{n}\bar{\epsilon}_1(K(n)) \xrightarrow{P} 0$. Then*

$$\sqrt{n}(\hat{\theta}_{MAL} - \theta_0) \xrightarrow{d} N(0, \mathcal{L}^{-1})$$

Remark 3.5. *It is important to note that the positivity of the integration weights is not assumed at any point. Additionally, it should be observed that in practical applications, both n and $k = K(n)$ must be finite quantities, with $K(n)$ defined such that $K(n) \geq k_0(\delta)$, for the reason that otherwise $\bar{\ell} > \delta$ might not hold true.*

4 Simulations and Inferences

In this Section, we undertake simulation studies to evaluate and compare the efficiency of the Multivariate (Multidimensional) Gauss-Hermite Quadrature (MGHQ) approximation method in analyzing multilevel data subject to measurement error in the covariates and misclassification in the binary response, i.e. the hybrid method with Monte Carlo Expectation Maximization (MCEM) algorithm, as a maximum simulated likelihood (MSL) method. We use (1), i.e. logistic random effects model to simulate data. We assume $(\beta_0, \beta_x, \beta_z)$ are the intercept and covariates's fixed effects coefficients, for $i = 1, \dots, m$ individuals and $j = 1, \dots, n_i$ occasions. We assume that z_{ij} is an error-free or exactly measured covariate generated as a normally distributed variable following $N(0, 4^2)$ distribution, which is treated fixed during

the simulation study. The measurement error model that we have considered is the classical additive model. In this case, we generate a surrogate variable W_{ij} for the error-prone covariate X_{ij} , as in (2), where e_{ij} is the measurement error variable independently and identically distributed following $N(0, \sigma_e^2)$, but we will change the magnitude of σ_e^2 (measurement error in X_{ij}) along with different scopes of misclassification rates during the simulation study. The response variables are generated from Bernoulli distribution. We generate 2000 different data sets, each simulated data contains $m = 500$ individuals with $n_i = 3$ replicate measurements at each occasion.

4.1 Simulation Results

To assess the performance of the MGHQ and MCEM approaches, we computed simulated results in terms of the mean absolute bias, root of mean squared error (RMSE) across simulation runs, and coverage probability. The notation MGHQk is used to denote the MGHQ scenario with k quadrature points. To mitigate approximation error, the number of quadrature points was progressively increased from 2 to 20. Table 1 presents the estimated parameter values alongside the RMSEs, which serve as a standard metric for illustrating the convergence behavior of the two methods under consideration. The results in Table 1 indicate that the MGHQ approach yields accurate parameter estimates even with a small number of quadrature points, with additional points producing only marginal changes. The MCEM algorithm achieved convergence after 500 iterations. Based on the simulation outcomes, the MGHQ method exhibits lower bias in parameter estimation, reduced RMSEs, and superior coverage probabilities for nominal confidence intervals compared to the MCEM approach.

It is important to recognize that the convergence rate of an approximation method reflects the necessary growth rate of the link function $K(n)$ to ensure asymptotic normality. Therefore, we have conducted an analysis on the convergence of $\sqrt{n}\epsilon(K(n))$ for various link functions. Based on our simulation results (not presented herein), it was observed that the MGHQ method exhibits superior performance relative to the MCEM approach when utilizing a logarithmic link function.

Furthermore, the magnitude of the approximation error is intrinsically dependent on the choice of the quadrature parameter, k . This error component inevitably contributes to an increase in the variance of the estimator. In practical applications, a pragmatic strategy to mitigate this error is to progressively increase the value of k . Additionally, enlarging the sample size provides an effective means to reduce sampling variability and attain more accurate approximation results.

Table 1: True values, Parameter estimations and RMSEs with small degree of measurement error variance for two simulated and approximated methods (RMSEs in parentheses)

Parameters							
Method	β_0	β_x	β_z	σ_τ^2	σ_e^2	γ_{01}	γ_{10}
True Value	0.02	0.6	0.5	2.4	0.5	0.05	0.05
MCEM	0.028 (0.060)	0.687 (0.086)	0.5069 (0.062)	2.707 (0.505)	0.665 (0.166)	0.040 (0.040)	0.039 (0.041)
MGHQ2	0.020 (0.055)	0.587 (0.084)	0.486 (0.050)	2.257 (0.480)	0.655 (0.158)	0.046 (0.023)	0.042 (0.022)
MGHQ3	0.017 (0.002)	0.591 (0.075)	0.489 (0.054)	2.309 (0.512)	0.651 (0.155)	0.049 (0.026)	0.042 (0.021)
MGHQ4	0.025 (0.010)	0.577 (0.087)	0.485 (0.058)	2.265 (0.567)	0.650 (0.156)	0.039 (0.026)	0.043 (0.021)
MGHQ5	0.020 (0.010)	0.570 (0.087)	0.485 (0.058)	2.265 (0.567)	0.650 (0.156)	0.049 (0.026)	0.042 (0.021)
MGHQ6	0.020 (0.010)	0.571 (0.087)	0.485 (0.058)	2.265 (0.567)	0.650 (0.156)	0.049 (0.026)	0.042 (0.021)
MGHQ7	0.020 (0.010)	0.570 (0.087)	0.485 (0.058)	2.265 (0.567)	0.650 (0.156)	0.049 (0.026)	0.042 (0.021)
MGHQ8	0.020 (0.010)	0.570 (0.087)	0.485 (0.058)	2.265 (0.567)	0.650 (0.156)	0.049 (0.026)	0.042 (0.021)
MGHQ9	0.020 (0.010)	0.570 (0.087)	0.485 (0.058)	2.265 (0.567)	0.650 (0.156)	0.049 (0.026)	0.042 (0.021)
MGHQ10	0.020 (0.010)	0.507 (0.087)	0.485 (0.058)	2.265 (0.567)	0.650 (0.156)	0.049 (0.026)	0.042 (0.021)
MGHQ15	0.020 (0.010)	0.500 (0.087)	0.485 (0.058)	2.265 (0.567)	0.650 (0.156)	0.049 (0.026)	0.042 (0.021)
MGHQ20	0.020 (0.010)	0.500 (0.087)	0.485 (0.058)	2.265 (0.567)	0.650 (0.156)	0.049 (0.026)	0.042 (0.021)

Discussion and Results

This paper investigates maximum approximated likelihood (MAL) estimators that extend the framework of Monte Carlo simulated likelihood (MSL) estimators for likelihood inference in generalized linear mixed models with repeatedly measured covariates subject to measurement error and misclassified binary responses. A principal advantage of MSL lies in the asymptotic properties guaranteed under broad regularity conditions due to its reliance on Monte Carlo simulation methods. However, the substantial computational burden required to attain sufficient approximation accuracy often renders this approach impractical. Consequently, more precise numerical approximation techniques, such as Gaussian quadrature, have become standard practice. The MAL estimators developed herein demonstrate robustness in handling measurement error and misclassification, which are common issues in real-world multilevel data. This robustness enhances the practical applicability of the generalized linear mixed model framework in complex settings. We have delineated a set of conditions on the approximation method that ensure the consistency and asymptotic efficiency of MAL estimators. Overall, simulation results demonstrate that the Multivariate Gauss-Hermite Quadrature (MGHQ) method yields superior parameter estimates compared to the simulation-based MCEM approach. Moreover, the MGHQ method significantly reduces computational time while maintaining or improving estimation accuracy, thereby making it a more viable choice for large-scale or high-dimensional data analyses.

References

- Ahangari, M., Golalizadeh, M., and Rezaei Ghahroodi, Z. (2019), Likelihood inference in the random effects logistic regression with response misclassification and Covariate subject to measurement error, *Journal of Statistical Research of Iran*, **16**(1), 255-286.
- Buonaccorsi, J.P. (2010), *Measurement Error, Models, Methods and Applications*, New York: CRC Press.
- Griebel M., Heiss F., Oettershagen, J. and Weiser, C. (2019), *Maximum Approximated Likelihood Estimation*, Bonn, Institute for Numerical Simulation.
- Heiss F. and Winschel V. (2008), Likelihood approximation by numerical integration on sparse grids, *Journal of Econometrics*, **144**, 62-80.
- Neuhaus, J.M. (2002), Analysis of clustered and longitudinal binary data subject to response misclassification, *Biometrics*, **58**(3), 675-683.
- Newey W. and McFadden D. (1994), *Large Sample Estimation and Hypothesis Testing*, *Handbook of Econometrics*, Amsterdam, Elsevier.
- Paulino, C. D., Soares, P. and Neuhaus, J. (2003), Binomial regression with misclassification, *Biometrics*, **59**(3), 670-675.
- Roy, S. (2012), Accounting for response misclassification and covariate measurement error using a random effect logit model, *Communications in Statistics-Simulation and Computation*, **41**(9), 1623-1636.
- Tang, L., Lyles, R.H., King, C.C., Hogan, J.W. and Lo, Y. (2015), Regression analysis for differentially misclassified correlated binary outcomes, *Journal of the Royal Statistical Society. Series C, Applied Statistics*, **64**(3), 433-449.
- Xie, X., Xue, X. and Strickler, H.D. (2017), Generalized linear mixed model for binary outcomes when covariates are subject to measurement errors and detection limits, *Statistics in Medicine*, **37**, 119-136.
- Yi, G.Y. (2017), *Statistical Analysis with Measurement Error or Misclassification*. New York: Springer.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



The Law of Iterated Logarithm for the Smooth Estimator of Distribution Function in Length-Biased Data

Masoud Ajami^{*1}, Mahboubeh Akbari²

¹ Department of Statistics, Faculty of Mathematical Sciences, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran.

² Department of Statistics, University of Pretoria, Pretoria, 0002, Gauteng, South Africa.

Abstract

This paper establishes the law of iterated logarithm (LIL) for the smooth estimator of distribution function in length-biased sampling. We prove that under certain regularity conditions, the smooth estimator achieves the same LIL behavior as the non-smooth Cox estimator with rate $2\mu\gamma$. The main result extends previous work on uniform consistency by providing precise asymptotic behavior of the supremum deviation between the smooth estimator and the true distribution function.

Keywords: Length-biased data, smooth estimator, law of iterated logarithm, distribution function, kernel estimation.

Mathematics Subject Classification (2020): 62G05, 62G20.

1 Introduction

Suppose X is a non-negative random variable with probability density function $f(\cdot)$. Length-biased data, which are obtained from length-biased distribution, arise in many practical situations including survival analysis, econometrics, renewal processes, biomedicine and physics. For instance, if the probability of an item being selected for the sample is proportional to its length (X), then the distribution of the observed length is length-biased.

In the case of length-biased data we denote Y_w by Y and $g_w(t, 1) = g(t)$ for simplicity. Therefore,

$$g(t) = \frac{tf(t)}{\mu},$$

where $\mu = E(X)$ is assumed finite.

Let f and g have distribution functions (d.f.) F and G , respectively. It is clear that the following relationship between F and its length-biased distribution G holds:

$$G(t) = \frac{1}{\mu} \int_0^t x dF(x), \quad t \geq 0.$$

^{*}Corresponding author, m.ajami@vru.ac.ir

Throughout this paper we assume that G is continuous on $\mathbb{R}^+ = [0, \infty)$. From this it can be concluded that F is also continuous. An elementary calculation shows that F is determined uniquely by G , namely

$$F(t) = \mu \int_0^t y^{-1} dG(y), \quad t \geq 0.$$

Let Y_1, \dots, Y_n be a sample from G . The empirical estimator of F proposed by [Cox \(1969\)](#) is

$$\begin{aligned} F_n^c(t) &= \mu_n \int_0^t y^{-1} dG_n(y) \\ &= \frac{\mu_n}{n} \sum_{i=1}^n \frac{1}{Y_i} I(Y_i \leq t), \end{aligned} \quad (1)$$

where

$$\begin{aligned} \mu_n^{-1} &= \int_0^\infty y^{-1} dG_n(y) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{Y_i}, \end{aligned}$$

and G_n is the empirical d.f. estimator of G given by

$$G_n(t) = \frac{1}{n} \sum_{i=1}^n I(Y_i \leq t),$$

where $I(A)$ denotes the indicator of event A .

[Vardi \(1982\)](#) showed that Cox's estimator is a non-parametric maximum likelihood estimator for the target distribution $F(\cdot)$. [Horváth \(1985\)](#) established strong uniform consistency without rate of $F_n^c(\cdot)$ and showed that under the condition $E(Y^{-1}) < \infty$,

$$\lim_{n \rightarrow \infty} \sup_{0 < t < \infty} |F_n^c(t) - F(t)| = 0 \quad a.s. \quad (2)$$

The empirical estimator $F_n^c(\cdot)$ is a step function and discontinuous. In the absence of length-biased data, where F_n is an empirical d.f. of a random sample from F distribution, strong consistency of the smoothed empirical d.f. estimator has been proved by [Fernholz \(1991\)](#). Also, its weak convergence has been proved by [Van der Vaart \(2000\)](#). Since smoothed estimators have better performance compared to non-smoothed estimators, we define the kernel smoothed version of $F_n^c(\cdot)$ through the convolution of two functions $F_n^c(\cdot)$ and CDF $L(\cdot)$.

Using (1) we can define the smoothed estimator F_n^s through the convolution of two functions F_n^c and d.f. L , i.e.

$$\begin{aligned} F_n^s(t) &= \int_0^\infty L\left(\frac{t-y}{h_n}\right) dF_n^c(y) \\ &= \frac{\mu_n}{n} \sum_{i=1}^n \frac{1}{Y_i} L\left(\frac{t-Y_i}{h_n}\right), \end{aligned} \quad (3)$$

where h_n is an arbitrary sequence of smoothing parameters (or bandwidths) that satisfies $\lim_{n \rightarrow \infty} h_n = 0$ and L is the cumulative form of a kernel density function K , i.e.

$$L(t) = \int_0^t K(x) dx.$$

Smoothed estimators have better performance compared to non-smoothed estimators. [Jahanshahi et al. \(2017\)](#) investigated uniform consistency and asymptotic normality of (3) and constructed a one-sample Kolmogorov-type goodness of fit test with this estimator for length-biased data. Also [Zamini et al. \(2024\)](#) established a Berry-Esseen type bound for the smoothed estimator in this setting. Recently [Ajami et al. \(2024\)](#) have proved the extended Glivenko-Cantelli theorem for $F_n^c(\cdot)$ and $F_n^s(\cdot)$.

The purpose of this paper is to establish the law of iterated logarithm (LIL) for the smooth estimator $F_n^s(\cdot)$ in length-biased sampling, generalizing the strong consistency result in (2). The proof relies on key assumptions about the kernel function, bandwidth, and distribution properties, which are outlined below:

Assumptions.

A1. The kernel function K is symmetric, of bounded variation on $(-1, 1)$ with total variation $V_K < \infty$. In addition $K(t) = 0$ if $t \notin (-1, 1)$ and satisfies the following conditions:

$$\begin{aligned} \int_{-1}^1 K(t) dt &= 1, \\ \int_{-1}^1 tK(t) dt &= 0, \\ \int_{-1}^1 t^2 K(t) dt &= m < \infty, \end{aligned}$$

A2. $\int_0^\infty u^{-2} G^{1/r}(u) du < \infty$, for some $r > 2$.

A3. $\sup_{0 < t < \infty} |f'(t)| \leq c$.

A4. $E(Y^{-2}) = \int_0^\infty y^{-2} dG(y) < \infty$.

A5. $\int_0^\infty u^{-2-\delta} dG(u) < \infty$, for some $\delta > 0$.

Discussion on the Assumptions. Assumptions **A1** and **A2** are widely used in kernel estimation. Assumption **A3** is not a strong assumption for h_n but it is an essential assumption for validation of the results. If we choose $h_n = n^{-\alpha}$ for $3^{-1} < \alpha < 2^{-1}$, **A1** is satisfied.

2 Main Results

In this section we discuss the LIL property for the smooth estimator F_n^s . Before that, in the following Lemma, we derive the property of the LIL for the non-smooth Cox's estimator.

Lemma 2.1. Suppose that Assumption (A4) is satisfied. If we let

$$b_n = (2 \log \log n)^{-1/2},$$

and

$$\alpha_n(t) = \sqrt{n}(F_n^c(t) - F(t))$$

then

$$\limsup_{n \rightarrow \infty} \sup_{0 < t < \infty} b_n |\alpha_n(t)| = 2\mu\gamma \quad a.s. \quad (4)$$

where $\gamma = \int_0^\infty y^{-2} dG(y) - \mu^{-2}$.

Proof. We define $H(t) = \int_0^t y^{-1} dG(y)$ and $H_n(t) = \int_0^t y^{-1} dG_n(y)$, so we can write

$$H_n(t) - H(t) = \frac{1}{n} \sum_{i=1}^n [Y_i^{-1} I(Y_i \leq t) - E(Y^{-1} I(Y \leq t))], \quad (5)$$

and define

$$\xi_{(t)}^2 = \text{Var}(Y_i^{-1} I(Y_i \leq t)) = \int_0^t y^{-2} dG(y) - \left(\int_0^t y^{-1} dG(y) \right)^2. \quad (6)$$

And also we define

$$\gamma^2 = \xi_{(\infty)}^2 = \int_0^\infty y^{-2} dG(y) - \mu^{-2}. \quad (7)$$

Now according to the classical LIL

$$\limsup_{n \rightarrow \infty} \sup_{0 < t < \infty} \frac{\sqrt{n}[H_n(t) - H(t)]}{\sqrt{2\xi_{(t)}^2 \log \log n}} = 1 \quad a.s., \quad (8)$$

and

$$\limsup_{n \rightarrow \infty} \frac{\sqrt{n}[\mu - \mu_n]}{\sqrt{2\gamma\mu \log \log n}} = 1 \quad a.s., \quad (9)$$

so under Assumption (A4) we have $E(Y^{-2}) < \infty$ then we can write

$$(F_n^c(t) - F(t)) = (1) + (2) + (\mu_n - \mu)[H_n(t) - H(t)] \quad (10)$$

in which

$$(1) = \mu[H_n(t) - H(t)], \quad (11)$$

and

$$(2) = (\mu_n - \mu)H(t). \quad (12)$$

On the other hand, based on (8) and (9) one can see that

$$\limsup_{n \rightarrow \infty} \sup_{0 < t < \infty} [\mu - \mu_n][H_n(t) - H(t)] = O\left(\frac{\log \log n}{n}\right) \quad a.s. \quad (13)$$

Now based on the LIL property for (1) and (2), it can be shown that

$$\begin{aligned} \left\{\frac{n}{\log \log n}\right\}^{1/2} |F_n^c(t) - F(t)| &= \mu \left\{\frac{n}{\log \log n}\right\}^{1/2} [H_n(t) - H(t)] \\ &+ \left\{\frac{n}{\log \log n}\right\}^{1/2} (\mu_n - \mu) + O\left(\frac{\sqrt{\log \log n}}{n}\right) \\ &= \sqrt{2}\mu\xi_{(t)}^2 + \sqrt{2}\mu\gamma + O\left(\frac{\sqrt{\log \log n}}{n}\right) \quad a.s. \end{aligned} \quad (14)$$

So, based on (14) the proof is completed. \square

Lemma 2.2. Suppose that Assumptions (A2), (A3) and (A5) are satisfied. Then

$$\sup_{t,s \in J_n} |\alpha_n(t) - \alpha_n(s)| = O(n^{-\frac{1}{4}} \log n), \quad a.s., \quad (15)$$

where $J_n = \{t, s : |t - s| \leq Cn^{-\frac{1}{2}}\}$, and C is a positive number.

Proof. cf. Lemma 2 Fakoor et al. (2011). \square

Theorem 2.3. Under Assumptions (A1), (A3), $\gamma < \infty$ and

$$\lim_{n \rightarrow \infty} \frac{\sqrt{n}h_n^2}{\sqrt{\log \log n}} = 0, \quad (16)$$

then

$$\limsup_{n \rightarrow \infty} \sup_{0 < t < \infty} \left\{\frac{n}{2 \log \log n}\right\}^{1/2} |F_n^s(t) - F(t)| = 2\mu\gamma \quad a.s.$$

Proof. Using integration by parts, for any $t > 0$ and large n , one can see that

$$\begin{aligned} \sqrt{n}(F_n^s(t) - F(t)) &= \sqrt{n} \int_{-1}^1 \{F_n^c(t - h_n u) - F(t)\} K(u) du \\ &= \int_{-1}^1 \{\alpha_n(t - h_n u) - \alpha_n(t)\} K(u) du \\ &+ \int_{-1}^1 \alpha_n(t) K(u) du + \sqrt{n} \int_{-1}^1 \{F(t - h_n u) - F(t)\} K(u) du \\ &= \int_{-1}^1 \{V_1(t) + V_2(t) + V_3(t)\} K(u) du. \end{aligned} \quad (17)$$

First, based on Lemma 2.1, we obtain that

$$\limsup_{n \rightarrow \infty} \sup_{t > 0} b_n \left| \int_{-1}^1 V_2(t) K(u) du \right| = \limsup_{n \rightarrow \infty} \sup_{0 < t < \infty} b_n \alpha_n(t) = 2\mu\gamma \quad a.s., \quad (18)$$

and secondly according to (16) and using Taylor expansion, we have

$$\limsup_{n \rightarrow \infty} \sup_{t > 0} b_n \left| \int_{-1}^1 V_3(t) K(u) du \right| = \sup_{t > 0} f''(t) \frac{\sqrt{n}h_n^2}{b_n} \left| \int_{-1}^1 u^2 K(u) du \right| = 0. \quad (19)$$

Finally based on Lemma 2.2

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sup_{t > 0} b_n \left| \int_{-1}^1 V_1(t) K(u) du \right| &\leq \limsup_{n \rightarrow \infty} b_n \sup_{t > 0} |\alpha_n(t - uh_n) - \alpha_n(t)| \left| \int_{-1}^1 K(u) du \right| \\ &\leq \limsup_{n \rightarrow \infty} \frac{\log n}{n^{\frac{1}{4}} \sqrt{2 \log \log n}}, \end{aligned}$$

so

$$\limsup_{n \rightarrow \infty} \sup_{t > 0} b_n \left| \int_{-1}^1 V_1(t) K(u) du \right| = 0 \quad a.s. \quad (20)$$

Now, from (17) and using the triangle inequality, it can be stated that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sup_{t > 0} b_n n^{1/2} \left| F_n^s(t) - F(t) \right| &\leq \limsup_{n \rightarrow \infty} \sup_{t > 0} b_n \left| \int_{-1}^1 V_2(t) K(u) du \right| \\ &\quad - \limsup_{n \rightarrow \infty} \sup_{t > 0} b_n \left| \int_{-1}^1 \{V_1(t) + V_3(t)\} K(u) du \right|. \end{aligned} \quad (21)$$

So from (18)-(20) it is obvious that

$$\limsup_{n \rightarrow \infty} \sup_{0 < t < \infty} \left\{ \frac{2n}{\log \log n} \right\}^{1/2} |F_n^s(t) - F(t)| \leq 2\mu\gamma \quad a.s. \quad (22)$$

On the other hand, to prove the lower bound, using the reversed triangle inequality we derive from (17),

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sup_{t > 0} b_n n^{1/2} \left| F_n^s(t) - F(t) \right| &\geq \limsup_{n \rightarrow \infty} \sup_{t > 0} b_n \left| \int_{-1}^1 V_2 K(u) du \right| \\ &\quad - \limsup_{n \rightarrow \infty} \sup_{t > 0} b_n \left| \int_{-1}^1 \{V_1 + V_3\} K(u) du \right|. \end{aligned} \quad (23)$$

By reapplying formulas (18)-(20), we have

$$\limsup_{n \rightarrow \infty} \sup_{0 < t < \infty} \left\{ \frac{2n}{\log \log n} \right\}^{1/2} |F_n^s(t) - F(t)| \geq 2\mu\gamma \quad a.s. \quad (24)$$

So based on (22) and (24), the proof is completed. \square

References

- Ajmi M., Zamini R., and Jahanshahi S. M. A. (2024), The extended Glivenko-Cantelli property for Kernel-Smoothed estimator of the cumulative distribution function in the length-biased sampling, *Journal of Mahani Mathematical Research*, **13**(2), 535–545.
- Cox D. R. (1969), Some sampling problems in technology. In: Johnson N. L. and Smith H. Jr., *New Developments in Survey Sampling*, Wiley, New York, pp. 506–527.
- Fernholz L. T. (1991), Almost sure convergence of smoothed empirical distribution functions, *Scandinavian Journal of Statistics*, **22**, 255–262.
- Horváth L. (1985), Estimation from a length-biased distribution, *Statistics & Risk Modeling*, **3**(1–2), 91–114.
- Jahanshahi S. M. A., Rad A. and Fakoor V. (2017), Goodness-of-fit test under length-biased sampling, *Communications in Statistics–Theory and Methods*, **46**(15), 7580–7592.
- Fakoor, V., Bolbolian, M. and Azarnoosh, H. (2011), Asymptotic behaviors of the Lorenz curve and Gini index in sampling from a length-biased distribution, *Statistics and Probability Letters*, **9**(81), 1425–1435.
- Van der Vaart A. W. (2000), *Asymptotic Statistics*, Cambridge University Press.
- Vardi Y. (1982), Nonparametric estimation in renewal processes, *The Annals of Statistics*, **10**(3), 772–785.
- Zamini R., Ajami M. and Fakoor V. (2024), Berry-Esseen bound for smooth estimator of distribution function under length-biased data, *Communications in Statistics–Theory and Methods*, **53**(5), 1800–1809.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Parameters Estimation for Two Lomax Exponential Distributions Under Joint Type-II Censored Data

Soheila Akbari Bargoshadi^{*1}, Hossein Bevrani², Maryam Maroufi³

¹Department of Mathematics, Urmia University, Urmia, Iran

²Department of Statistics, University of Kurdistan, Sanandaj, Iran

³Department of Statistics, Islmaic Azad University, Marvdasht Branch, Marvdasht, Iran

Abstract

In this article, we investigate parameter estimation for two Lomax exponential distributions under a joint Type-II censoring scheme. This distribution is a flexible model capable of representing lifetime data with both increasing and decreasing trends. Using the Expectation-Maximization algorithm, we derive the maximum likelihood estimators for the parameters. We also compute approximate confidence intervals based on the missing information matrix. To evaluate the performance of the proposed method, we conduct a Monte Carlo simulation study across various sample sizes and demonstrate its practical utility through the analysis of a real dataset.

Keywords: Expectation-Maximization algorithm, Lomax exponential distribution, Joint type-II censoring scheme.

Mathematics Subject Classification (2020): 62F10, 62F25, 62N02.

1 Introduction

In reliability analysis and life-testing experiments, researchers often terminate tests before all units fail to reduce costs and save time. This results in censored data, where only partial failure information is available. A key censoring method is Type-II censoring, where testing stops once a predefined number of failures occur (see [Lawless \(2003\)](#)). This approach is especially useful when comparing failure times between two populations under the same environmental conditions. For example, [Balakrishnan and Rasouli \(2008\)](#) and [Rasouli and Balakrishnan \(2010\)](#) studied the failure times of air conditioning systems in two Boeing 720 jets using this method. In these cases, two independent samples of sizes m and n are drawn from their respective populations and subjected to simultaneous life-testing under identical conditions. The experiment concludes after a specified number of failures, let's say r , where $1 \leq r \leq N$

^{*}Corresponding author: soheila.akbari61@gmail.com

and $N = m + n$. This method allows for population comparison based on observed failures and is known as a joint Type-II censoring scheme.

Previous research has extensively investigated joint Type-II censoring schemes and their applications. Notably, [Balakrishnan and Rasouli \(2008\)](#) developed exact inference methods for two exponential distributions under a joint Type-II censoring scheme. Further advancing this field, [Akbari and *et al.* \(2023\)](#) focused on statistical inference for two Burr-XII distribution using the same censoring framework. More recently, [Qiu and Gui \(2023\)](#) extended these methodologies by examining inference techniques for two Gumbel distributions under joint Type-II censoring scheme.

The Lomax Exponential (LE) distribution, introduced by [Ijaz and *et al.* \(2019\)](#), is a two-parameter modification of the classical Lomax distribution. This flexible model is particularly useful for analyzing lifetime data with both increasing and decreasing hazard rate patterns. [Ijaz and *et al.* \(2019\)](#) conducted a comprehensive study of the LE distribution statistical properties, examining its moments, entropy measures, hazard and survival functions, as well as its median, mode, and order statistics.

Despite these thorough investigations, research on the LE distribution under censoring schemes remains limited. Addressing this gap, [Akbari and *et al.* \(2024\)](#) recently developed maximum likelihood estimators (MLEs) and approximate confidence intervals (ACIs) for the LE distribution parameters under Type-II censoring scheme. Building on this work, we continue their research by examining joint type-II censored data from two LE populations. Section 2 presents the joint Type-II censoring scheme model for two LE populations. MLEs of the unknown parameters are derived using the Expectation-Maximization (EM) algorithm, and ACIs are computed based on the fisher information matrix. Additionally, a simulation study and a real data application are provided in Sections 3 and 4, respectively, to evaluate the performance of the estimation method.

2 Model specification and likelihood inferences

Let $\mathbf{X} = (X_1, \dots, X_m)'$ and $\mathbf{Y} = (Y_1, \dots, Y_n)'$ denote the lifetime vectors of random samples from the first and second populations, respectively, where both populations follow a LE distribution. The probability density function f_i and survival function \bar{F}_i ($i = 1, 2$) for the i^{th} population ($i = 1, 2$) are given by

$$f_1(x; \theta_1, \theta_2) = \frac{\theta_1}{\theta_2} (1+x)e^x \left[1 + \frac{xe^x}{\theta_2}\right]^{-(\theta_1+1)}, x > 0, \quad \theta_1 > 0, \quad \theta_2 > 0, \quad (1)$$

$$\bar{F}_1(x; \theta_1, \theta_2) = \left[1 + \frac{xe^x}{\theta_2}\right]^{-\theta_1}, x > 0, \quad \theta_1 > 0, \quad \theta_2 > 0, \quad (2)$$

$$f_2(y; \theta_3, \theta_4) = \frac{\theta_3}{\theta_4} (1+y)e^y \left[1 + \frac{ye^y}{\theta_4}\right]^{-(\theta_3+1)}, y > 0, \quad \theta_3 > 0, \quad \theta_4 > 0 \quad (3)$$

$$\bar{F}_2(y; \theta_3, \theta_4) = \left[1 + \frac{ye^y}{\theta_4}\right]^{-\theta_3}, y > 0, \quad \theta_3 > 0, \quad \theta_4 > 0, \quad (4)$$

where θ_i ($i = 1, 3$) and θ_j ($j = 2, 4$) control the shape and scale parameters of the distribution, respectively. Consider $N = m + n$ as the total number of selected samples from the two populations, where r denotes the total number of failure or observed units. Let $W_1 < W_2 < \dots < W_N$ represent the ordered statistics of the combined N units. Under the joint Type-II censoring scheme, the observed data is represented by the pair (\mathbf{W}, \mathbf{S}) , where $\mathbf{W} = (W_1, W_2, \dots, W_r)$ are the first r ordered failure times and $\mathbf{S} = (S_1, S_2, \dots, S_r)$ is the indicator vector such that $S_i = 1$ if W_i belongs to X -failures and $S_i = 0$ if W_i belongs to Y -failures for $i = 1, 2, \dots, r$. Let $M_r = \sum_{i=1}^r S_i$ and $N_r = \sum_{i=1}^r (1 - S_i) = r - M_r$ denote the

number of failures from the X and Y samples, respectively. The likelihood function for the observed data based on the joint density function of (\mathbf{W}, \mathbf{S}) is

$$L(\boldsymbol{\theta}; \mathbf{w}, \mathbf{s}) = \frac{m!n!}{(m-m_r)!(n-n_r)!} \prod_{i=1}^r [f_1(w_i; \theta_1, \theta_2)]^{s_i} [f_2(w_i; \theta_3, \theta_4)]^{1-s_i} \\ \times [\bar{F}_1(w_r; \theta_1, \theta_2)]^{m-m_r} [\bar{F}_2(w_r; \theta_3, \theta_4)]^{n-n_r},$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)'$ and f_i and \bar{F}_i ($i = 1, 2$) are defined in equations (1)-(4). Then logarithm of the likelihood function without the constant term is given by

$$l_1 = \ln L(\boldsymbol{\theta}; \mathbf{w}, \mathbf{s}) \propto m_r \ln(\theta_1) + \theta_1 m \ln(\theta_2) + \sum_{i=1}^r \ln(w_i + 1) + \sum_{i=1}^r w_i + n_r \ln(\theta_3) + \theta_3 n \ln(\theta_4) \\ - (\theta_1 + 1) \sum_{i=1}^r s_i \ln(\theta_2 + w_i e^{w_i}) - (\theta_3 + 1) \sum_{i=1}^r (1 - s_i) \ln(\theta_4 + w_i e^{w_i}) \\ - \theta_1 (m - m_r) \ln(\theta_2 + w_r e^{w_r}) - \theta_3 (n - n_r) \ln(\theta_4 + w_r e^{w_r}). \quad (5)$$

To obtain the MLEs of the parameters θ_i ($i = 1, 2, 3, 4$), differentiation of equation (5) with respect to the parameters is performed. This leads to obtaining the likelihood equations in the following form

$$\frac{\partial l_1}{\partial \theta_1} = \frac{m_r}{\theta_1} + m \ln(\theta_2) - (m - m_r) \ln(\theta_2 + w_r e^{w_r}) - \sum_{i=1}^r s_i \ln(\theta_2 + w_i e^{w_i}) = 0, \quad (6)$$

$$\frac{\partial l_1}{\partial \theta_2} = \frac{m \theta_1}{\theta_2} - \frac{\theta_1 (m - m_r)}{\theta_2 + w_r e^{w_r}} - (\theta_1 + 1) \sum_{i=1}^r s_i (\theta_2 + w_i e^{w_i})^{-1} = 0, \quad (7)$$

$$\frac{\partial l_1}{\partial \theta_3} = \frac{n_r}{\theta_3} + n \ln(\theta_4) - (n - n_r) \ln(\theta_4 + w_r e^{w_r}) - \sum_{i=1}^r (1 - s_i) \ln(\theta_4 + w_i e^{w_i}) = 0, \quad (8)$$

$$\frac{\partial l_1}{\partial \theta_4} = \frac{n \theta_3}{\theta_4} - \frac{\theta_3 (n - n_r)}{\theta_4 + w_r e^{w_r}} - (\theta_3 + 1) \sum_{i=1}^r (1 - s_i) (\theta_4 + w_i e^{w_i})^{-1} = 0. \quad (9)$$

As closed-form solutions are unavailable to solve the above equations, numerical methods are required to compute the MLEs. The EM algorithm is used as an iterative estimation approach for censored data, with implementation details provided in Subsection 2.1.

2.1 The EM algorithm

The EM algorithm is particularly useful for parameter estimation in models with incomplete data, such as missing or censored observations, which commonly occur in real-world applications. This method was initially presented by [Dempster and *et al.* \(1977\)](#). This iterative method consists of two key steps: the expectation step (E-step) and the maximization step (M-step). During the E-step, the algorithm calculates the expected value of the censored variables based on the observed data, while in the M-step, it maximizes this expected value to determine the parameters. Censored data is displayed by $(\mathbf{Z}, \boldsymbol{\delta})$ where $\mathbf{Z} = (Z_1, \dots, Z_{N-r})$ and $\boldsymbol{\delta} = (\delta_1, \dots, \delta_{N-r})$ such that $\delta_j = 1$ and $\delta_j = 0$, if Z_j belongs to X and Y , respectively, where $j = 1, 2, \dots, N - r$. It is clear that $\sum_{j=1}^{N-r} \delta_j = m - M_r$ and $\sum_{j=1}^{N-r} (1 - \delta_j) = n - N_r$. The combined observed and censored data are used to represent the completed dataset as $\mathbf{C} = (\mathbf{W}, \mathbf{S}, \mathbf{Z}, \boldsymbol{\delta})$. According to complete data, the likelihood function is given by

$$L_C(\boldsymbol{\theta}; \mathbf{w}, \mathbf{s}, \boldsymbol{\delta}, \mathbf{z}) = \prod_{i=1}^r [f_1(w_i; \theta_1, \theta_2)]^{s_i} [f_2(w_i; \theta_3, \theta_4)]^{1-s_i} \prod_{j=1}^{N-r} [f_1(z_j; \theta_1, \theta_2)]^{\delta_j} [f_2(z_j; \theta_3, \theta_4)]^{1-\delta_j}, \quad (10)$$

then, logarithm of the likelihood function based on equation (10) is as follows:

$$\begin{aligned}
 l_2 = \ln L_C(\boldsymbol{\theta}; \mathbf{w}, \mathbf{s}, \boldsymbol{\delta}, \mathbf{z}) &= m \ln(\theta_1) + m\theta_1 \ln(\theta_2) + n \ln(\theta_3) + n\theta_3 \ln(\theta_4) + \sum_{i=1}^r \ln(w_i + 1) \\
 &+ \sum_{i=1}^r w_i + \sum_{j=1}^{N-r} \ln(z_j + 1) + \sum_{j=1}^{N-r} z_j - (\theta_1 + 1) \left[\sum_{i=1}^r s_i \ln(\theta_2 + w_i e^{w_i}) + \sum_{j=1}^{N-r} \delta_j \ln(\theta_2 + z_j e^{z_j}) \right] \\
 &- (\theta_3 + 1) \left[\sum_{i=1}^r (1 - s_i) \ln(\theta_4 + w_i e^{w_i}) + \sum_{j=1}^{N-r} (1 - \delta_j) \ln(\theta_4 + z_j e^{z_j}) \right]. \quad (11)
 \end{aligned}$$

Following the E-step, the conditional expectations from derivatives of equation (11) with respect to the unknown parameters given the observed data are calculated as follows:

$$\begin{aligned}
 E\left(\frac{\partial l_2}{\partial \theta_1} | z_j > w_r\right) &= \frac{m}{\theta_1} + m \ln(\theta_2) - \sum_{i=1}^r s_i \ln(\theta_2 + w_i e^{w_i}) \\
 &- \sum_{j=1}^{N-r} \delta_j E[\ln(\theta_2 + z_j e^{z_j}) | z_j > w_r] = 0, \quad (12)
 \end{aligned}$$

$$E\left(\frac{\partial l_2}{\partial \theta_2} | z_j > w_r\right) = \frac{m\theta_1}{\theta_2} - (\theta_1 + 1) \left[\sum_{i=1}^r \frac{s_i}{\theta_2 + w_i e^{w_i}} + \sum_{j=1}^{N-r} \delta_j E\left(\frac{1}{\theta_2 + z_j e^{z_j}} | z_j > w_r\right) \right] = 0, \quad (13)$$

$$\begin{aligned}
 E\left(\frac{\partial l_2}{\partial \theta_3} | z_j > w_r\right) &= \frac{n}{\theta_3} + n \ln(\theta_4) - \sum_{i=1}^r (1 - s_i) \ln(\theta_4 + w_i e^{w_i}) \\
 &- \sum_{j=1}^{N-r} (1 - \delta_j) E[\ln(\theta_4 + z_j e^{z_j}) | z_j > w_r] = 0, \quad (14)
 \end{aligned}$$

$$\begin{aligned}
 E\left(\frac{\partial l_2}{\partial \theta_4} | z_j > w_r\right) &= \frac{n\theta_3}{\theta_4} - (\theta_3 + 1) \left[\sum_{i=1}^r \frac{1 - s_i}{\theta_4 + w_i e^{w_i}} \right. \\
 &\left. + \sum_{j=1}^{N-r} (1 - \delta_j) E\left(\frac{1}{\theta_4 + z_j e^{z_j}} | z_j > w_r\right) \right] = 0. \quad (15)
 \end{aligned}$$

To compute the conditional expectations in equations (12)-(15), we require the conditional probability density function of $Z_j = z_j$ given $z_j > w_r$. This follows a left-truncated LE distribution (Ng and *et al.* (2002)), given by

$$f_{Z_j|w_r}(z_j | z_j > w_r) = \left\{ \frac{f_1(z_j; \theta_1, \theta_2)}{\bar{F}_1(w_r; \theta_1, \theta_2)} \right\}^{\delta_j} \left\{ \frac{f_2(z_j; \theta_3, \theta_4)}{\bar{F}_2(w_r; \theta_3, \theta_4)} \right\}^{1-\delta_j}. \quad (16)$$

According to equation (16), conditional probability density function reduces to the first component if $\delta_j = 1$ and to the second component if $\delta_j = 0$. The following conditional expectations can be calculated using equation (16) as follows:

$$\begin{aligned}
 E[\ln(\theta_2 + z_j e^{z_j}) | z_j > w_r] &= \ln(\theta_2 + w_r e^{w_r}) + \frac{1}{\theta_1}, \quad E\left[\frac{1}{\theta_2 + z_j e^{z_j}} | z_j > w_r\right] = \frac{\theta_1}{(\theta_1 + 1)(\theta_2 + w_r e^{w_r})}, \\
 E[\ln(\theta_4 + z_j e^{z_j}) | z_j > w_r] &= \ln(\theta_4 + w_r e^{w_r}) + \frac{1}{\theta_3}, \quad E\left[\frac{1}{\theta_4 + z_j e^{z_j}} | z_j > w_r\right] = \frac{\theta_3}{(\theta_3 + 1)(\theta_4 + w_r e^{w_r})}.
 \end{aligned}$$

Based on the M- step, let $\hat{\theta}_i^h$ denote the MLEs of parameter θ_i ($i = 1, 2, 3, 4$) at the h^{th} iteration. Then, the MLEs for the θ_i at the $(h + 1)^{th}$ iteration are given by

$$\hat{\theta}_1^{(h+1)} = \frac{m_r}{-m \ln(\hat{\theta}_2^h) + \sum_{i=1}^r s_i \ln(\hat{\theta}_2^h + w_i e^{w_i}) + (m - m_r) \ln(\hat{\theta}_2^h + w_r e^{w_r})}, \quad (17)$$

$$\hat{\theta}_2^{(h+1)} = \frac{m \hat{\theta}_1^{(h+1)}}{(\hat{\theta}_1^{(h+1)} + 1) \sum_{i=1}^r s_i (\hat{\theta}_2^h + w_i e^{w_i})^{-1} + (m - m_r) \hat{\theta}_1^{(h+1)} [\hat{\theta}_2^h + w_r e^{w_r}]^{-1}}, \quad (18)$$

$$\hat{\theta}_3^{(h+1)} = \frac{n_r}{-n \ln(\hat{\theta}_4^h) + \sum_{i=1}^r (1 - s_i) \ln(\hat{\theta}_4^h + w_i e^{w_i}) + (n - n_r) \ln(\hat{\theta}_4^h + w_r e^{w_r})}, \quad (19)$$

$$\hat{\theta}_4^{(h+1)} = \frac{n \hat{\theta}_3^{(h+1)}}{(\hat{\theta}_3^{(h+1)} + 1) \sum_{i=1}^r (1 - s_i) (\hat{\theta}_4^h + w_i e^{w_i})^{-1} + (n - n_r) \hat{\theta}_3^{(h+1)} [\hat{\theta}_4^h + w_r e^{w_r}]^{-1}}. \quad (20)$$

The iteration process in equations (17)-(20) continues until the desired convergence is reached when $|\hat{\theta}_1^{(h+1)} - \hat{\theta}_1^h| + |\hat{\theta}_2^{(h+1)} - \hat{\theta}_2^h| + |\hat{\theta}_3^{(h+1)} - \hat{\theta}_3^h| + |\hat{\theta}_4^{(h+1)} - \hat{\theta}_4^h| < \varepsilon$, for a small value of $\varepsilon > 0$.

2.2 Approximation confidence interval estimation

This subsection presents the computation of the fisher information matrix, a crucial component for constructing the variance-covariance matrix needed to derive confidence intervals for the unknown parameters. Following the missing information principle introduced by Louis (1982), we formulate the observed information matrix as follows:

$$\mathbf{I}_{\mathbf{W},\mathbf{S}}(\boldsymbol{\theta}) = \mathbf{I}_C(\boldsymbol{\theta}) - \mathbf{I}_{\mathbf{Z},\delta|\mathbf{W},\mathbf{S}}(\boldsymbol{\theta}), \quad (21)$$

where $\mathbf{I}_{\mathbf{W},\mathbf{S}}(\boldsymbol{\theta})$, $\mathbf{I}_C(\boldsymbol{\theta})$ and $\mathbf{I}_{\mathbf{Z},\delta|\mathbf{W},\mathbf{S}}(\boldsymbol{\theta})$ are the observed, complete, and missing information matrices, respectively, and $\mathbf{C} = (X_1, \dots, X_m, Y_1, \dots, Y_n)$. Based on the two LE distributions, the complete information matrix is presented in the following form

$$\mathbf{I}_C(\boldsymbol{\theta}) = \begin{bmatrix} \frac{m}{\theta_1^2} & -\frac{m}{\theta_2(\theta_1+1)} & 0 & 0 \\ -\frac{m}{\theta_2(\theta_1+1)} & \frac{m\theta_1}{\theta_2^2(\theta_1+2)} & 0 & 0 \\ 0 & 0 & \frac{n}{\theta_3} & -\frac{n}{\theta_4(\theta_3+1)} \\ 0 & 0 & -\frac{n}{\theta_4(\theta_3+1)} & \frac{n\theta_3}{\theta_4^2(\theta_3+2)} \end{bmatrix}, \quad (22)$$

and the missing information matrix can be computed as follows:

$$\mathbf{I}_{\mathbf{Z},\delta|\mathbf{W},\mathbf{S}}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{m-n_r}{\theta_1^2} & \frac{m_r-m}{(\theta_1+1)(\theta_2+w_r e^{w_r})} & 0 & 0 \\ \frac{m_r-m}{(\theta_1+1)(\theta_2+w_r e^{w_r})} & \frac{\theta_1(m-n_r)}{(\theta_1+2)(\theta_2+w_r e^{w_r})^2} & 0 & 0 \\ 0 & 0 & \frac{n-n_r}{\theta_3^2} & \frac{n_r-n}{(\theta_3+1)(\theta_4+w_r e^{w_r})} \\ 0 & 0 & \frac{n_r-n}{(\theta_3+1)(\theta_4+w_r e^{w_r})} & \frac{\theta_3(n-n_r)}{(\theta_3+2)(\theta_4+w_r e^{w_r})^2} \end{bmatrix}. \quad (23)$$

Using Relation 21 along with equations (22) and (23), we obtain the observed information matrix. A $(1 - \alpha)100\%$ ACIs for θ_i is then constructed by inverting the observed information matrix $\mathbf{I}_{\mathbf{W},\mathbf{S}}(\hat{\boldsymbol{\theta}})$ as $\hat{\theta}_i \pm z_{\alpha/2} \sqrt{\text{var}(\hat{\theta}_i)}$ ($i = 1, 2, 3, 4$). $z_{\alpha/2}$ denotes the upper $\alpha/2$ quantile of the standard normal distribution, $\text{var}(\hat{\theta}_i)$ is the i^{th} diagonal element of $(\mathbf{I}_{\mathbf{W},\mathbf{S}}(\hat{\boldsymbol{\theta}}))^{-1}$ for $i = 1, 2, 3, 4$ and $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4)'$ is the vector of MLEs of the vector $\boldsymbol{\theta}$.

3 Simulation Study

A Monte Carlo simulation study was performed in R (R Core Team (2020)) to evaluate the performance of the estimation methodology. The simulation involved 1000 replications ($n_s =$

10^3), with data generated from the quantile functions $Q(u, \theta_1 = 0.5, \theta_2 = 1)$ and $Q(u, \theta_3 = 0.8, \theta_4 = 1.5)$, where $Q(u, a, b) = W(b((1-u)^{-1/a} - 1))$, $W(\cdot)$ denotes the Lambert w function and $u \sim \text{Uniform}(0,1)$. The study examined various Type-II joint censoring schemes and sample sizes, specified as: $(m, n, r) = \{(25, 20, 25), (25, 20, 30), (40, 40, 25), (40, 40, 30)\}$, where m and n denote the sizes of the first and second samples respectively, and r is the censoring point. The MLEs of the parameters were computed using the EM algorithm. Additionally, the ACIs were calculated for the unknown parameters via the missing information matrix. Table 1 shows the bias and mean squared error (MSE) and Table 2 shows the upper and lower limits (L and U, respectively), and interval length (IL) for the parameters. The bias of the estimated parameters is determined as $\text{bias}(\hat{\theta}_i) = \hat{\theta}_i - \theta_i$, ($i = 1, 2, 3, 4$) and the MSE is defined as

$$\text{MSE}(\hat{\theta}_i) = \frac{1}{n_s} \sum_{j=1}^{n_s} (\hat{\theta}_{i,j} - \theta_i)^2, \quad i = 1, 2, 3, 4.$$

According to Table 1, it is evident that EM estimations have demonstrated favorable performance in terms of bias and MSE. Moreover, based on the mentioned Tables, as the values of m , n , and r increase, there is a general reduction in bias, MSE and IL.

Table 1: Bias and MSE of EM estimates

(m, n)	r	$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_3$		$\hat{\theta}_4$	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
(25, 20)	25	0.1415	0.0809	0.3675	1.0545	0.1272	0.1333	0.2650	1.4726
	30	0.1129	0.0616	0.3856	0.9116	0.0749	0.1051	0.1643	0.4734
(40, 40)	25	0.1057	0.0419	0.2510	0.1187	0.1125	0.0823	0.1104	0.0864
	30	0.1011	0.0352	0.1775	0.0670	0.0700	0.0436	0.1166	0.0479

Table 2: 95% L and U bound estimates and IL for parameters based on the EM method

(m, n)	r	θ_1			θ_2			θ_3			θ_4		
		L	U	IL	L	U	IL	L	U	IL	L	U	IL
(25, 20)	25	0.1193	1.1065	0.9872	0.3699	2.4013	2.0314	0.2379	1.5119	1.2740	0.6334	2.6952	2.0618
	30	0.1766	1.1065	0.9299	0.0000	2.8125	2.8125	0.4966	1.3577	0.8611	1.2736	2.2565	0.9829
(40, 40)	25	0.2798	0.9224	0.6426	0.4932	1.8617	1.3684	0.4768	1.2633	0.7866	1.1212	2.1121	0.9910
	30	0.3264	0.8849	0.5585	0.6998	1.8022	1.1024	0.5355	1.2896	0.7541	1.1983	2.0225	0.8242

4 Application to Real Data

In this section, we evaluate the proposed estimation method using a real dataset from [Ijaz and et al. \(2019\)](#), which records losses caused by wind catastrophes in 1977. The dataset was randomly split into two subsets, as shown in Table 3. To assess whether both subsets follow the LE distribution, we performed the Kolmogorov-Smirnov (K-S) test using the "fitdistrplus" package in R. The results, presented in Table 4, indicate a good fit for the LE distribution based on the p-values. Furthermore, Table 5 demonstrates that as the value of r increases, both bias and MSE decrease. Additionally, Table 6 shows that higher values of r , generally, lead to a reduction in IL while maintaining coverage of the true dataset parameters.

Table 3: Losses due to wind catastrophes

Group	Data																			
X	27	2	17	5	2	32	43	3	4	2	2	6	6	4	3	2	2	2	2	24
Y	2	2	2	2	3	3	4	5	5	5	6	6	8	8	9	15	22	23	25	

Table 4: MLEs and K-S test results for real dataset

Data			K-S distance	p-value
X	$\hat{\theta}_1=0.1014$	$\hat{\theta}_2=3.9103$	0.27086	0.1063
Y	$\hat{\theta}_3=0.1318$	$\hat{\theta}_4=11.7951$	0.12695	0.9194

Table 5: Bias and MSE of estimations based on the EM method for real dataset

		$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_3$		$\hat{\theta}_4$	
(m, n)	r	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
(20, 19)	31	-0.0014	0.0002	-0.0103	0.0003	-0.0318	0.0010	-0.0951	0.0090
	35	0.0006	0.0001	-0.0003	0.0002	-0.0018	0.0001	-0.0151	0.0002

Table 6: 95% L and U bound estimates and its IL based on the EM method for real dataset

		θ_1			θ_2			θ_3			θ_4		
(m, n)	r	L	U	IL	L	U	IL	L	U	IL	L	U	IL
(20, 19)	31	0.0607	0.1393	0.0786	0.0000	10.6097	10.6097	0.0000	0.2353	0.2353	0.0000	34.6607	34.6607
	35	0.0634	0.1406	0.0771	0.0000	10.5356	10.5356	0.0000	0.2656	0.2656	0.0000	30.9332	30.9332

Discussion and Results

This paper addresses the estimation of unknown parameters for two LE distributions under a joint Type-II censoring scheme. Since the MLEs lack a closed-form solution, the EM algorithm was employed. ACIs were subsequently derived using the missing information matrix. Simulation results indicated that both the MSE and IL generally decrease as the values of m , n and r increase. Furthermore, the proposed method was applied to a real dataset, confirming the accuracy of the simulation findings.

References

- Akbari B.S., Bevrani H. and Arabi B.R. (2023), Statistical inference under joint type-II censored data from two Burr-XII populations, *Communications in Statistics-Simulation and Computation*, 1-21.
- Akbari B.S., Bevrani H. and Gholami H. (2024), *Parameter Estimation for Lomax Exponential Distribution with Type-II Censored Data*, 17th Iranian Statistics Conference, University of Birjand, <https://conf.birjand.ac.ir/isc17>.
- Balakrishnan N. and Rasouli A. (2008), Exact likelihood inference for two exponential populations under joint type-II censoring, *Computational Statistics and Data Analysis*, 52(5), pp.2725-2738.

- Dempster A.P., Laird N.M. and Rubin D.B. (1977), Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society: Series B (Methodological)*, **39**, 1, 1-22.
- Ijaz M., Asim S.M. and Alamgir. (2019), *Lomax exponential distribution with an application to real-life data*, PloS one, **14**, 12, p.e0225827.
- Lawless J.F. (2003), *Statistical models and methods for lifetime data*, 2nd ed, New York: John Wiley.
- Louis T.A. (1982), Finding the observed information matrix when using the EM algorithm, *Journal of the Royal Statistical Society: Series B (Methodological)*, **44**, 2, 226-233.
- Ng H.K.T., Chan, P.S. and Balakrishnan N. (2002), Estimation of parameters from progressively censored data using EM algorithm, *Computational Statistics and Data Analysis*, **39**, 4, 371-386.
- Qiu Y. and Gui W. (2023), Statistical inference for two Gumbel type-II distributions under joint type-II censoring scheme. *Axioms*, 12(6), p.572.
- Rasouli A. and Balakrishnan N. (2010), Exact likelihood inference for two exponential populations under joint progressive type-II censoring. *Communications in Statistics-Theory and Methods*, 39(12), pp.2172-2191.
- Team R.C. (2020). C R: *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, <https://www.R-project.org/>. [Google Scholar].



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



From probability uncertainty to nonlinear PDEs

Bahar Akhtari^{*1,2}

¹Department of Applied Mathematics and Computer Science, Faculty of Mathematics and Statistics,
University of Isfahan, Isfahan, Iran

² School of Mathematics, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran

Abstract

The problem of pricing a contingent claim/annuity contract under model uncertainty is determined by G -conditional expectation of the associated discounted payoff. In this paper, by generalizing a nonlinear Feynman–Kac formula, we show that this sublinear expectation is a viscosity solution of a nonlinear parabolic PDE. This result demonstrates a nice and deep relation between the theory of probability and statistics under uncertainty and second order fully nonlinear parabolic equations.

Keywords: Model uncertainty, G -Brownian motion calculus, Nonlinear Feynman-Kac theorem

Mathematics Subject Classification (2020): 60G65, 60H30

1 Introduction

It is important to measure the uncertain quantities in the real world, such as uncertainties in statistics, measures of risk, and also fully nonlinear super-hedging in finance. A strong mathematical response to such problems is probability theory. However, in some circumstances, probability measure can not be determined in a certain way due to the imperfect knowledge. This situation, where estimators should consider a non-dominated family of probability measures rather than a specific prior, is called *probability model uncertainty* or *ambiguity* and *Knightian uncertainty* in the theory of economics. The book Knight (1921) is a comprehensive discussion about risk and uncertainty. The word "Knightian uncertainty" stems from the name of *Frank Knight* due to his contribution in developing the concept of uncertainty.

The G -expectation theory is one of the frequently used tools to study the attractive and challenging problems including financial pricing and risk management under model uncertainty. In this framework, the notation of sublinear expectation $\hat{\mathbb{E}}$ is indicated which is a monotone, sub additive, constant preserving functional defined on a linear space of random variables.

The pricing of contingent claims in fixed income markets or in insurance modeling under probability uncertainty has been investigated in Holzermann (2021), Holzermann (2022) in the G -setting. Moreover, the authors in Hu and et al. (2014) present a relation between

*Corresponding author, b.akhatri@mcs.ui.ac.ir, b.akhtari@ipm.ir

G -conditional expectation and solutions of (nonlinear) PDEs where the Lipschitz continuity of some functionals is assumed. In the setting of short rate models under a fixed probability measure, a Feynman–Kac formula is proposed in Filipovic (2009) which starting from the PDE associated to the underlying SDE, and proving that its solution can be seen as the conditional expectation of the discounted payoff.

In this paper, inspired by Filipovic (2009), we provide a Feynman–Kac formula under probability uncertainty to include an additional linear term in the associated PDE due to discounting whose coefficients are given in a setting not satisfying the hypothesis of Hu and et al. (2014).

1.1 Preliminaries

The G -expectation theory provides very rich and interesting tools to study the model uncertainty. In this section, we briefly discuss the essential preliminaries of this calculus. The reader can refer to Peng (2019) for more details.

Let Ω be a given set and we denote by \mathcal{H} set of random variables which is a vector lattice of real valued functions defined on Ω . Here, \mathcal{H} is a linear space such that $X \in \mathcal{H}$ implies $\varphi(X) \in \mathcal{H}$ for $\varphi \in C_{l, \text{lip}}(\mathbb{R})$ where $C_{l, \text{lip}}(\mathbb{R})$ denotes the linear space of functions satisfying

$$\varphi(x) - \varphi(y) \leq C_\varphi(1 + |x|^m + |y|^m)$$

for each $x, y \in \mathbb{R}$ and some $C_\varphi > 0$ and $m \in \mathbb{N}$ depending on φ .

Definition 1.1. A sublinear expectation $\hat{\mathbb{E}}$ is a functional $\hat{\mathbb{E}}: \mathcal{H} \rightarrow \mathbb{R}$ satisfying the following properties

1. **Monotonicity:** For all $X, Y \in \mathcal{H}$ and $X \leq Y$, we have $\hat{\mathbb{E}}[X] \leq \hat{\mathbb{E}}[Y]$;
2. **Constant preserving:** For all $c \in \mathbb{R}$, $\hat{\mathbb{E}}[c] = c$;
3. **Sub-additivity:** For all $X, Y \in \mathcal{H}$, we have $\hat{\mathbb{E}}[X + Y] \leq \hat{\mathbb{E}}[X] + \hat{\mathbb{E}}[Y]$;
4. **Positive homogeneity:** For all $X \in \mathcal{H}$, we have $\hat{\mathbb{E}}[\lambda X] = \lambda \hat{\mathbb{E}}[X]$ for all $\lambda \geq 0$.

The triple $(\Omega, \mathcal{H}, \hat{\mathbb{E}})$ is called sublinear expectation space.

Definition 1.2. A random variable X is said to be independent from a random variable Y if $\hat{\mathbb{E}}[\varphi(X, Y)] = \hat{\mathbb{E}}[\hat{\mathbb{E}}[\varphi(X, y)]y = Y]$, for $\varphi \in C_{l, \text{lip}}(\mathbb{R} \times \mathbb{R})$.

Definition 1.3. Two random variables X and Y are identical distributed if $\hat{\mathbb{E}}[\varphi(X)] = \hat{\mathbb{E}}[\varphi(Y)]$ for all $\varphi \in C_{l, \text{lip}}(\mathbb{R})$. In this case, we write $X \sim Y$.

Definition 1.4. A random variable X in a sublinear expectation $(\Omega, \mathcal{H}, \hat{\mathbb{E}})$ is called G -normal distributed if for $a, b \geq 0$ it holds

$$aX + b\bar{X} \sim \sqrt{a^2 + b^2}X,$$

where $a, b \geq 0$ and \bar{X} is an independent copy of X . The letter G denotes the function $G: \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$G(x) := \frac{1}{2}(\bar{\sigma}^2 x^+ - \underline{\sigma}^2 x^-), \quad 0 \leq \underline{\sigma}^2 \leq \bar{\sigma}^2. \quad (1)$$

We denote G -normal distributed X by $X \sim \mathcal{N}(0, [\underline{\sigma}^2, \bar{\sigma}^2])$ where $\bar{\sigma}^2 = \hat{\mathbb{E}}[X^2]$ and $\underline{\sigma}^2 = -\hat{\mathbb{E}}[-X^2]$.

Definition 1.5. Let $G: \mathbb{R} \rightarrow \mathbb{R}$ a given monotonic and sublinear function. A stochastic process $B = (B_t)_{t \geq 0}$ on $(\Omega, \mathcal{H}, \hat{\mathbb{E}})$ is called a G -Brownian motion if the following conditions hold:

1. $B_0 = 0$;
2. $B_t \in \mathcal{H}$ for all $t \geq 0$;

3. For each $t, s \geq 0$, $B_{t+s} - B_t$ is independent of $(B_{t_1}, \dots, B_{t_n})$ for each $n \in \mathbb{N}$ and $0 \leq t_1 < \dots < t_n \leq t$. Moreover, $(B_{t+s} - B_t)s^{-\frac{1}{2}}$ is G -normally distributed.

Briefly speaking, a G -Brownian motion is a continuous process with independent and stationary increments under a given sublinear expectation $\hat{\mathbb{E}}$.

Definition 1.6. Fix now a time horizon $T > 0$ and introduce the space $\Omega_T := C^0([0, T]; \mathbb{R})$ of all \mathbb{R} -valued continuous paths $\omega = (\omega_t)_{t \in [0, T]}$ with $\omega_0 = 0$. We equip this space with the uniform convergence on compact intervals topology and denote by $\mathcal{F} = \mathcal{B}(\Omega_T)$ the Borel σ -algebra. Moreover, the canonical process $B = (B_t)_{t \in [0, T]}$ on Ω_T is given by $B_t(\omega) := \omega_t$ for $\omega \in \Omega_T$ and $t \in [0, T]$. We define $\Omega_t := \{\omega_i : \omega \in \Omega_T\}$ and $\mathcal{F}_t := \mathcal{B}(\Omega_t)$ for $t \in [0, T]$.

Definition 1.7. Introduce $Lip(\Omega_t)$ as

$$Lip(\Omega_t) := \{\varphi(B_{t_1}, \dots, B_{t_n}) \mid n \in \mathbb{N}, t_1, \dots, t_n \in [0, t], \varphi \in C_{l, lip}(\mathbb{R}^n)\}$$

for $t \in [0, T]$. Then the space $L_G^p(\Omega_t)$ is the completion of $Lip(\Omega_t)$, $t \geq 0$, under the norm $\|\xi\| = (\hat{\mathbb{E}}[|\xi|^p])^{\frac{1}{p}}$ for $p \geq 1$. Next, we introduce the definition of the G -expectation and the G -conditional expectation.

Definition 1.8. A G -expectation $\hat{\mathbb{E}}$ is a sublinear expectation on $(\Omega, Lip(\Omega_T))$ defined as follows. For $X \in Lip(\Omega_T)$ of the form $X = \varphi(B_{t_1} - B_{t_0}, \dots, B_{t_n} - B_{t_{n-1}})$, $0 \leq t_0 < t_1 < \dots < t_n \leq T$, $\varphi \in C_{l, lip}(\mathbb{R}^n)$, we set

$$\hat{\mathbb{E}}[X] := \mathbb{E}[\varphi(\xi_1 \sqrt{t_1 - t_0}, \dots, \xi_n \sqrt{t_n - t_{n-1}})],$$

where ξ_1, \dots, ξ_n are 1-dimensional random variables on a sublinear expectation space $(\tilde{\Omega}, \tilde{\mathcal{H}}, \mathbb{E})$ such that for each $i = 1, \dots, n$, ξ_i is G -normally distributed and independent of $(\xi_1, \dots, \xi_{i-1})$.

The corresponding canonical process $B = (B_t)_{t \in [0, T]}$ is a G -Brownian motion on the sublinear expectation space $(\Omega_T, Lip(\Omega_T), \hat{\mathbb{E}})$.

Definition 1.9. Let $X \in Lip(\Omega_T)$ have the representation $X = \varphi(B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}})$, $\varphi \in C_{l, lip}(\mathbb{R}^n)$, $0 \leq t_1 < \dots < t_n \leq T$. Then the G -conditional expectation under \mathcal{F}_{t_j} is defined as

$$\begin{aligned} \hat{\mathbb{E}}_{t_j}[X] &:= \hat{\mathbb{E}}[X | \mathcal{F}_{t_j}] \\ &= \psi(B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_j} - B_{t_{j-1}}) \end{aligned}$$

for $j = 1, \dots, n-1$ where $\psi(x) := \hat{\mathbb{E}}[\varphi(x, B_{t_{j+1}} - B_{t_j}, \dots, B_{t_n} - B_{t_{n-1}})]$. The G -expectation and the G -conditional expectation can be extended to sublinear operators $\hat{\mathbb{E}}[\cdot] : L_G^p(\Omega) \rightarrow \mathbb{R}$ and $\hat{\mathbb{E}}_t[\cdot] : L_G^p(\Omega_T) \rightarrow L_G^p(\Omega_t)$, for $t \in [0, T]$. It is shown that the G -expectation is an upper expectation, i.e., there exists a weakly compact set of probability measures \mathbb{P} such that

$$\hat{\mathbb{E}}[X] = \sup_{P \in \mathbb{P}} E_P[X]$$

for each $X \in L_G^1(\Omega_T)$. As we see that $\hat{\mathbb{E}}$ provides a robust way to measure the risky positions under model uncertainty. Related to this set \mathbb{P} , we introduce the Choquet capacity defined by

$$c(A) := \sup_{P \in \mathbb{P}} P(A)$$

for $A \in \mathcal{B}(\Omega_T)$.

Definition 1.10. For $p \geq 1$, we denote by $M_G^{p,0}(0, T)$ the space of simple integrands. Specifically, for a given partition $\{t_0, \dots, t_N\}$ of $[0, T]$, $N \in \mathbb{N}$, we define an element $\eta \in M_G^{p,0}(0, T)$ by

$$\eta_t(\omega) = \sum_{j=0}^{N-1} \xi_j(\omega) 1_{[t_j, t_{j+1}]}(t)$$

where $\xi_i \in L_G^p(\Omega_{t_i})$, $i = 0, \dots, N-1$.

Definition 1.11. For $p \geq 1$, we let $M_G^p(0, T)$ be the completion of $M_G^{p,0}(0, T)$ under the norm

$$\left(\int_0^T \hat{\mathbb{E}}[|\eta_t|^p] dt \right)^{\frac{1}{p}}.$$

Note that $M_G^{p,0}(0, T) \subseteq M_G^p(0, T)$. Similar as in the classical Itô's case, the integral with respect to B is first defined for the simple integrands in $M_G^{2,0}$ by the mapping

$$I(\eta) = \int_0^T \eta_s dB_s : M_G^{2,0} \rightarrow L_G^2(\Omega_T).$$

Then this mapping can be continuously extended to $I : M_G^2 \rightarrow L_G^2(\Omega_T)$. As in the classical case the quadratic variation of the G -Brownian motion is defined as the process $\langle B \rangle = (\langle B \rangle_t)_{t \in [0, T]}$ with

$$\langle B \rangle_t = B_t^2 - 2 \int_0^t B_s dB_s$$

for $t \in [0, T]$. A very interesting new phenomenon of the G -Brownian motion is that its quadratic variation process $\langle B \rangle$ is also a continuous process with independent and stationary increments, and thus can be still regarded as a Brownian motion.

The integral with respect to the quadratic variation

$$\int \eta_s d\langle B \rangle_s,$$

is introduced first for $\eta \in M_G^{1,0}(0, T)$ and then extended to $M_G^1(0, T)$.

2 Main Result

In this section, we prove a generalized Feynman–Kac formula in the G -setting, in presence of a linear term in the associated PDE. For this aim, we consider the G -Itô process $X_t^{r,x} = (X_t^{r,x})_{t \in [0, T]}$ given by

$$X_t^{r,x} = x + \int_r^t f(s, X_s^{r,x}) ds + \int_r^t g(s, X_s^{r,x}) d\langle B \rangle_s + \int_0^t h(s, X_s^{r,x}) dB(s), \quad r \leq t \leq T, \quad (2)$$

where $f, g, h : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ are deterministic functions such that $f(t, \cdot), g(t, \cdot), h(t, \cdot)$ are Lipschitz-continuous functions for every $t \in [0, T]$ and $f(\cdot, x), g(\cdot, x), h(\cdot, x)$ are continuous in t for every $x \in \mathbb{R}$. We aim to examine the G -conditional expectation follows as

$$\hat{\mathbb{E}}[\varphi(T, X_T) e^{-\int_r^T X_s ds} | X_r = x] \quad (3)$$

for any $(r, x) \in [0, T] \times (0, \infty)$ where φ is a bounded real-valued continuous function. Note that φ represents the payoff of an interest rate derivative or of a life insurance liability in applications.

An analogous of a Feynman–Kac formula within the framework of G -expectation is given in [Hu and et al. \(2014\)](#). This result connects the solution of a G -BSDE to the unique viscosity solution of a fully nonlinear PDE, whose nonlinearity is given by a term that depends on the function G representing the uncertainty about the volatility. The authors are able to deal with viscosity (i.e., non necessarily classical) solutions to the associated PDEs assuming the Lipschitz continuity of the coefficients of the G -BSDE. However, due to the presence of the discounting term $e^{-\int_r^T X_s ds}$ in (3), these conditions are not satisfied in our framework, so that we cannot apply this formula directly. Moreover, [Hu and et al. \(2014\)](#) start from a G -conditional expectation and prove that this solves a fully nonlinear PDE.

In the sequel, we aim to extend the approach in [Filipovic \(2009\)](#), which starts from PDE, to the G -setting. In particular, we assume that X is quasi-surely strictly positive and solves the given G -SDE (2) and we establish a relation between a nonlinear PDE and G -conditional expectation of a discounted payoff.

2.1 Feynman-Kac formula for a bounded payoff

In this section, we provide a Feynman–Kac formula under volatility uncertainty to include an additional linear term in the associated PDE due to discounting. For this aim, we present a theorem related to the bounded payoffs.

Assumption 2.1. *The function φ belongs to $C([0, T] \times \mathbb{R}^+)$ and is bounded by a constant $M_0 > 0$.*

Theorem 2.2. *Let $\varphi \in C([0, T] \times \mathbb{R}^+)$ satisfy Assumption 2.1. Given $(r, x) \in [0, T] \times (0, \infty)$, assume that the process $X^{r,x} = (X_t^{r,x})_{t \in [r, T]}$ in (2) is quasi-surely strictly positive. Define the function $u : [0, T] \times (0, \infty) \rightarrow \mathbb{R}$ by*

$$u(r, x) := \hat{\mathbb{E}}_r[\varphi(T, X_T^{r,x})e^{-\int_r^T X_s^{r,x} ds}].$$

Then u is a viscosity solution of the PDE

$$\begin{aligned} u_t + 2G(u_x g(t, x) + \frac{1}{2}u_{xx}h^2(t, x)) + f(t, x)u_x - xu &= 0, \quad (t, x) \in Q, \\ u &= \varphi, \quad (t, x) \in \partial'Q \end{aligned} \quad (4)$$

where G is defined by (1), $Q = (0, T) \times (0, \infty)$ and $\partial'Q = \{T\} \times \mathbb{R}^+ \cup (0, T) \times \{0\}$.

2.1.1 The sketch of the proof

In order to perform the analysis illustrated above, we have to use both stochastic calculus in the G -setting and the theory of nonlinear PDEs. In general the PDE (4) does not admit a classical $C^{1,2}$ solution. That is because, we substitute the coefficients of the original PDE with some bounded and $C^{1,2}$ cut-off functions depending on a parameter $\varepsilon \in (0, 1)$. In this way, we are able to show that the approximated PDE admits a viscosity solution u_ε defined on an unbounded domain, which is additionally $C^{1,2}$ inside a bounded domain D_ε depending on ε by following Krylov (1987).

For this aim, for fixed $\varepsilon \in (0, 1)$, we define the intervals $D_\varepsilon := (\varepsilon, \varepsilon^{-1})$ and the random time

$$\tau_\varepsilon^{r,x} := \inf\{s \in [r, T] : X_s^{r,x} \in [\varepsilon, \varepsilon^{-1}]\} \wedge T,$$

where we use the convention $\inf \emptyset := +\infty$. Then we can apply the G -Itô's formula to prove that

$$u_\varepsilon(r, x) = \hat{\mathbb{E}}_r[u_\varepsilon(T \wedge \tau_\varepsilon^{r,x}, X_{T \wedge \tau_\varepsilon^{r,x}}^{r,x})e^{-\int_r^{T \wedge \tau_\varepsilon^{r,x}} (X_s^{r,x} + \varepsilon) ds}],$$

is the viscosity solution of the regularized PDE corresponding to the PDE (4) based on this fact that $u_\varepsilon \in C^{1,2}$. The reader can find more details in Akhtari and et al. (2023). Finally, we prove that

$$\lim_{\varepsilon \rightarrow 0} |u_\varepsilon(r, x) - u(r, x)| = 0$$

for any $(r, x) \in [0, T] \times (0, \infty)$ which finishes the proof.

Discussion and Results

We proved that the G -conditional expectation of a discounted payoff is a limit of the viscosity solutions of a family of regularized PDEs which is also $C^{1,2}$ on a bounded domain depending on the regularizing parameter with terminal condition φ at time $T > 0$ when the regularity parameter goes to zero. Indeed, such a limit is itself a viscosity solution of the limit PDE.

References

- Akhtari, B., Biagini, F., Andrea Mazzon, A., Oberpriller, K. (2023), Generalized Feynman–Kac formula under volatility uncertainty, *Stochastic Processes and their Applications*, **166**, 104083.
- Filipovic. D. (2009), *Term-Structure Models: A Graduate Course*, **31**, (Springer Finance), Springer.
- Hölzermann, H. (2021), The Hull-White Model under volatility uncertainty, *Quantitative Finance*, **21(11)**, 1921–1933.
- Hölzermann, H. (2022), Term structure modeling under volatility uncertainty, *Mathematics Financial Economics*, **16**, 317–343.
- Hu, M., Ji, S., Peng, S., Song, Y. (2014), Comparison theorem, Feynman-Kac formula and Girsanov transformation for BSDEs driven by G-Brownian motion, *Stochastic Processes and their Applications*, **124(2)**, 1170–1195.
- Knight. F. H. (1921), *Risk, uncertainty and profit*, **31**, Houghton Mifflin.
- Krylov. N. (1987), *Nonlinear elliptic and parabolic equations of the second order*, Kluwer, 1987.
- Peng S. (1972), *Nonlinear Expectations and Stochastic Calculus Under Uncertainty: With Robust CLT and G-Brownian Motion*, Probability Theory and Stochastic Modelling 95. Springer.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes

August 30-31, 2025
University of Kurdistan



Review on complete convergence for m-NOD random sequences

Mohammad Amini ^{*1}, Habib Naderi ²

¹Ferdowsi University of Mashhad

²University of Sistan And Baluchestan

Abstract

In this paper, we review the complete convergence for m-NOD sequence. Also, we extend Jajte's technique, to study the rate of complete convergence for weighted sequence of m-NOD random sequence. Moreover, we make a simulation study to illustrate the asymptotic behavior in the sense of the rate of complete convergence.

*m-amini@um.ac.ir



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



A Review of Quantum-Based Distance Metrics and Their Applications in Modern Finance

Hamid Arian *

York University

Abstract

Classical financial models often rely on distance metrics, like the Pearson correlation, that struggle to capture the complex, non-linear dynamics and high-dimensional nature of modern markets. This paper reviews an emerging alternative: quantum-based distance metrics (see Arian et al. (2025)). We survey the theoretical foundations of this approach and inspect key quantum metrics like the Trace Distance, Bures Distance, and Quantum Jensen-Shannon Divergence (QJSD). We then explore the practical applications of these tools in finance, including enhanced portfolio diversification through quantum-inspired clustering, market regime detection for risk management, and the development of quantum kernels for machine learning-based trading strategies. The paper discusses the potential advantages of this quantum formalism, such as its inherent ability to capture non-linearity and its robustness in high-dimensional settings, while also addressing the significant current challenges, including the data encoding problem and the need for greater interpret-ability.

Keywords: Quantum Finance Distance Metrics Portfolio Optimization Risk Management Quantum Machine Learning.

*arian@risklab.ai



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



**Seminar On Probability
and Stochastic Processes**
August 30-31, 2025
University of Kurdistan



A Novel Bayesian Two-Stage Approach for Dynamic Prediction Using Joint Modeling of Multiple Longitudinal Markers and Time-to-Event Data

Taban Baghfalaki^{*1}

¹Department of Mathematics, The University of Manchester, Manchester, UK

Abstract

In many clinical and epidemiological studies, researchers collect multiple longitudinal measurements alongside time-to-event outcomes. Understanding the association between these two types of data is often key, but analyzing them together using joint models becomes increasingly challenging as the number of longitudinal markers grows. Existing R packages can handle joint modeling, but they struggle with high-dimensional data due to long computation times and convergence issues. To address this, we introduce a novel Bayesian two-stage approach for joint modeling with multiple longitudinal markers. Like the standard two-stage method, we begin by modeling the longitudinal and survival components separately. However, instead of fitting mixed models in the first stage, we fit one-marker joint models for each longitudinal marker and the event time. This allows us to derive subject-specific predictions, such as current values and slopes, that are adjusted for informative dropout. In the second stage, we fit a proportional hazards model that includes these predicted values and/or slopes as time-dependent covariates. To properly account for uncertainty from the first stage, we use multiple imputation when estimating the Cox model. This two-stage approach enables scalable and flexible dynamic prediction, even with many markers-something traditional joint modeling can't easily handle. We assess the method's performance through simulation studies and apply it to two real-world datasets: the publicly available PBC2 data and a dementia study involving seventeen longitudinal markers. Our results show that the approach is both practical and effective for high-dimensional longitudinal survival data.

Keywords: Bayesian paradigm; Joint modeling; Multi-marker Longitudinal data; Regression calibration; Risk prediction.

Mathematics Subject Classification (2020): 62N01, 62J12, 62F15.

^{*}Corresponding author, taban.baghfalaki@manchester.ac.uk



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Allocated Optimal Strategies to the Wealth Process of a Diffusion Risk Model

Abouzar Bazyari*

Department of Statistics, Faculty of Intelligent Systems Engineering and Data Science,
Persian Gulf University, Bushehr, Iran

Abstract

In this paper, we obtain a closed-form solution for a diffusion risk model which optimally allocates his/her wealth among the following coefficients: a risky asset, a free bank account, and a stock. The following relationships between the optimal amount invested in the security and the risk premium are obtained: (i) for a risk premium greater than one, the investor will optimally invest a positive amount in the bond, and (ii) for a risk premium equal to one, the insurance company will optimally invest nothing in the bond. Although the inclusion of a credit-related financial product in the portfolio selection if risk models is more realistic, no closed-form solutions to date are given in the literature when a recovery value is considered in the event of a default. We solve the optimal portfolio problem of insurance company for the representative investor with a specified utility function. Moreover, the implications of the analytic result for asset allocation are discussed and carried out sensitivity analysis by adopting benchmark parameter values.

Keywords: Itô's formula, Optimal Strategy, Portfolio optimization, Defaultable security, Credit risk, Recovery of market value.

Mathematics Subject Classification (2020): 62G32, 62F99, 62E20.

1 Introduction

Risk theory came into being and was developed in order to provide a basis for the existence and significance of the insurance system. One of the most fundamental problems in risk theory is the ruin problem. It analyzes the behavior of a stochastic process, that represents the evolution of the capital of an insurance company. As a result, optimal portfolio problems with defaultable securities have become an important area of research. [Hou and Jin \(2002\)](#) and [Hou \(2003\)](#) address the optimal portfolio problem of the investor by giving the investor the ability to allocate her wealth among a stock, a default-free bank account, and a credit-risky financial instrument.

[Lacker and Zariphopoulou \(2019\)](#) obtained the optimal investment strategies for n agents with

*Corresponding author, ab_bazyari@pgu.ac.ir

relative performance concerns under Coefficient of Constant Absolute Risk-Aversion (CARA) and Coefficient of Constant Relative Risk-Aversion (CRRA) utilities. Aydoğan and Steffensen (2024) determined the optimal investment strategies for the agent under the specific situations arising from the pure-diffusion and pure-jump market cases. Bäuerle and Göll (2024) have recently extended their studies by analyzing the effect of the price impact on equilibrium strategies. Deng et al. (2024) introduced another extension of Lacker and Zariphopoulou (2019) in which they investigate a mean-field game strategy in a partially observable market. Some other studies derived the optimal strategies assuming stochastic volatility, for instance, Zhu et al. (2019), Kraft (2020) and Zariphopoulou (2024).

In the present paper, we derive the optimal amount of wealth allocated to a defaultable security as well as to a stock and to a default-free money market account. The following relationships between the optimal amount invested in the defaultable security and the default-risk premium are obtained: (i) for a default-risk premium greater than one, the investor will optimally invest a positive amount in the defaultable bond, and (ii) for a risk premium equal to one, the investor will optimally invest nothing in the defaultable bond. A risk premium greater than one can be interpreted as the market fairly pricing the default risk in the defaultable bond. On the other hand, a default-risk premium equal to one, can be interpreted as the market not pricing the default risk in the defaultable bond.

The rest of the paper is organized as follows: Section 2 presents the dynamics of a defaultable bond, a stock as well as a savings account under two equivalent probability measures Q and P which will be used for the portfolio optimization in Section 3. Implications of the analytic result for asset allocation are discussed in Section 4. Section 5 carries out sensitivity analysis by adopting benchmark parameter values from the literature. Section 6 summarizes the results and concludes the paper.

2 Dynamics of Financial Securities

In the sequel to obtain our goal, we will work on the probability space (Ω, \mathcal{F}, P) , which is endowed with the information filtration $\mathcal{F} = \{\mathcal{F}_t\}_{t \geq 0}$ which carries all stochastic quantities and right continuity and it is often called an enlarged filtration given by $\mathcal{F}_t = \mathcal{G}_t \vee \mathcal{H}_t$. In the probability space, P denotes the martingale probability measure. The filtration \mathcal{G}_t is assumed to be generated by the Winner process which shows the change of stock price and the filtration \mathcal{H}_t shows a Poisson process that is used to denote the arrival of risks. In the probability space Q denotes the martingale probability measure (or risk neutral measure), which is assumed to be equivalent to the some real-world probability measure which we denote by P .

Now we are ready to define a default process.

Definition 2.1. *A nondecreasing right continuous process which makes discrete jumps at a random time τ is called default process. We denote a default process by $H_t = 1_{\tau \leq t}$, where 1 represents the indicator function which takes the value of one if there is a jump and zero otherwise and $H_u^- = \lim_{s \rightarrow u} H_s$.*

The default process H is assumed to be a Poisson process with a constant intensity h .

Definition 2.2. *The martingale default process is thus given by the following equation:*

$$M_t = H_t - \int_0^t (1 - H_{u-}) h du.$$

The stochastic differential equation (SDE) of compensated Poisson process defined above is $dM_t = dH_t - h(1 - H_{t-})dt$. We extensively utilize this stochastic differential equation (SDE) to derive the dynamics of the defaultable bond.

2.1 Dynamics of bond price process and stock price under Q

We assume that there exists a defaultable zero coupon bond with a maturity date T_1 . The value of the bond after default is assumed to be zero. In the case of default the investor recovers a fraction of the market value of the defaultable bond just prior to default. The loss rate is denoted by ζ which is assumed to be constant and take value between zero and one. The arrival rate of the default h^Q of the Poisson process is also assumed to be constant. Table 1 summarizes the parameter's values.

Table 1: Parameter Definition and values	
<i>Symbol</i>	<i>Denition</i>
h^Q	risk neutral default intensity
$\frac{h^Q}{h^P}$	default risk premium
γ	risk aversion parameter
ζ	write-down rate
μ	instantaneous rate of return of stock
σ	instantaneous volatility of stock's return
r	risk free short rate of interest

We will now define the price process for a bond under the assumption of constant interest rate and intensity under the risk neutral measure Q as follows:

$$p(t, T_1) = 1_{\tau < t} e^{-(r+\delta)(T_1-t)} + 1_{\tau \leq t} (1 - \zeta) e^{-(r+\delta)(T_1-t)} e^{(t-\tau)}. \quad (1)$$

In the equation (1), $p(t, T_1)$ being a fictitious security because it is not really a traded security. It actually consists of two components, which are mutually exclusive, i.e., the component of the left hand side is the actual defaultable bond price given that default has yet to occur, whereas the component of the right hand side represents the value (or recovered amount) of an already defaulted bond. By accounting for the second component we are implicitly modeling the difference between the bond price if there is a full recovery and fractional recovery at default. The explicit inclusion of the recovery is the first innovation of this paper and the measure probability Q dynamics of the price process of the defaultable bond is derived in the following lemma.

Lemma 2.3. *The Q dynamics for the price process of the defaultable bond at time t is represented in terms of the compensated jump process M^Q , which is Q martingale process, such as:*

$$dp(t, T_1) = rp(t, T_1)dt - \zeta e^{-(r+\delta)(T_1-t)} dM^Q. \quad (2)$$

Proof. Applying the Itô's formula on equation (1). □

Now we consider that the agent can invest money in the bank at the constant short rate of interest r , i.e. he/she has access to the risk-free asset B_t with

$$dB_t = rB_t dt. \quad (3)$$

The investor can also invest in a risky stock with price process S_t , the dynamics under Q is given by the following diffusion equation:

$$dS_t = S_t(rdt + \sigma dW_t^Q), \quad (4)$$

where σ instantaneous volatility of the equity return process is assumed to be constant. Note that the expected return under Q for each security is the risk free interest rate.

We state the theorem without proof.

Theorem 2.4. For positive interest rate r , the dynamics of bond price is represented as the exponential form of SDE as follows:

$$\begin{aligned} dp(t, T_1) &= (1 - H_t)(r + \delta - \delta\Delta)p(t, T_1)dt + H_t r p(t, T_1)dt \\ &\quad - (1 - H_{t-})\zeta p(t, T_1)dM_t^Q \\ &= p(t^-, T_1)(r dt + (1 - H_t)\delta(1 - \Delta)dt - (1 - H_{t-})\zeta dM_t^Q), \end{aligned} \quad (5)$$

where

$$p(t, T_1) = \begin{cases} e^{-(r+\delta)(T_1-t)}, & \text{if } \tau > t, \\ (1 - \zeta)e^{-(r+\delta)(T_1-t)}e^{(t-\tau)}, & \text{if } \tau \leq t. \end{cases}$$

The instantaneous expected return of the defaultable bond, given in equation (5), can be seen to consist of two components similar to what Yu (2002) argues (see equation (15) in Yu (2002)). The first component is the return on an otherwise identical default-free bond. The second is the difference between the risk neutral credit spread and the real world credit spread provided that default has not already occurred by time t . The implication of Jarrow's et al. (2005) argument is that the second component will disappear. In order to see this clearly we explicitly express the difference of the credit spreads in terms of the intensities under Q and P such as:

$$\delta - \delta\Delta = \zeta h^Q - \zeta \Delta h^Q = \zeta h^Q - \zeta h^P = \left(\frac{h^Q}{h^P} - 1\right)\zeta h^P,$$

where $\delta^P = \zeta h^P$ is the credit spread under the real world probability measure. Therefore, if h^Q is equal to h^P then the second term in equation (5) vanishes provided that default has not already occurred by time t . In such a case, the investor requires no compensation for bearing the risk of the default event. Through simple parameterization of the default event risk we can show the difference of the credit spreads in terms of the default event risk premium, namely,

$$\delta - \delta\Delta = \left(\frac{h^Q}{h^P} - 1\right)\zeta h^P = \left(\frac{1}{\Delta} - 1\right)\delta^P,$$

where $\frac{1}{\Delta} = \frac{h^Q}{h^P}$. Thus the compensation for the default event risk premium becomes $(\frac{1}{\Delta} - 1)$. We can also express the dynamics of stock price under P as follows:

$$dS_t = S_t(\mu dt + \sigma dW^P). \quad (6)$$

The stock price grows at a instantaneous rate of μ and has the instantaneous volatility of σ under the real world probability measure P . Note that we change the drift term of the dynamics of stock price, but do not change the volatility of it through the change of risk neutral probability measure to real world probability measure. On the other hand, the dynamics of a savings account has the same representation under both measure Q and P since it is non-stochastic.

The pricing dynamics of each security under the real world probability measure P such as equation (3), (5), and (6) are utilized in the next section.

3 Optimal portfolio problems using the utility function

We use stochastic optimal control theory to find the optimal allocations among a savings account, a stock (or stock index), and a defaultable bond for the investor whose objective is to maximize her conditional expected utility of the terminal wealth over the investment horizon $[0, T]$. We assume throughout that $T < T_1$, where T_1 is the time of maturity of the defaultable bond. We consider that short rate of interest is constant and non-negative.

To start, let N_t^s be the number of shares of stocks the investor longs $N_t^s \geq 0$ or shorts $N_t^s < 0$ at time t . Similarly, N_t^P and N_t^B are the number of shares of the defaultable bond and a savings

account she buys or sells at time t for all $t \in [0, T]$. Then the process $N = (N^s, N^p, N^B)$ is called a portfolio process. Now we define $K_t(N)$ as the wealth of the portfolio process $N = (N^s, N^p, N^B)$ at time t , that is

$$K_t(N) = N_t^s S_t + N_t^B B_t + N_t^p p(t, T_1), \quad t \in [0, T].$$

The wealth of the portfolio process N starting with $\nu > 0$ at time t will be denoted as $K_t^\nu(N)$. Now, let $(\pi_t^s, \pi_t^p, \pi_t^B)$ be defined as

$$\pi_t^s = \frac{N_t^s S_t}{K_{t-}(N)}, \quad \pi_t^p = \frac{N_t^p P(t^-, T_1)}{K_{t-}(N)}, \quad \pi_t^B = \frac{N_t^B B_t}{K_{t-}(N)}.$$

Note that $\pi_t = (\pi_t^s, \pi_t^p, \pi_t^B)$ represents the fraction of the wealth invested in each asset at time t and is called fractional strategy or simply strategy in the sequel. For all $t \in [0, T]$ we shall consider the following wealth equation:

$$\begin{aligned} dK_t(\pi) &= K_{t-}(\pi) \left((\pi_t^s \mu + (\delta(1 - H_t)(1 - \Delta)\pi_t^p + r) + r(1 - \pi_t^s - \pi_t^p)) dt \right. \\ &\quad \left. + \pi_t^s \sigma dW_t^P - \pi_t^p \zeta dM_t^P \right), \end{aligned} \quad (7)$$

with $K_0(\pi) = \nu$, where $dM_r^P = (1 - H_t)dM_r$, and $h^P = h^P \Delta$.

3.1 Optimization Problems

We consider the following joint dynamics of the wealth equation $K_t(\pi)$ given in (7) and for the factor equation H_t for $t \in [0, T]$,

$$\begin{aligned} dK_t(\pi) &= K_{t-}(\pi) \left((\pi_t^s \mu + (\delta(1 - H_t)(1 - \Delta)\pi_t^p + r) + r(1 - \pi_t^s - \pi_t^p)) dt \right. \\ &\quad \left. + \pi_t^s \sigma dW_t^P - \pi_t^p \zeta dM_t^P \right), \end{aligned} \quad (8)$$

where $dH_t = h^P(1 - H_t)dt + dM_t^P$, with $K_0(\pi) = \nu$, $H_0 = 0$. Moreover, we use $h^P = h^P \Delta$ and $\delta = h^P \zeta$.

To present the optimization problem, we consider the utility function as

$$U(x) = \frac{x^\gamma}{\gamma}, \quad 0 < \gamma < 1, \quad x \geq 0. \quad (9)$$

We shall study the optimization problems in case of the given utility function (9) such as:

$$O(t, \nu, z) = \sup_{(\pi^s, \pi^p) \in \Pi_t(G)} E_P \left(\frac{K_T^\gamma(\pi)}{\gamma} \mid K_t(\pi) = \nu, H_t = z \right), \quad (10)$$

for all $(t, \nu, z) \in (0, T) \times (0, \infty) \times \{0, 1\}$ over all admissible strategies π . Let us define the following processes as

$$\pi_t^{s*} = \frac{(\mu - r)}{\sigma^2(1 - \gamma)}, \quad t \in [0, T],$$

$$\pi_t^{p*} = \begin{cases} \frac{1}{\zeta} \left(1 - \left(\frac{1}{\Delta} \right)^{\frac{1}{1-\gamma}} e^{\left(-\frac{h^P}{1-\gamma} - \frac{h^P}{\Delta} \frac{\gamma}{1-\gamma} \right)(T-t)} + e^{\left(-\frac{h^P}{1-\gamma} - \frac{h^P}{\Delta} \frac{\gamma}{1-\gamma} \right)(T-t)} \right), & \text{if } 0 \leq t < \tau \wedge T, \\ 0, & \text{if } t \in ((\tau \wedge T), T), \end{cases}$$

and

$$\pi_t^{B*} = \begin{cases} 1 - \pi_t^{s*} - \pi_t^{p*}, & \text{if } 0 \leq t < \tau \wedge T, \\ 1 - \pi_t^{s*}, & \text{if } t \in ((\tau \wedge T), T). \end{cases}$$

Then the process π defined above is an optimal strategy. Now we are ready to state the optimality theorem.

Theorem 3.1. (*Optimal strategies of risk model*). *The process π^* defined above is an optimal strategy.*

Note that the optimal trading strategies are independent of the wealth since the utility function is a homogeneous function of wealth. Next we derive the optimal value function $O(.,.,.)$. This is often called the indirect utility function since it measures the highest attainable expected utility the investor can derive from her current wealth in the current state of the world.

Theorem 3.2. (*Indirect utility function*). *Given the utility function (9), we obtain the optimal objective function $O(t, \nu, z)$ as*

$$O(t, \nu, z) = \begin{cases} (f(t))^{1-\gamma} \times \left(\frac{\nu^\gamma}{\gamma}\right), & \text{if } z = 0, \\ e^{\left(\frac{1}{2} \frac{\gamma}{1-\gamma} \frac{(\mu-r)^2}{\sigma^2} + r\gamma\right)(T-t)} \times \left(\frac{\nu^\gamma}{\gamma}\right), & \text{if } z = 1, \end{cases}$$

where

$$\begin{aligned} f(t) &= e^{\left(\frac{1}{2} \frac{\gamma}{(1-\gamma)^2} \frac{(\mu-r)^2}{\sigma^2} + \frac{r\gamma}{1-\gamma}\right)(T-t)} \\ &\times \left(e^{\left(-\frac{h^P}{1-\gamma} - \frac{h^P}{\Delta} \frac{\gamma}{1-\gamma}\right)(T-t)} + \left(e^{\left(-\frac{h^P}{1-\gamma} - \frac{h^P}{\Delta} \frac{\gamma}{1-\gamma}\right)(T-t)} - e^{-\frac{r\gamma^2}{1-\gamma}(T-t)}\right) \right. \\ &\times h^P \left(\frac{1}{\Delta}\right)^{\frac{1}{1-\gamma}} \frac{1-\gamma}{h^P(\gamma-\Delta) + r\Delta\gamma^2}. \end{aligned}$$

Proof. In the following, we prove that $O(t) \geq O(t, \nu, z)$ for all $\nu \in S_t$, and $O_1(t) = O_1(u, \nu^*, z^*)$. Using the quasi variational inequality, for any $q > 0$ we have

$$-\eta O(R^{(\nu, z^*)}(t)) + K^{(e_1, e_2^*)} O(R^{(\nu, z^*)}(t)) \leq -q(R^{(\nu, z^*)}(t)).$$

On the other hand, since $-\alpha_2 \leq O'(t) \leq \alpha_1$, therefore for $z \in \Delta_1$ and $z \leq t$, the inequality

$$O(R^{(\nu, z^*)}(t)) - O(R^{(\nu, z^*)}(t-)) \leq \alpha_1 (A_1^{s_1}(z) - A_1^\nu(t-)),$$

holds, and for $z \in \Delta_2$, we have

$$O(R^{(\nu, z^*)}(z)) - O(R^{(\nu, z^*)}(z-)) = \alpha_2 (A_2^{s_2^*}(t) - A_2^{z^*}(t-)).$$

On the other hand, we obtain

$$\begin{aligned} O(t) &\geq E \left[\int_0^t e^{-\eta t} q_1(R^{(\nu, z^*)}) dt - \alpha_1 \int_0^t e^{-\eta t} dA_1^\nu(t) \right. \\ &\quad \left. + \alpha_2 \int_0^t e^{-\eta t} dA_2^{z^*}(t) \right. \\ &\quad \left. + e^{-\eta z} O(R^{(\nu, z^*)}(t)) \right] = O(t, \nu, z^*). \end{aligned}$$

By the same procedure, we will obtain that $O(t) \geq O(t, \nu, z^*)$ and this completes the proof. \square

This Theorem shows that $O_\nu = \frac{\partial}{\partial \nu} O(t, \nu, z) > 0$ and $O_{\nu\nu} = \frac{\partial^2}{\partial \nu^2} O(t, \nu, z) < 0$. This indicates that $O(t, \nu, z)$ is an concave function of initial wealth ν .

4 Implication for asset allocation

We investigate the implications of the optimal strategies.

4.1 Optimal strategy for a stock

The optimal strategy for the stock is constant irrespective of pre-default or post-default. The reason is that the stock has no correlation with the defaultable bond, which means there is no need to hedge for the default. Thus the optimal strategy for the stock is invariant to the default event risk.

Let us decompose the optimal strategy π^{s*} for the stock as follows:

$$\pi^{s*} = -\frac{O_\nu}{\nu O_{\nu\nu}} \frac{\mu - r}{\sigma^2} = -\frac{1}{\gamma - 1} \frac{\mu - r}{\sigma^2}, \quad (11)$$

The optimal strategy π^{s*} for the stock is interpreted as the product of the relative risk tolerance $-\frac{O_\nu}{\nu O_{\nu\nu}}$ (i.e. the inverse of the relative risk aversion) of the indirect utility function and market price of equity risk (or Sharpe ratio of a stock) $\frac{\mu - r}{\sigma}$ normalized by the volatility of the stock return σ . Thus the demand for equity contains only so-called “myopic” term despite the fact that the investment opportunity set is stochastic in this economy. As usual, the myopic demand for the stock decreases as the risk aversion coefficient γ is close to zero. This is reasonable since close to zero indicates that the relative risk aversion $1 - \gamma$ of the investor whose utility function is close to one.

4.2 Optimal strategy for a bond

It is clear from Lemma 2.4 that the drift and volatility of the return of the defaultable bond changes discretely at default. That is, the stochastically evolving default process H , which is the only risk source of randomness for the defaultable bond, is captured in the drift and volatility of the return of the defaultable bond. We can interpret the change that occurs at default as a change in the investment opportunity set of the investor. In the default literature it is commonly assumed that the investor knows whether the default event has occurred or not. Thus we can simply analyze the optimal defaultable bond investment on the pre and post default space. From the analytical result it is clear that the pre-default optimal strategy of the defaultable bond depends on the investment planning horizon. The investor allocates more when she has longer planning horizon.

It can be easily shown that the probability of default in our framework is greater for longer planning horizon. Nevertheless, we find that the investor allocates greater amount of wealth the longer the amount of time remaining. Although this result at first appears to be perverse it can be explained in terms of the default risk premium. Recall from Section 2 that if default follows a Poisson process then the pre-default price of the defaultable bond is given $p(t, T_1) = e^{(r+\delta)(T_1-t)}$, where $T < T_1$. We can see that as time t approaches to T , both the investment horizon $T - t$ and the time to maturity of the defaultable bond $T_1 - t$ tend to zero. This leads to an increase in the price of the defaultable bond providing no default has yet to occur. In other words, the defaultable bond is getting relatively cheap as the investment horizon increases. Since the default event risk premium $\frac{h^Q}{h^P}$ is assumed to be the same throughout the planning horizon and the time to maturity then the only risk the investor faces with regards to her return in the defaultable bond is the default itself. Thus the investor would naturally invest greater amount at longer planning horizon since for given probability of default the bond naturally appreciates in value as it achieves par value at maturity providing there is no default. Following the buy low sell high rule, the investor should invest more in the defaultable bond when its price is low, that is, for longer maturity date, all else equal.

5 Sensitivity analysis on the model parameters

In this section we analyze the behavior of the optimal strategy of the defaultable bond in particular. Since the optimal strategy of the defaultable security is a function of the default risk premium, the loss rate, the time to maturity, the risk aversion, the default intensity, and the interest rate. We quantify the optimal investment for the defaultable bond by adopting parameter values from the literature. Hou (2003) and Driessen (2005) also provide parameter values. Since a higher default risk premium induces a positive amount of investment in the defaultable security, we are not surprised to find a positive relationship between optimal investment in the defaultable security and the default risk premium. Figure 1 shows that the investor buys more defaultable security as the default risk premium increases. Note that the optimal investment in the defaultable bond is increasing with a decreasing rate when the default risk premium increases.

In Figure 1, we also observe that if the loss rate is small then the investor responds more to

the increase in the default risk premium. As the loss rate increases, however, the investor still enjoys the high risk premium, but with much care. The concave shape of optimal investment of the defaultable security is getting flatter as ζ increases. When the loss rate (recovery rate) is very low (high), the marginal increase of the optimal strategy of the defaultable bond is huge. Thus the investor enjoys the priced default event risk premium since she can recover a lot even in the case of default. Note that if the default risk premium is equal to one the investor does not invest in the defaultable security, which is shown as the horizontal line.

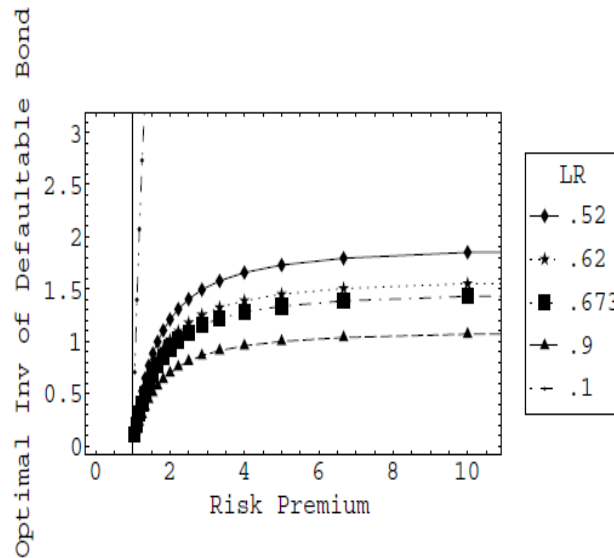


Figure 1: Optimal investment of a bond versus the risk premium is 1, 0.52, 0.62, 0.673 and 0.9 given $\gamma = 0.5$.

Next we analyze the relationship between the loss rate and the optimal strategy. Since the higher loss rates mean less recovery values the optimal investment in the defaultable security shows a decaying pattern. Figure 2 indicates that when the default risk premium is high the investor decreases the investment in the defaultable security, but still invests more compared to the case of low default risk premium as the loss rate increases.

Discussion and Results

In this paper we address the optimal portfolio problem with a defaultable security. We adopt the intensity based approach to modeling defaultable securities. In addition, we assume the recovery of market value throughout the paper. This is a key distinguishing feature, as most papers ignore the problem of a positive recovery value in the event of a default or they just simply assume zero recovery. Thus this paper improves upon the existing literature by explicitly modeling the recovery amount in the dynamics of the defaultable bond price. The theory of stochastic optimal control is utilized as opposed to the martingale approach to solve the investor's problem. We obtained closed form solutions for the optimal strategies for a default-free bank account, a stock, and a defaultable bond.

Acknowledgment

The author would like to thank the reviewers for providing insightful comments that improved the presentation of this paper.

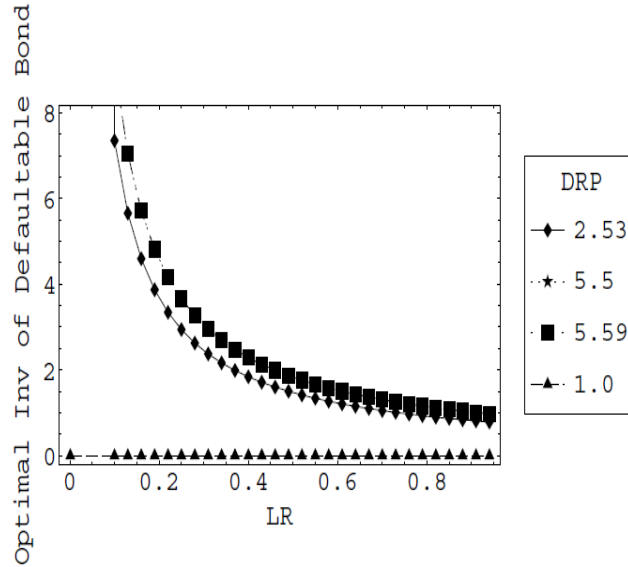


Figure 2: Optimal investment of a defaultable bond versus the loss rate when risk premium is 1, 2.53, 5.5, 5.59 given $\gamma = 0.5$.

References

- Aydoğan, B. and Steffensen, M. (2024), *Optimal investment strategies under the relative performance in jump-diffusion markets*, Decisions in Economics and Finance, <https://doi.org/10.1007/s10203-024-00499-1>.
- Bäuerle, N. and Göll, T. (2024), *Nash equilibria for relative investors with (non)linear price impact*, Mathematics and Financial Economics, 18, 27-48.
- Deng, C., Su, X. and Zhou, C. (2024), *Relative wealth concerns with partial information and heterogeneous priors*, SIAM Journal on Financial Mathematics, 15(2), 360-398.
- Driessen, J. (2005), *Is Default Event Risk Priced in Corporate Bonds?*, Review of Financial Studies, 18, 165-195.
- Hou, Y. and Jin, X. (2002), *Optimal Investment with Default Risk*, FAME research pape 46, Switzerland.
- Hou, Y. (2003), *Integrating Market Risk and Credit Risk: A Dynamic Asset allocation perspective*, Working Paper, Switzerland. Yale University.
- Kraft, H. (2020), Meyer-Wehmann, A., Seifried, F.T.: *Dynamic asset allocation with relative wealth concerns in incomplete markets*, Journal of Economic Dynamics and Control, 113, 103857.
- Lacker, D., Zariphopoulou, T. (2019), *Mean-field and n-agent games for optimal investment under relative performance criteria*, Mathematical finance, 29, 1003-1038.
- Zhu, H., Cao, M. and Zhang, C. (2019), *Time-consistent investment and reinsurance strategies for meanvariance insurers with relative performance concerns under the heston model*, Finance Research Letters, 30, 280-291.
- Zariphopoulou, T. (2024), *Mean field and n-player games in itô-diffusion markets under forward performance criteria*, Probability, Uncertainty and Quantitative Risk, 9(2), 123-148.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Case deletion diagnostic and Generalized Cook's distance Using Liu Estimator in Linear Mixed Measurement Error Models

Sh. Borhani¹, F. Ghapani^{2,*} and R. Jafaraghaie²

¹ Department of Statistics, Ahv.C., Islamic Azad University,
Ahvaz, Iran

² Department of Mathematics and Statistics, Sho.C., Islamic
Azad University, Shoushtar, Iran

Abstract

In this paper, we focused on identifying influential observations using Liu corrected likelihood estimator (LCLE) in linear mixed measurement error models when mul-ticollinearity is present. We considered an extension of Cook's distance to determine influential observations based on the case deletion model. Finally, simulation study was provided to illustrate the performance of the influence measures.

Keywords: Diagnostic · Influential observations -Cook's distance -Liu estimator

1 Introduction

Consider the following linear mixed model with measurement errors

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{U}\mathbf{b} + \boldsymbol{\varepsilon}, \quad \mathbf{X} = \mathbf{Z} + \boldsymbol{\Delta}, \quad (1)$$

where \mathbf{Z} and $\mathbf{U} = [\mathbf{U}_1 | \mathbf{U}_2 | \dots | \mathbf{U}_m]$ are $n \times p$ and $n \times q$ matrices of "regressors", respectively, where \mathbf{U}_i is an $n \times q_i$ known design matrix of the random effect factori, $\mathbf{b}' = (b'_1, b'_2, \dots, b'_m)$, $\boldsymbol{\Delta}' = [\delta_1, \delta_2, \dots, \delta_n]$ and $\mathbf{b} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma})$ that $\boldsymbol{\Sigma}$ is a block diagonal matrix with the i th block being $\gamma_i \mathbf{I}_{q_i}$ for $\gamma_i = \frac{\sigma_i^2}{\sigma^2}$, so that \mathbf{y} has a multivariate normal distribution with $E(\mathbf{y}) = \mathbf{Z}\boldsymbol{\beta}$ and $Var(\mathbf{y}) = \sigma^2 \mathbf{V}$, in which $\mathbf{V} = \mathbf{I}_n + \sum_{i=1}^m \gamma_i \mathbf{U}_i \mathbf{U}_i' = \mathbf{I}_n + \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}'$. Then the corrected likelihood estimates (CLE) of estimates are given by $\hat{\boldsymbol{\beta}} = [\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} - tr(\mathbf{V}^{-1})\boldsymbol{\Lambda}]^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \mathbf{A}^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ where $\mathbf{A} = [\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} - tr(\mathbf{V}^{-1})\boldsymbol{\Lambda}]$, $\hat{\sigma}^2 = n^{-1} [(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) - tr(\mathbf{V}^{-1})\hat{\boldsymbol{\beta}}' \boldsymbol{\Lambda} \hat{\boldsymbol{\beta}}]$ and $\hat{\sigma}_i^2 = \frac{\tilde{\mathbf{b}}_i' \tilde{\mathbf{b}}_i - tr(\hat{\mathbf{D}}_i' \hat{\mathbf{D}}_i) \hat{\boldsymbol{\beta}}' \boldsymbol{\Lambda} \hat{\boldsymbol{\beta}}}{[q_i - tr(\mathbf{T}_{ii})]}$; $i = 1, \dots, m$

in which $\hat{\mathbf{D}}_i = \hat{\gamma}_i \mathbf{U}_i' \mathbf{V}^{-1} = \frac{\hat{\sigma}_i^2}{\hat{\sigma}^2} \mathbf{U}_i' \mathbf{V}^{-1}$ and $\tilde{\mathbf{b}} = \boldsymbol{\Sigma} \mathbf{U}' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ is random effects prediction and \mathbf{T}_{ij} is ij th block of matrix \mathbf{T} (see Zhong et al., 2002 and Zare et al., 2012).

*Corresponding author. fateme.Ghapai@iau.ac.ir

Linear mixed models have tremendous flexibility to fit statistical models and are used to analyze data over a wide range of topics. It is a very common problem to deal with highly intercorrelated explanatory variables. There are several ways to overcome this problem such as Stein estimator (Stein, 1956) and Liu estimator (Liu, 1993). Sensitivity and residual analysis of the underlying assumptions constitute important tools for evaluating the fit of any statistical model to given data. For this end, based on Liu (1993) we consider the linear restriction $d\hat{\beta} = \beta + \mathbf{e}$, $\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_p)$ where $0 < d < 1$ is the Liu biasing parameter and $\hat{\beta}$ is the corrected likelihood estimator (CLE) in linear mixed measurement error model. The LCLE of β and σ^2 will be obtained as (Ghapani, 2020). $\hat{\beta}_d = \mathbf{A}_p^{-1}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} + d\hat{\beta})$ and

$$\hat{\sigma}_d^2 = n^{-1} \left[(\mathbf{y} - \mathbf{X}\hat{\beta}_d)' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\beta}_d) - \text{tr}(\mathbf{V}^{-1}) \hat{\beta}_d' \mathbf{\Lambda} \hat{\beta}_d + (d\hat{\beta} - \hat{\beta}_d)' (d\hat{\beta} - \hat{\beta}_d) \right]$$

where $\mathbf{A}_p = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} - \text{tr}(\mathbf{V}^{-1})\mathbf{\Lambda} + \mathbf{I}_p)$. Also, the Liu predictor of \mathbf{b} is given by

$$\tilde{\mathbf{b}}_d = \Sigma \mathbf{U}' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\beta}_d).$$

Theorem 1.1. $\hat{\beta}_d$ has an asymptotically normal distribution with mean $AE(\hat{\beta}_d) = \mathbf{G}_d \beta$ and covariance matrix $AVar(\hat{\beta}_d) = \mathbf{G}_d (\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1} (\mathbf{B} + \sigma^2 \mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}) (\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1} \mathbf{G}_d$, which $\mathbf{B} = n\sigma^2 + \beta' \mathbf{Z}' \mathbf{V}^{-1} \mathbf{Z} \beta$ and $\mathbf{G}_d = (\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z} + \mathbf{I}_p)^{-1} (\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z} + d\mathbf{I}_p)$.

Proof. (see Ghapani, 2020) □

Corollary 1.2. $\hat{\beta}$ has an asymptotically normal distribution with mean $AE(\hat{\beta}) = \beta$ and covariance matrix $AVar(\hat{\beta}) = (\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1} (\mathbf{B} + \sigma^2 \mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}) (\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1}$.

Theorem 1.3. There exists $0 < l_i = \frac{\lambda_i \alpha_i^2 - \sigma^2 (2\lambda_i + 1)}{\lambda_i \alpha_i^2 + \sigma^2} < d < 1$, $i = 1, \dots, p$ such that $\mathbf{\Delta}$ is a positive definite matrix. (explanation: The spectral decomposition of the matrix $\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z} > 0$ is $\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z} > \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}'$ where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$, $\lambda_i > 0$ and λ_i are eigenvalues of $\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}$, there exists some orthogonal matrix \mathbf{Q} , We denote $\alpha = \mathbf{Q}'\beta$, then, we have $\beta = \mathbf{Q}\alpha$.)

Proof. (see Ghapani, 2020) □

2 Diagnostic methods

To determine the observations that behave abnormally compared to other observations, various methods have used by statisticians. We study some of these measures in the context of linear mixed measurement error models under the LCLE.

2.1 Case deletion diagnostic

To quantify the effects of deleting the i th observation on estimation of parameters, a fundamental approach is case deletion model that, for the Liu estimator with the i th observation, deleted in linear mixed measurement error models, is defined as

$$\begin{aligned} \mathbf{y}_{(i)} &= \mathbf{Z}_{(i)}\beta + \mathbf{U}_{(i)}\mathbf{b} + \boldsymbol{\varepsilon}_{(i)}, \quad \mathbf{X}_{(i)} = \mathbf{Z}_{(i)} + \mathbf{\Delta}_{(i)}, \quad i = 1, 2, \dots, n \\ d\hat{\beta} &= \beta + \mathbf{e}, \end{aligned} \tag{2}$$

Theorem 2.1. For model (2), we have

$$\begin{aligned} \hat{\beta}_{d(i)} &\simeq \hat{\beta}_d - \mathbf{A}_p^{-1} \frac{\mathbf{X}'_i \mathbf{c}_i \hat{\mathbf{v}}_{di}}{r_{pii}} \\ \hat{\sigma}_{d(i)}^2 &\simeq \frac{n}{n-1} \hat{\sigma}_d^2 - \mathbf{c}'_i (\mathbf{y} - \mathbf{X}\hat{\beta}_d) \frac{\hat{\mathbf{v}}_{di}}{(n-1)r_{pii}} = \frac{n}{n-1} \hat{\sigma}_d^2 - \frac{\hat{\mathbf{v}}_{di}^2}{(n-1)r_{pii}} \\ \tilde{\mathbf{b}}_{d(i)} &\simeq \tilde{\mathbf{b}}_d - \Sigma \mathbf{U}'_{pi} \frac{\hat{\mathbf{v}}_{di}}{r_{pii}} \end{aligned}$$

where $\hat{\beta}_{d(i)}$, $\hat{\sigma}_{d(i)}^2$ and $\tilde{\mathbf{b}}_{d(i)}$ denote the estimates of β , σ^2 and \mathbf{b} when the i th case is deleted, respectively. Also, \mathbf{c}'_i and \mathbf{r}'_{pi} are i th rows of \mathbf{V}^{-1} and $\mathbf{R}_p = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}\mathbf{A}_p^{-1}\mathbf{X}'\mathbf{V}^{-1}$, respectively, c_{ii} and r_{pii} denotes the i th diagonal elements of \mathbf{V}^{-1} and \mathbf{R}_p , $\hat{\mathbf{v}}_{di} = \mathbf{c}'_i (\mathbf{y} - \mathbf{X}\hat{\beta}_d)$ is the i th elements of $\hat{\mathbf{v}}_d$.

Proof. When the i th row is deleted from the data set, we have

$$\hat{\beta}_{d(i)} = (\mathbf{X}_{(i)}' \mathbf{V}_{[i]}^{-1} \mathbf{X}_{(i)} - \text{tr}(\mathbf{V}_{[i]}^{-1})\mathbf{\Lambda} + \mathbf{I}_p)^{-1} (\mathbf{X}_{(i)}' \mathbf{V}_{[i]}^{-1} \mathbf{y}_{(i)} + d\hat{\beta}),$$

where $\mathbf{V}_{[i]}^{-1}$ denotes inverse matrix \mathbf{V} with the i th row and column removed. Using $(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}$ (Rao et al. 2008, Theorem A. 18), $\hat{\beta}_{d(i)}$ can be written as \square

$$\begin{aligned} \hat{\beta}_{d(i)} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} - \frac{\mathbf{X}'\mathbf{c}_i\mathbf{c}_i'\mathbf{X}}{c_{ii}} - \text{tr}(\mathbf{V}^{-1})\mathbf{\Lambda} + c_{ii}\mathbf{\Lambda} + \mathbf{I}_p)^{-1} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} - \frac{\mathbf{X}'\mathbf{c}_i\mathbf{c}_i'\mathbf{y}}{c_{ii}} + d\hat{\beta}) \\ &= (\mathbf{A}_p^{-1} + \mathbf{A}_p^{-1} \frac{\mathbf{X}'\mathbf{c}_i}{c_{ii}} (\mathbf{I}_n - \frac{\mathbf{c}_i'\mathbf{X}\mathbf{A}_p^{-1}\mathbf{X}'\mathbf{c}_i}{c_{ii}})^{-1} \mathbf{c}_i'\mathbf{X}\mathbf{A}_p^{-1}) (\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} + d\hat{\beta} - \frac{\mathbf{X}'\mathbf{c}_i\mathbf{c}_i'\mathbf{y}}{c_{ii}}) + \mathbf{O}_p(n^{-1}) \\ &\simeq \beta_d - \mathbf{A}_p^{-1} \frac{\mathbf{X}'\mathbf{c}_i\hat{\mathbf{v}}_{di}}{r_{pii}}. \end{aligned}$$

As the same way

$$\begin{aligned} (n-1)\hat{\sigma}_{d(i)}^2 &= (\mathbf{y}_{(i)} - \mathbf{X}_{(i)}\hat{\beta}_{d(i)})' \mathbf{V}_{[i]}^{-1} (\mathbf{y}_{(i)} - \mathbf{X}_{(i)}\hat{\beta}_{d(i)}) - \text{tr}(\mathbf{V}_{[i]}^{-1})\hat{\beta}_{d(i)}\mathbf{\Lambda}\hat{\beta}_{d(i)} \\ &\quad - (d\hat{\beta} - \hat{\beta}_{d(i)})'(d\hat{\beta} - \hat{\beta}_{d(i)}) \end{aligned}$$

Then,

$$\hat{\sigma}_{d(i)}^2 \simeq \frac{n}{n-1}\hat{\sigma}_d^2 - \mathbf{c}_i'(\mathbf{y} - \mathbf{X}\hat{\beta}_d) \frac{\hat{\mathbf{v}}_{di}}{(n-1)r_{pii}} = \frac{n}{n-1}\hat{\sigma}_d^2 - \frac{\hat{\mathbf{v}}_{di}^2}{(n-1)r_{pii}} \quad (3)$$

And $\tilde{\mathbf{b}}_{d(i)} = \Sigma\mathbf{U}'_{(i)}\mathbf{V}_{[i]}^{-1}(\mathbf{y}_{(i)} - \mathbf{X}_{(i)}\hat{\beta}_{d(i)}) \simeq \tilde{\mathbf{b}}_d - \Sigma\mathbf{U}'(\mathbf{c}_i - \mathbf{V}^{-1}\mathbf{X}\mathbf{A}_p^{-1}\mathbf{X}'\mathbf{c}_i) \frac{\hat{\mathbf{v}}_{di}}{r_{pii}}$
 $\quad = \tilde{\mathbf{b}}_d - \Sigma\mathbf{U}'\mathbf{r}_{pi} \frac{\hat{\mathbf{v}}_{di}}{r_{pii}}$

2.2 Generalized Cook's distance

Cook's distance is a convenient measure of influence, which based on Cook (1977) is denoted by $CD_i(\beta)$ and in linear mixed measurement error models under LCLE is given as

$$CD_i(\beta) = (\hat{\beta}_d - \hat{\beta}_{d(i)})' \mathbf{M} (\hat{\beta}_d - \hat{\beta}_{d(i)}),$$

where $\mathbf{M} = \hat{\sigma}_d^{-2} [\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} - \text{tr}(\mathbf{V}^{-1})\mathbf{\Lambda} + \mathbf{I}_p]$. Then, we have

$$CD_i(\beta) = \frac{(\hat{\beta}_d - \hat{\beta}_{d(i)})' [\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} - \text{tr}(\mathbf{V}^{-1})\mathbf{\Lambda} + \mathbf{I}_p] (\hat{\beta}_d - \hat{\beta}_{d(i)})}{\hat{\sigma}_d^2},$$

then we have approximately $CD_i(\beta) = \frac{(c_{ii} - r_{pii})\hat{\mathbf{v}}_{di}^2}{\hat{\sigma}_d^2 r_{pii}} + \mathbf{O}_p(n^{-1})$.

2.3 Cook's distance for random effects

The Cook's distance for random effects is define as

$$CD_i(\mathbf{b}) = \frac{(\tilde{\mathbf{b}}_d - \tilde{\mathbf{b}}_{d(i)})' (\mathbf{U}'\mathbf{U} + \Sigma^{-1}) (\tilde{\mathbf{b}}_d - \tilde{\mathbf{b}}_{d(i)})}{\hat{\sigma}_d^2}.$$

Since $\tilde{\mathbf{b}}_{d(i)} \simeq \tilde{\mathbf{b}}_d - \Sigma\mathbf{U}'\mathbf{r}_{pi} \frac{\hat{\mathbf{v}}_{di}}{r_{pii}}$.

Simulation study

To assess the performance of the $CD_i(\beta_d)$ and $CD_i(\mathbf{b}_d)$ in terms of type I error and power of the test, we carried out the following parametric bootstrap simulation study. The i th simulated data set was generated as

$$\begin{aligned} \mathbf{y}_i &= \mathbf{Z}\beta + \mathbf{U}\mathbf{b}_i + \varepsilon_i, \quad i = 1, \dots, 1000 \\ \mathbf{X}_i &= \mathbf{Z} + \Delta \end{aligned} \quad (4)$$

We consider the following combinations for simulation: $m = 4$ or $m = 10$, $n_i = 7$, $z_{i1} \sim N(3, 1)$, $z_{i2} \sim N(3, 1)$ and $z_{i3} = z_{i2} + v$ in which $v \sim N(0, 1)$ for $i = 1, \dots, n$ and $\mathbf{b}_i \sim N(0, \sigma_1^2)$ for $\sigma_1^2 = 0.16$ and 0.25 and the random errors $\varepsilon_i \sim N(0, \sigma^2)$ for $\sigma^2 = 0.36$ and $\sigma^2 = 0.64$

and $\mathbf{\Lambda} = \text{diag}(0, 0.05, 0.05, 0.05)$. Furthermore, for each simulated data set, $CD_i(\boldsymbol{\beta}_d)$ and $CD_i(\mathbf{b}_d)$ were calculated for the first observation. The choice of the first observation was arbitrary. Similar to Ghapani (2019), we generate an empirical distribution of the test statistic under the null hypothesis, the data sets for $j = 1, \dots, 1000$ were simulated as

$$\begin{aligned} \mathbf{y}_{ij}^* &= \hat{\mathbf{Z}}_{dj} \hat{\boldsymbol{\beta}}_{dj} + \mathbf{U} \mathbf{b}_{ij}^* + \boldsymbol{\varepsilon}_{ij}^*, \\ \mathbf{X}_{ij}^* &= \hat{\mathbf{Z}}_{dj} + \boldsymbol{\Delta}_{ij}, \end{aligned}$$

where $\boldsymbol{\varepsilon}_{ij}^*$, \mathbf{b}_{ij}^* and $\boldsymbol{\Delta}_{ij}$ have a normal distribution with zero mean and variances $\hat{\sigma}_i^2 \mathbf{I}_n$, $\hat{\sigma}_{1di}^2 \mathbf{I}_q$ and $\mathbf{I}_n \otimes \mathbf{\Lambda}$ respectively. Also, $\hat{\boldsymbol{\beta}}_{di}$, $\hat{\mathbf{Z}}_{di}$, $\hat{\sigma}_{1di}^2$ and $\hat{\sigma}_{di}^2$ are the corrected Liu estimates of $\boldsymbol{\beta}$, \mathbf{Z} , σ^2 and σ_1^2 from model (4). The $CD_i(\boldsymbol{\beta}_d)$ and $CD_i(\mathbf{b}_d)$ were performed for the first observation of each simulated data and $100(1 - \alpha)$ percentiles from the empirical distribution of test statistic were used as the threshold value of the test statistic of model (4). The probability of a type I error estimate for the $CD_i(\boldsymbol{\beta}_d)$, $CD_i(\mathbf{b}_d)$ and $\alpha = 0.05$ were calculated as the number of data sets for which the $CD_i(\boldsymbol{\beta}_d)$ and $CD_i(\mathbf{b}_d)$ exceeded the $100(1 - \alpha)$ percentiles of the empirical distribution, divided by the number of replicates. The results are listed in Table 1. It appears that in general the type I error of the $CD_i(\boldsymbol{\beta}_d)$ and $CD_i(\mathbf{b}_d)$ for different combinations of parameters are close to the nominal value of 0.05. In the next step, in order to evaluate the relative sensitivity of the $CD_i(\boldsymbol{\beta}_d)$ and $CD_i(\mathbf{b}_d)$, we consider:

$z_{11} \sim N(5, 3)$, $z_{12} \sim N(5, 3)$ and $z_{13} = z_{12} + v$ with $\varepsilon_1 = 0.5$,

generated as a point with high leverage for the first observation and again for each combination of parameters, 1000 data sets are generated from the following model

$$\begin{aligned} \mathbf{y}_i &= \mathbf{Z}_i \boldsymbol{\beta} + \mathbf{U} \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \\ \mathbf{X}_i &= \mathbf{Z}_i + \boldsymbol{\Delta}_i \end{aligned} \quad i = 1, \dots, 1000$$

The parameters remained as in the evaluation of type I error. Again, for each simulated data set, we derive the Liu estimate of parameters and the $CD_i(\boldsymbol{\beta}_d)$ and $CD_i(\mathbf{b}_d)$ for the first observation. The power of the $CD_i(\boldsymbol{\beta}_d)$ and $CD_i(\mathbf{b}_d)$ were calculated as the number of data sets for which the $CD_i(\boldsymbol{\beta}_d)$ and $CD_i(\mathbf{b}_d)$ exceeded the percentile $100(1 - \alpha)$ of the empirical distribution, divided by the number of replicates. According to Table 1, we found that the estimates obtained for the linear mixed measurement error models under LCLE are more powerful than the CLE. Moreover, we can see that power of the $CD_i(\boldsymbol{\beta}_d)$ and $CD_i(\mathbf{b}_d)$ also increase as the sample size increases.

Table1: Comparison of the power of Cook's distance for CLE and LCLE

n	(σ_1^2, σ^2)	$CD(\boldsymbol{\beta})$		$CD(\boldsymbol{\beta}_d)$		$CD(\mathbf{b})$		$CD(\mathbf{b}_d)$	
		Siglevel	Power	Siglevel	Power	Siglevel	Power	Siglevel	Power
70	(0.16, 0.36)	0.048	0.634	0.031	0.676	0.047	0.113	0.043	0.123
70	(0.25, 0.64)	0.040	0.599	0.030	0.635	0.036	0.083	0.044	0.118
28	(0.16, 0.36)	0.064	0.422	0.04	0.456	0.063	0.101	0.040	0.129
28	(0.25, 0.64)	0.067	0.418	0.039	0.450	0.070	0.048	0.036	0.094

Discussion

In this paper, we have presented diagnostic methods based on LCLE in linear mixed measurement error models. Based on LCLE, we obtained case deletion diagnostics for the detection of the influential observations in these models. The final numerical results and simulation studies have demonstrated that the diagnostic methods presented in this paper are useful to detect influential observations.

References

- [1] Cook, R.D. (1977). Detection of influential observations in linear regression. *Technometrics*, 19, 15-18.

- [2] Ghapani, F. (2020). Stochastic restricted liu estimator in linear mixed measurement error models. *Commun. Statist. Simul. Computat.*, 51(4), 1220-1233.
- [3] Liu, K. (1993). A new class of biased estimate in linear regression, *Comm. Stat. Theory Methods*, 22, 393-402.
- [4] Stein, C. (1956). Inadmissibility of the usual estimator for mean of multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.* 1, 197-206.
- [5] Stein, C. (1956). Inadmissibility of the usual estimator for mean of multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.* 1, 197-206.
- [6] Agarwal, R.P. O'Regan, D. Sahu, D.R. , *Fixed Point Theory for Lipschitzian-type Mappings with Applications Series*,: Topological Fixed Point Theory and Its Applications, vol. 6. Springer, New York (2009).
- [7] Zare, K. and Rasekh, A. (2011). Diagnostic measures for linear mixed measurement error models, *Sort.*, 35, 125-144.
- [8] are, K., Rasekh, A. and Rasekhi, A. (2012). Estimation of variance components in linear mixed measurement error models. *Stat. Papers.* 53, 849-863.
- [9] Zhong, X. P., Wei, B. C. and Fung, W. K. (2000). Influence analysis for linear measurement error models. *Ann. Inst. Stat. Math* ,52, 367-379.
- [10] Nakamura, T. (1990). Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika*, 77, 127-137.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Some Inequalities for Absolute Moments of Feasible Acceptable Random Variables

Negar Eghbal^{*1}

¹Department of Statistics, Shahrood University of Technology

Abstract

We introduce a new class of dependent random variables based on the notion of acceptability in terms of the characteristic function. Such a view ensures that this new definition always makes sense, which is not the case for the classical acceptability given by the moment generating function. We then extend some of the results of inequalities for absolute moments of sums of independent random variables, for this class of dependent random variables.

Keywords: Dependent Random Variables, Characteristic Function, Feasible Acceptable Random Variables.

Mathematics Subject Classification (2020): 60E10, 60E15.

1 Introduction

Ushakov (2011) proposed several inequalities for absolute moments of sums and differences of independent random variables, in terms of characteristic functions. Such moment inequalities could be applied to obtain weak and strong laws of large numbers and their convergence rates (see, for example, Fazekas and Pecsora, 2017 and references therein). Further, by using the results of Ushakov (2011), Ushakov and Ushakov (2017) introduced some non-parametric tests for assessing the homogeneity of two or more distributions.

Although the independent case is classical in the literature, the use of dependent random variables can be more relevant. There are different structures for dependent random variables. One of the new exciting classes of dependent random variables is *acceptable*, which was first introduced in Antonini et al. (2008). The concept of acceptability is defined as follows; Let $\{X_n; n \geq 1\}$ be a sequence of random variables defined on a fixed probability space (Ω, \mathcal{F}, P) . We say that a finite family of random variables $\{X_j; j = 1, 2, \dots, n\}$ is acceptable if for any real t ,

$$\mathbb{E} \exp \left\{ t \sum_{j=1}^n X_j \right\} \leq \prod_{j=1}^n \mathbb{E} \exp \{ t X_j \}.$$

A sequence of random variables $\{X_n; n \geq 1\}$ is acceptable if every finite subfamily is acceptable. Obviously, the class of acceptable random variables contains independent random

^{*}Corresponding author, n.eghbal@shahroodut.ac.ir

variables as a particular case. More importantly, this class includes negatively associated random variables (Alam and Saxena, 1981) (and hence negatively dependent random variables (Lehmann, 1966)) with a finite Laplace transform (or finite moment generating function) as well. Some statisticians provided weaker representations of acceptability, such as extended acceptability (Choi and Baek, 2013) or wide acceptability (Wang et al. 2011). See also Christofides et al. (2016), Ndiaye and Lo (2013), and Sung et al. (2011) for more definitions and details.

In all proposed versions for acceptability, the authors implicitly assumed that the random variables have a finite moment generating function (for all t or at least near zero, according to Sung et al. (2011)). However, it is not always the case. A trivial example is a sequence of random variables that follow a Cauchy distribution. Unlike the moment generating function, the characteristic function of any real-valued random variable always exists. Following this point, our first intention is to provide a new concept of acceptability, named feasible acceptability, employing the characteristic function. Such a new definition is always valid. Our second purpose, which is more interesting, is to extend some of the results of Ushakov (2011) to the proposed class of feasible acceptable random variables.

2 Feasible Acceptability

Definition 2.1. A finite family of random variables $\{X_j; j = 1, 2, \dots, n\}$ is feasible acceptable if for any real t ,

$$\mathbb{E} \exp \left\{ it \sum_{j=1}^n X_j \right\} \leq \prod_{j=1}^n \mathbb{E} \exp \{ it X_j \}$$

where i denotes the imaginary unit with $i^2 = -1$. A sequence of random variables $\{X_n; n \geq 1\}$ is feasible acceptable if every finite subfamily is feasible acceptable.

In the spirit of the characteristic function, a finite family of random variables $\{X_j; j = 1, 2, \dots, n\}$ is feasible acceptable if

$$\varphi_{\sum_{j=1}^n X_j}(t) \leq \prod_{j=1}^n \varphi_{X_j}(t),$$

where $\varphi_X(t) = E(e^{itX})$ is the characteristic function of random variable X . Indeed, it is straightforward to develop Definition 2.1 to cover similar extended versions of acceptability, such as those introduced by Sung et al. (2011) and Wang et al. (2011).

The Fourier transformation $e^{itx} = \cos(tx) + i \sin(tx)$ conveys the realization of the random variable X from the real line to the complex plane, and by the characteristic function, we should work with complex numbers. Although complex numbers do not have a natural ordering similar to real numbers, however, in this work, we suppose the random variables and their (partial) sums are symmetric about zero for which the characteristic function is a real function. Hence, we can use the natural linear ordering of real numbers.

The random variables given in the following example belong to the class of feasible acceptable random variables. Note that in this example, the random variables and their partial sums are symmetric about zero.

Example 2.2. Let (X_1, \dots, X_n) follows a multivariate normal distribution with the zero-mean vector and the covariance matrix $\Sigma = (\sigma_{jk})$ where for $j = 1, \dots, n$, $\sigma_{jj} \geq 0$ and for $j, k = 1, \dots, n$, $\sigma_{jk} = \sigma_{kj}$. It is easy to see $X_j \sim N(0, \sigma_{jj})$, $j = 1, \dots, n$, and

$$\sum_{j=1}^n X_j \sim N(0, \sum_{j=1}^n \sum_{k=1}^n \sigma_{jk}).$$

Hence,

$$\varphi_{\sum_{j=1}^n X_j}(t) = \exp \left\{ -\frac{t^2}{2} \sum_{j=1}^n \sum_{k=1}^n \sigma_{jk} \right\} \quad \& \quad \prod_{j=1}^n \varphi_{X_j}(t) = \prod_{j=1}^n \exp \left\{ -\frac{t^2}{2} \sigma_{jj} \right\}.$$

So (X_1, \dots, X_n) is a vector of feasible acceptable random variables if $\sum_{j \neq k} \sigma_{jk} > 0$.

According to a referee's comment, the symmetry of individual random variable does not guarantee the symmetry of their sum. Hence, to ensure that the partial sums of symmetric random variables about zero are also symmetric, we use the definition of joint symmetry for the random variables.

Definition 2.3. *The random variables X_1, X_2, \dots, X_n are jointly symmetric if and only if they have the same joint distribution as $-X_1, -X_2, \dots, -X_n$. In particular, if the X_i 's are independent and each one symmetric, then they are jointly symmetric.*

Behboodian (1990) proved that a necessary and sufficient condition for symmetry of sums of random variables $\{X_j; j = 1, 2, \dots, n\}$ is the joint symmetry of them.

Lemma 2.4. *(Behboodian, 1990). The random variables X_1, X_2, \dots, X_n are jointly symmetric if and only if any linear combination $\sum_{j=1}^n a_j X_j$, with real a_j 's is symmetric.*

The joint symmetry could be a relevant condition in practical situations and hence not restrictive; To this end, let us discuss the Farlie-Gumbel-Morgenstern (FGM) family of distributions. For any pair of random variables X and Y with distribution functions $F_1(x)$ and $F_2(y)$ respectively, the FGM family of bivariate distributions is given by

$$F(x, y) = F_1(x)F_2(y)[1 + \theta(1 - F_1(x))(1 - F_2(y))], \quad -1 \leq \theta \leq 1$$

originally considered by Farlie (1960), Gumbel (1960), and Morgenstern (1956). Johnson and Kotz (1975) extended the family to m variables for $m > 2$. By the following lemma, we will show that if the random variables X and Y are symmetric and their joint distribution belongs to the FGM family, they are jointly symmetric.

Lemma 2.5. *Let X and Y be two symmetric random variables about zero, with distribution functions $F_1(x)$ and $F_2(y)$, respectively. If the joint distribution function of (X, Y) belongs to the FGM family, therefore they are jointly symmetric about zero.*

Proof. By using Definition 2.3, we should show

$$F(x, y) = \bar{F}(-x, -y),$$

for every real values x and y , in which $\bar{F}(x, y) = P(X \geq x, Y \geq y)$ is the joint survival function. We can write

$$\begin{aligned} \bar{F}(-x, -y) &= 1 - F_1(-x) - F_2(-y) + F(-x, -y) \\ &= 1 - F_1(-x) - F_2(-y) + F_1(-x)F_2(-y)[1 + \theta F_1(x)F_2(y)] \\ &= F_1(x)F_2(y) + \theta F_1(x)F_2(y)(1 - F_1(x))(1 - F_2(y)) \\ &= F(x, y). \end{aligned}$$

Hence, the proof is complete. \square

The following example illustrates the feasible acceptability of random variables that are jointly symmetric about zero.

Example 2.6. *Let X and Y be two identically distributed random variables with the support $\{-1, 0, 1\}$ for which $P(X = -1) = P(X = 1) = \frac{1}{4}$ and $P(X = 0) = \frac{1}{2}$. Further, suppose that the joint probability mass function of (X, Y) for $\{(-1, 0), (0, -1), (0, 1), (1, 0)\}$ is equal to $\frac{1}{4}$ and for the other values is equal to zero. Therefore,*

$$\varphi_{X+Y}(t) = \frac{e^{-it} + e^{it}}{2} = \cos t \quad \& \quad \varphi_X(t) = \varphi_Y(t) = \frac{e^{-it} + e^{it} + 2}{4} = \frac{\cos t + 1}{2}.$$

Then, using simple arguments, we can show that for every t , $\varphi_{X+Y}(t) \leq \varphi_X(t)\varphi_Y(t)$.

Due to the inherent properties of the characteristic function, the proposed feasible acceptability is not totally in agreement with the acceptability of Antonini et al. (2008). The following example explains that the acceptability does not necessarily imply feasible acceptability and vice versa.

Example 2.7. Let X and Y be two identically distributed random variables with the support $\{-1, 1\}$ for which $P(X = -1) = P(X = 1) = \frac{1}{2}$. Let also the joint probability mass function of (X, Y) for $\{(-1, -1), (1, 1)\}$ is equal to $\frac{1}{4} - q$ and for $\{(-1, 1), (1, -1)\}$ is equal to $\frac{1}{4} + q$ in which $-\frac{1}{4} \leq q \leq \frac{1}{4}$. We can see that

$$\begin{aligned}\varphi_{X+Y}(t) &= \varphi_X(t)\varphi_Y(t) - q(e^{2it} + e^{-2it} - 2) \\ &= \varphi_X(t)\varphi_Y(t) - 2q(\cos 2t - 1)\end{aligned}$$

and

$$M_{X+Y}(t) = M_X(t)M_Y(t) - q(e^{2t} + e^{-2t} - 2)$$

where $M_X(t) = E(e^{tX})$ is the moment generating function of random variable X . Therefore, for $-\frac{1}{4} \leq q < 0$, $\varphi_{X+Y}(t) \leq \varphi_X(t)\varphi_Y(t)$, while $M_{X+Y}(t) \geq M_X(t)M_Y(t)$. Reversely, for $0 < q \leq \frac{1}{4}$, $\varphi_{X+Y}(t) \geq \varphi_X(t)\varphi_Y(t)$, while $M_{X+Y}(t) \leq M_X(t)M_Y(t)$.

3 Main Results

Many inequalities for absolute moments of independent random variables, under different conditions, have been introduced. A powerful tool to achieve such inequalities is the characteristic function that has attracted the interest of several statisticians like Hsu (1951), Von Bahr (1965), Von Bahr and Esseen (1965) and Ushakov (2011). In this paper, we will establish the same results of Ushakov (2011) for the feasible acceptable random variables.

The following two lemmas play an essential role in our main results.

Lemma 3.1. (Von Bahr and Esseen, 1965). Let X be a random variable with the distribution function $F(x)$ and the characteristic function $\varphi(t)$. If $E|X|^p < \infty$, then

$$E|X|^p = K(p) \int_{-\infty}^{\infty} \frac{1 - \Re(\varphi(t))}{|t|^{p+1}} dt, \quad 0 < p < 2,$$

where \Re stands for the real part and $K(p) = \frac{\Gamma(p+1)}{\pi} \sin \frac{p\pi}{2}$.

Lemma 3.2. (Von Bahr, 1965). Let $H(x)$ be a function of bounded variation on $(-\infty, \infty)$ with the finite absolute moment

$$\delta_p = \int_{-\infty}^{\infty} |x|^p |dH(x)| < \infty, \quad p > 0, \quad p \neq \text{even integer}$$

and with moments

$$\gamma_j = \int_{-\infty}^{\infty} x^j dH(x), \quad j = 0, 1, \dots, j \leq p$$

and let $h(t)$ be the corresponding Fourier-Stieltjes transform. Then

$$\begin{aligned}\int_{-\infty}^{\infty} |x|^p dH(x) &= \frac{\Gamma(p+1)}{\pi} \cos \frac{(p+1)\pi}{2} \\ &\quad \int_{-\infty}^{\infty} [\Re(h(t)) - \sum_{\ell=0}^s [(-1)^\ell \gamma_{2\ell} t^{2\ell} / (2\ell)!] / |t|^{p+1}] dt\end{aligned}$$

where s is the integer part of $\frac{p}{2}$.

By using Lemma 3.2, we can give the following corollary.

Corollary 3.3. *Let X be a random variable with the distribution function $F(x)$ and the characteristic function $\varphi(t)$. Suppose that $E|X|^p < \infty$, where $p > 0$ and p is not an even integer. For nonnegative integer ℓ , denote $\alpha_\ell = EX^\ell$. Then*

$$E|X|^p = C(p) \int_{-\infty}^{\infty} \frac{\Re(\varphi(t)) - \sum_{\ell=0}^s (-1)^\ell \alpha_{2\ell} t^{2\ell} / (2\ell)!}{|t|^{p+1}} dt,$$

where s is the integer part of $\frac{p}{2}$, and $C(p) = \frac{\Gamma(p+1)}{\pi} \cos \frac{(p+1)\pi}{2}$.

The following result improves Theorem 2 of Ushakov (2011) for the feasible acceptable random variables.

Theorem 3.4. *Let X and Y be two feasible acceptable symmetric random variables about zero with finite absolute moments of order p that their sum is also symmetric about zero. Further, let X' be an independent of X random variable having the same distribution as X , Y' an independent of Y random variable having the same distribution as Y .*

1. *If $0 < p < 2$ then*

$$E|X + X'|^p + E|Y + Y'|^p \leq 2E|X + Y|^p.$$

2. *If X and Y are uncorrelated random variables and $2 < p < 4$, then*

$$E|X + X'|^p + E|Y + Y'|^p \geq 2E|X + Y|^p.$$

Proof. For the first part, using Lemma 3.1 and our definition of feasible acceptability, we have

$$\begin{aligned} 2E|X + Y|^p &= 2K(p) \int_{-\infty}^{\infty} \frac{1}{|t|^{p+1}} (1 - \varphi_{X+Y}(t)) dt \geq 2K(p) \int_{-\infty}^{\infty} \frac{1}{|t|^{p+1}} (1 - \varphi_X(t)\varphi_Y(t)) dt \\ &\geq K(p) \left[\int_{-\infty}^{\infty} \frac{1}{|t|^{p+1}} (1 - \varphi_X^2(t)) dt + \int_{-\infty}^{\infty} \frac{1}{|t|^{p+1}} (1 - \varphi_Y^2(t)) dt \right] \\ &= E|X + X'|^p + E|Y + Y'|^p. \end{aligned}$$

Now, for the second part, using Corollary 3.3 and Definition 2.1, we can write

$$\begin{aligned} 2E|X + Y|^p &= 2C(p) \int_{-\infty}^{\infty} \frac{1}{|t|^{p+1}} \left(\varphi_{X+Y}(t) + E(X + Y)^2 \frac{t^2}{2} - 1 \right) dt \\ &\leq 2C(p) \int_{-\infty}^{\infty} \frac{1}{|t|^{p+1}} \left(\varphi_X(t)\varphi_Y(t) + (EX^2 + EY^2) \frac{t^2}{2} - 1 \right) dt \\ &\leq C(p) \int_{-\infty}^{\infty} \frac{1}{|t|^{p+1}} \left(\varphi_X^2(t) + \varphi_Y^2(t) + E|X + X'|^2 \frac{t^2}{2} + E|Y + Y'|^2 \frac{t^2}{2} - 2 \right) dt \\ &= C(p) \int_{-\infty}^{\infty} \frac{1}{|t|^{p+1}} \left(\varphi_X^2(t) + E|X + X'|^2 \frac{t^2}{2} - 1 \right) dt \\ &+ C(p) \int_{-\infty}^{\infty} \frac{1}{|t|^{p+1}} \left(\varphi_Y^2(t) + E|Y + Y'|^2 \frac{t^2}{2} - 1 \right) dt \\ &= E|X + X'|^p + E|Y + Y'|^p. \end{aligned}$$

□

Von Bahr and Esseen (1965) showed that if X_1, \dots, X_n be identically distributed random variables that are conditionally symmetric, then for $1 \leq p \leq 2$,

$$E|X_1 + \dots + X_n|^p \leq nE|X_1|^p. \quad (1)$$

The upper bound given in (1) holds for the independent random variables. Ushakov (2011) added a lower bound for the moment of the sum when the random variables are independent. We now obtain the same result for the feasible acceptable random variables whose sum is symmetric about zero. According to Lemma 2.4, the result holds for feasible acceptable random variables that are jointly symmetric as well. To this end, we use the following inequality obtained by Heathcote and Pitman (1972).

Lemma 3.5. For any characteristic function $\varphi(t)$ and any positive integer n ,

$$1 - \Re(\varphi(nt)) \leq n(1 - \Re(\varphi(t)))$$

for all real t .

Theorem 3.6. Let $\{X_j; j = 1, 2, \dots, n\}$ be feasible acceptable identically distributed random variables that are symmetric about zero. Further, suppose that their sum is symmetric about zero. Then, for any $1 < p < 2$,

$$n^{p-1} \mathbb{E}|X_1|^p \leq \mathbb{E}|X_1 + \dots + X_n|^p.$$

Proof. Using Lemma 3.1, Lemma 3.5, and the definition of feasible acceptability, we obtain

$$\begin{aligned} n\mathbb{E}|X_1 + \dots + X_n|^p &= K(p) \int_{-\infty}^{\infty} \frac{n}{|t|^{p+1}} (1 - \varphi_{X_1+\dots+X_n}(t)) dt \geq K(p) \int_{-\infty}^{\infty} \frac{n}{|t|^{p+1}} (1 - \varphi_{X_1}^n(t)) dt \\ &\geq K(p) \int_{-\infty}^{\infty} \frac{1}{|t|^{p+1}} (1 - \varphi_{X_1}(nt)) dt = \mathbb{E}|nX_1|^p = n^p \mathbb{E}|X_1|^p. \end{aligned}$$

□

The following result also extends Theorem 5 of Ushakov (2011) for the feasible acceptable random variables.

Theorem 3.7. Let X and Y be two feasible acceptable identically distributed symmetric random variables about zero that their sum is also symmetric about zero. For any $0 < p < 2$, we have

$$2^{p-1} \mathbb{E}|X|^p \leq \mathbb{E}|X + Y|^p. \quad (2)$$

Proof. Using lemmas 3.1 and 3.5, we obtain

$$\begin{aligned} 2\mathbb{E}|X + Y|^p &= K(p) \int_{-\infty}^{\infty} \frac{2}{|t|^{p+1}} (1 - \Re(\varphi_{X+Y}(t))) dt \geq K(p) \int_{-\infty}^{\infty} \frac{2}{|t|^{p+1}} (1 - \Re(\varphi_X^2(t))) dt \\ &\geq K(p) \int_{-\infty}^{\infty} \frac{2}{|t|^{p+1}} (1 - \Re^2(\varphi_X(t))) dt \geq K(p) \int_{-\infty}^{\infty} \frac{1 - \Re(\varphi_{2X}(t))}{|t|^{p+1}} dt = \mathbb{E}|2X|^p. \end{aligned}$$

□

It follows from (2) that if $p \geq 1$, then $\mathbb{E}|X|^p \leq \mathbb{E}|X + Y|^p$. However, this is not necessary true when $p < 1$. The following example explains it.

Example 3.8. Let X and Y be two identically distributed random variables with the support $\{-1, 1\}$ for which $P(X = -1) = P(X = 1) = \frac{1}{2}$. Let also the joint probability mass function of (X, Y) for $\{(-1, -1), (1, 1)\}$ is equal to $\frac{3}{8}$ and for $\{(-1, 1), (1, -1)\}$ is equal to $\frac{1}{8}$. Then, they are also two feasible acceptable random variables. Indeed,

$$\mathbb{E}|X|^p = 1 \quad \& \quad \mathbb{E}|X + Y|^p = 2^p \frac{3}{4}.$$

Hence, for example for $p = \frac{1}{3}$,

$$\mathbb{E}|X|^p > \mathbb{E}|X + Y|^p.$$

References

- Alam K., and Saxena K. M. L. (1981), Positive dependence in multivariate distributions, *Communications in Statistics: Theory and Methods*, **10**, 1183-1196.
- Antonini G. R., Kozachenko Y., and Volodin A. (2008), Convergence of series of dependent ϕ -subGaussian random variables, *Journal of Mathematical Analysis and Applications*, **338**, 1188-1203.
- Behboodian J. (1990), Some characterization theorems on symmetry, *Computational Statistics and Data Analysis*, **10**, 189-192.

- Choi J. Y. and Baek J. I. (2013), Exponential inequalities and complete convergence of extended acceptable random variables. *Journal of Applied Mathematics and Informatics*, **31**, 417-424.
- Christofides T. C., Fazekas I. and Hadjikyriakou M. (2016). Conditional acceptability of random variables, *Journal of Inequalities and Applications*, **149**.
<https://doi.org/10.1186/s13660-016-1093-1>.
- Farlie D. J. G. (1960), The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, **47**, 307-323.
- Fazekas I. and Pecsora S. (2017), General Bahr-Esseen inequalities and their applications, *Journal of Inequalities and Applications*, **191**. <https://doi.org/10.1186/s13660-017-1468-y>.
- Gumbel E. J. (1960), Bivariate exponential distributions, *Journal of American Statistical Association*, **55**, 698-707.
- Heathcote, C. R. and Pitman J. W. (1972), An inequality for characteristic functions, *Bulletin of the Australian Mathematical Society*, **16**, 1-9.
- Hsu P. L. (1951), Absolute moments and characteristic function, *Journal of Chinese Mathematical Society*, **1**, 259-280.
- Johnson N. L. and Kotz S. (1975), On some generalized farlie-gumbel-morgenstern distributions-II regression, correlation and further generalizations, *Communication in Statistics-Theory and Methods*, **6**, 485-496.
- Lehmann E. (1966), Some concepts of dependence, *The Annals of Mathematical Statistics*, **37**, 1137-1153.
- Morgenstern D. (1956), Einfache Beispiele zweidimensionaler Verteilungen, *Mitteilungsblatt für mathematische statistik, Würzburg*, **8**, 234-235.
- Ndiaye C. H. and Lo G. S. (2013), A note on a new exponential bound for M-acceptable random variables, *Afrika Statistika*, **8**, 575-581.
- Sung S. H., Srisuradetchai P. and Volodin A. (2011), A note on the exponential inequality for a class of dependent random variables, *Journal of the Korean Statistical Society*, **40**, 109-114.
- Ushakov N. G. (2011), Some inequalities for absolute moments, *Statistics and Probability Letters*, **81**, 2011-2015.
- Ushakov N. G. and Ushakov V. G. (2017), Permutation tests for homogeneity based on some characterizations, *Communication in Statistics: Theory and Methods*, **46**, 7692-7702.
- Von Bahr B. (1965), On the convergence of moments in the central limit theorem, *The Annals of Mathematical Statistics*, **36**, 808-818.
- Von Bahr D. and Esseen C. G. (1965), Inequalities for the r th absolute moment of a sum of random variables, *The Annals of Mathematical Statistics*, **36**, 299-303.
- Wang Y., Li Y. and Gao Q. (2011), On the exponential inequality for acceptable random variables, *Journal of Inequalities and Applications*, **40**. <https://doi.org/10.1186/1029-242X-2011-40>.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



An interactive R Shiny application for learning times series analysis

Motala Charles Frances^{1,*}, Mahdi Salehi^{2,†}, Andriette Bekker¹ and Mohammad Arashi³

¹Department of Statistics, University of Pretoria, Pretoria, South Africa

²Department of Mathematics and Statistics, University of Neyshabur, Neyshabur, Iran

³ Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran

Abstract

Many students and professionals find understanding time series analysis challenging because the time series concepts are too theoretical and require a strong mathematical background and reasoning skills. Thus, visualization of time series concepts or scenarios will aid in understanding the concepts. In this paper we discuss the development of the time series component within the [Advanced Modelling \(AM\) R Shiny Application](#). We make use of the Shiny web-based application framework R to develop the AM application, an interactive web application designed for learning multivariate analysis, machine learning, and time series modelling, with the main objective focused on time series. The time series component provides a user-friendly interface for individuals interested in analysing and gaining insights from the time series data. This component of the AM application incorporates interactive elements to facilitate a practical learning experience, making it particularly valuable for students and professionals seeking to enhance their understanding of the time series analysis concepts. We also explore the unique features of the time series component of the AM application, emphasizing its potential to support open learning and contribute to the advancement of knowledge. Finally, we will discuss the results of fitting the ARIMA model using the windspeed column from the air population data for Graz, Austria.

Keywords: ARIMA, Modelling, R, Shiny, Simulation, Time Series Modelling.

Mathematics Subject Classification (2020): xxAxx, xxBxx, xxCxx.

1 Introduction

In today's world, the field of statistical education is experiencing significant changes driven by the emergence of extensive datasets and advancements in artificial intelligence [Peng and Weng \(2022\)](#). In this evolving context, the teaching of statistics encounters unique opportunities and difficulties. The convergence of theoretical principles and practical application

*Corresponding author, francesmotala@gmail.com

†Speaker, salehi.sms@neyshabur.ac.ir

holds particular significance in shaping the education of undergraduate students majoring in statistics, equipping them to derive meaningful insights from the growing domain of data.

In this changing context, time series analysis stands out as a fundamental statistical approach, playing a crucial role in extracting insights from data points that are observed sequentially over time. Mathematically expressed as a sequence of observations $\{x_t\}_{t=-\infty}^{\infty}$, where the variable t represents the elapsed time [Cochrane \(1997\)](#); [Hipel and McLeod \(1994\)](#); [Raicharoen et al. \(2003\)](#). Time series analysis goes beyond specific fields, being used in economics, finance, epidemiology, and environmental science [Robert et al. \(2006\)](#). Its main objective is to improve our comprehension of temporal patterns, trends, and inherent structures within data, providing a systematic framework for investigating the dynamic nature of phenomena [Velicer and Fava \(2003\)](#).

However, delving into time series analysis can be challenging for both students and practitioners. Traditional teaching approaches that depend on textbooks and lectures may not effectively involve learners or offer hands-on experience with real-world datasets [Mustafa \(1996\)](#). The absence of user-friendly tools designed for time series analysis contributes to these difficulties. Acknowledging these challenges, there is a crucial demand for a creative and interactive educational resources that connect theoretical concepts with practical application. This demand is addressed by our time series component of the [AM application](#), a dynamic platform created to empower users in seamlessly simulating, analysing, and modelling time series datasets.

As an important tool in data science and research, time series analysis plays a significant role in making decisions based on evidence and enhancing understanding of temporal aspects in various phenomena (see for instance, [Salehi et al. \(2021\)](#)). This paper outlines the development and describes each function of the time series component of the AM application, with a focus on two key modules: time series simulations and time series modelling. Additionally, we conducted a systematic comparison between the AM application and the established time series analysis and visualization tool DATAEXPLOREFINES [Cruz et al. \(2023\)](#), aiming to evaluate and compare features to determine the strengths and weaknesses of each platform. Finally, we fit the ARIMA model to the column windspeed from the air population data for Graz, Austria spanning from 1 January 2014 to 5 March 2020 [Lovrić et al. \(2023\)](#). By tackling challenges associated with traditional teaching methods and offering a hands-on, interactive experience, this component of the AM application aims to contribute to the education of a new generation of analysts and researchers. Through adaptive educational approaches and an intuitive interface, the time series component of the application empowers users to navigate the complexities of time-varying data, fostering understanding that extends beyond disciplinary boundaries. The rest of this paper is organized as follows: in Section 1, we briefly introduce the R Shiny package. Section 3 introduces the time series module of the web-app developed. In Section 4, we illustrate the web-app by the monthly windspeed time series data. Finally, Section 5 concludes the paper.

2 R Shiny

R is a highly adaptable and robust tool for analysing and visualizing data. Shiny is a package within R, that simplifies the process of sharing interactive data analyses and graphics online [Beeley \(2016\)](#). R Shiny is a framework for building web-based data visualizations and dashboards using the R programming language [Sievert \(2020\)](#); [Wickham \(2021\)](#). It enables users to develop interactive applications without requiring expertise in web programming languages such as HTML, CSS, or JavaScript. With R Shiny, developers can use their R knowledge to craft dynamic and customizable web applications, allowing users to explore and interact with data in real-time.

The process of developing an R Shiny web application involves creating two main components: the user interface (UI) and the server logic. The UI is defined using R functions

that generate HTML and CSS code, specifying how the layout includes inputs (like sliders, dropdowns, or text fields) and outputs (such as tables, plots, or text) that users will interact with. The server logic on the other hand, written in R, manages data processing, computation, and the display of outputs based on user inputs. It utilizes reactive programming concepts, defining reactive expressions and values to establish connections between inputs and outputs. This enables the application to automatically update outputs when inputs change. The R Shiny application is then deployed on a web server, making it accessible to users through a web browser. The framework simplifies the creation of web applications with R, allowing researchers and data professionals to develop interactive tools for data analysis, visualization, exploration, and modelling. Figure 1 provides an illustration of the UI-Server model.

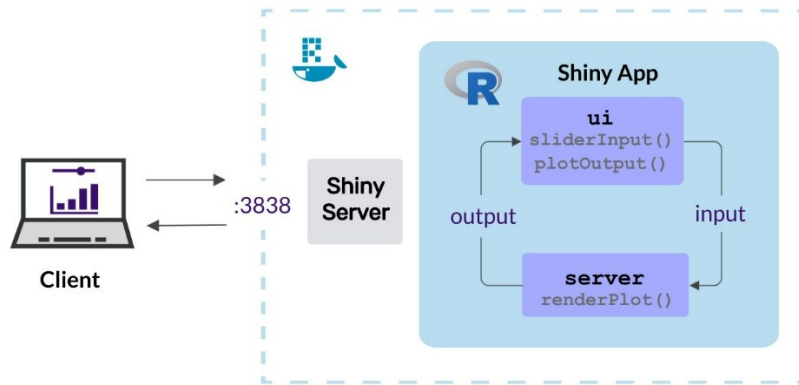


Figure 1: The R shiny web application UI-server model.

3 The Time Series Analysis R Shiny application

The time series component is a robust tool in the Advanced Modelling application. It consists of two modules: Time Series Simulations and Time Series Modelling. This analysis component offers end-users a platform to interact with, visualize, simulate, and model time series data based on their requirements. Figure 2 illustrates the workflow of these interactive modules, followed by the Advanced Modelling application interface in Figure 3.

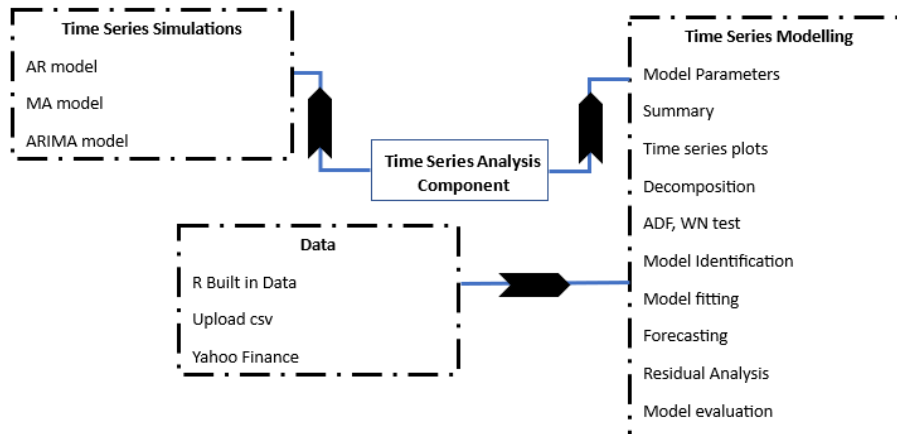


Figure 2: The Time Series Analysis component workflow.

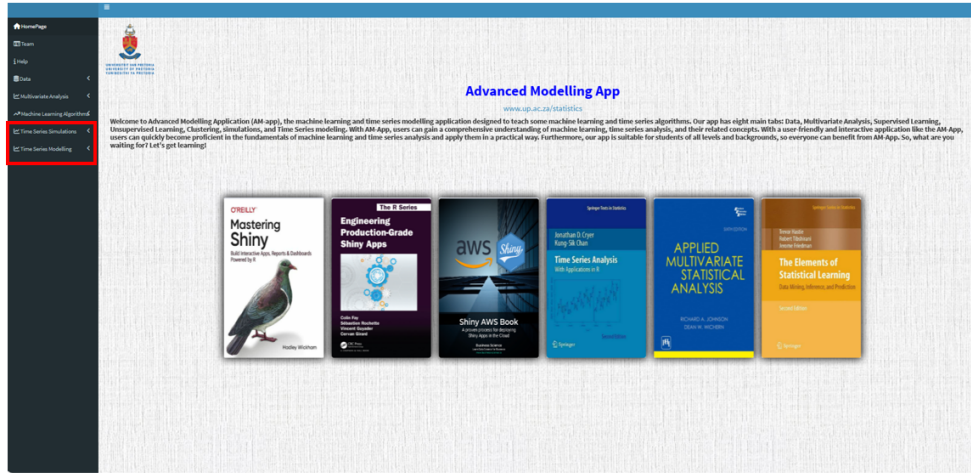


Figure 3: The Advanced Modelling Web-Based R Shiny application.

3.1 Data

Our application offers users with a versatile data analysis experience through three key features. Firstly, our end-users can easily import their own CSV files, ensuring a personalized analysis by seamlessly importing their datasets. Secondly, we offer our users with a diverse collection of pre-loaded datasets within the R Studio Programming Language, serving as valuable resources for users who do not have their own datasets. Finally, by integrating the 'quantmod' package, our users can access real-time or historical financial data from Yahoo Finance, facilitating in-depth financial time series analysis. This feature allows users to choose specific date ranges, scrape data for different stocks, and export the data in CSV, XLSX, or PDF format. Figure 4

refig:Yahoo provides an illustration of our user-friendly Yahoo Finance Data tab within the application.

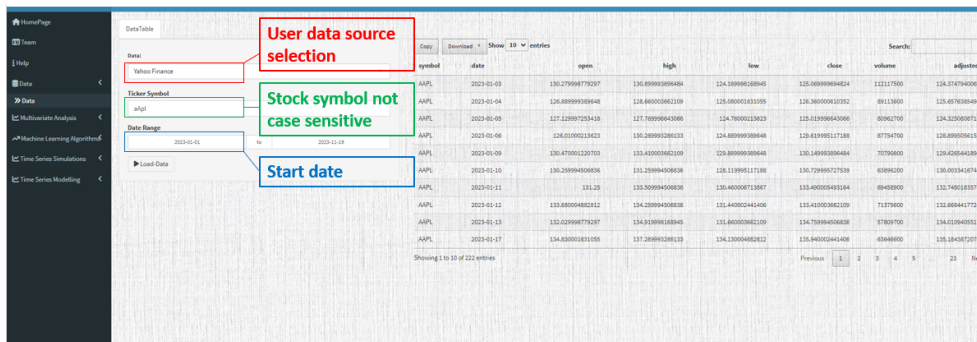


Figure 4: Yahoo Finance Data tab of the Advanced Modelling R Shiny application.

3.2 Module 1: Time Series Simulations

The Time Series Simulations module plays a crucial role in the AM application, this module offers a set of analytical tools to explore various time series models. This adaptable module is designed to help users with model validation, and forecasting using the time series simulated data. It consists of seven distinct tabs, each focusing on a specific time series model: white noise $a_t \sim N(0, \sigma_a^2)$, random walk $Z_t = Z_{t-1} + a_t$, autoregressive $Z_t = \theta_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t$, moving average $Z_t = \theta_0 + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$, autoregressive moving average $Z_t = \theta_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots + \theta_q a_{t-q}$,

autoregressive integrated $W_t = \theta_0 + \phi W_{t-1} + a_t$ where $W_t = Z_{t-1} - Z_t$ and $-1 < \phi < 1$, and integrated moving average $W_t = \theta_0 + a_t - \theta a_{t-1}$ where $W_t = Z_{t-1} - Z_t$ and $-1 < \theta < 1$. To further understand the definitions of parameters, please refer to the master dissertation paper (An interactive R shiny application for learning multivariate data analysis and time series modelling).

In Figure 5, we illustrate an example of 400 simulated random walk processes, each with a sample size of 150. In Figure 6, we present a simulation of a non-stationary $IMA(1,1)$ with parameters $\theta_0 = 0.2$, $\theta_1 = 0.6$, and $\sigma_a = 1$, represented as $Z_t - Z_{t-1} = \theta_0 - \theta_1 a_{t-1} + a_t$ where $a_t \sim N(0, \sigma_a^2)$.

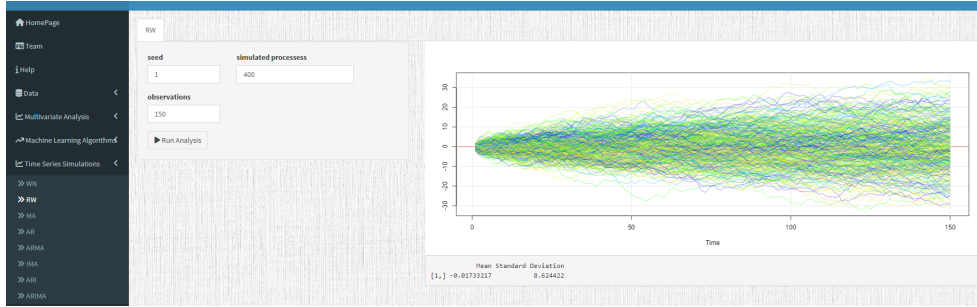


Figure 5: The Random Walk Simulation tab of the AM application.

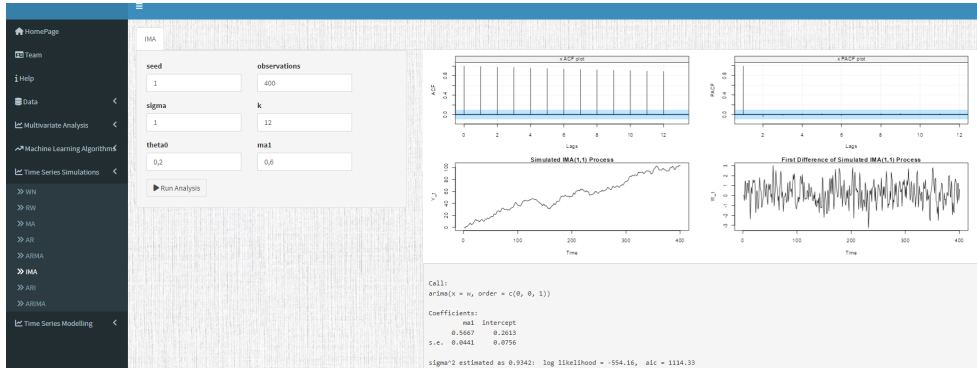


Figure 6: The Integrated Moving Average Simulation tab of the AM application.

3.3 Module 2: Time Series Modelling

The Time Series Modelling module is the final tool in our application, providing a set of analytical tools designed to fit time series models to the uploaded dataset. Through the use of the `auto.arima()` function, the module determines the best SARIMA model, allowing users to construct accurate models and forecast future values. Seasonal Autoregressive Integrated Moving Average (SARIMA), is a time series model that incorporates seasonality, trend, and noise components to make predictions. This versatile module assists users in constructing time series models and forecasting future values. The module includes ten user-friendly tabs, each serving a specific purpose: model parameters, summary statistics, time series plots, decomposition, tests, model identification, model fitting, forecasting, residual analysis, and model evaluation. Figures 9 and 10 illustrate the model fitting and residual analysis tabs, respectively.

4 ARIMA Modelling Results of the Monthly Wind-speed Time Series

The monthly windspeed time series data spanning from January 1, 2014, to March 15, 2020, was modeled using the Time Series Modelling component of the AM application. Figure 7 illustrates the windspeed series along with the series' Autocorrelation and Partial Autocorrelation plots. Figure 8 illustrates the Augmented Dickey-Fuller (ADF) and White Noise Tests tab of the AM application. Based on the ADF test with a p-value of 0.985, the windspeed series was found to be non-stationary. The Box-Ljung test, with a p-value less than 0.05, indicated that the observed time series is not a white noise.

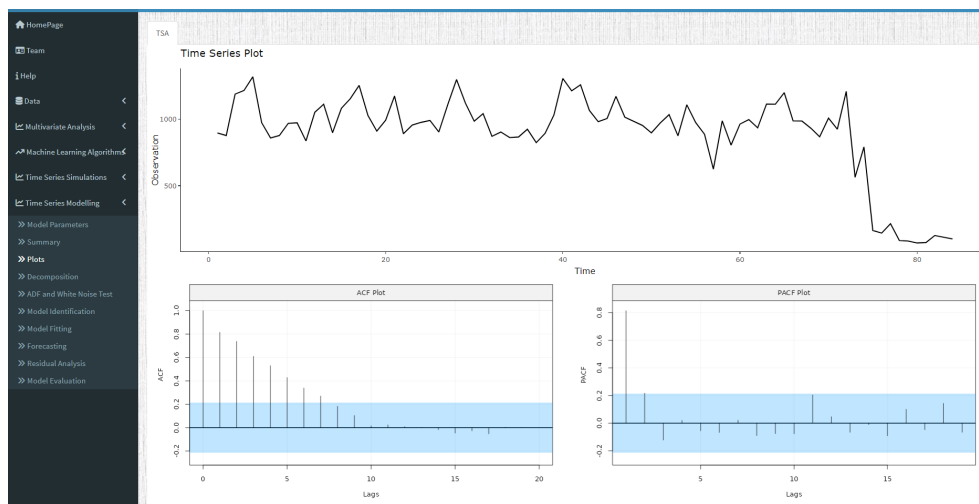


Figure 7: The Time Series plots tab of the Time Series Modelling application.

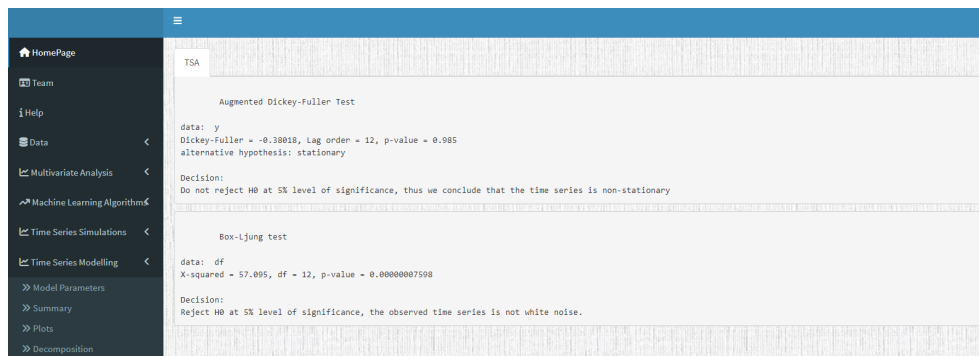


Figure 8: The Augmented Dickey-Fuller and White Noise Test tab of the Time Series Modelling application.

Utilizing the Time Series Modelling component, we fitted the auto ARIMA models and identified ARIMA(1,1,0) as the best model for the series, selected based on the minimum values of corrected AIC (1082.55) and BIC (1087.24). Figure 9 displays the original series values versus the fitted values over time. The adequacy of the ARIMA(1,1,0) model was verified using the Ljung-Box Q-statistics test, yielding a statistic of 5.4989 with a corresponding p-value of 0.7888. Since the p-value of the Ljung-Box statistics exceeded the critical value of 0.05, we rejected the null hypothesis of autocorrelation in the ARIMA(1,1,0) model residuals. This analysis demonstrated that the residuals were purely random, indicating no need for further modeling. Figure 10 presents the Residual Analysis of the ARIMA(1,1,0) model.

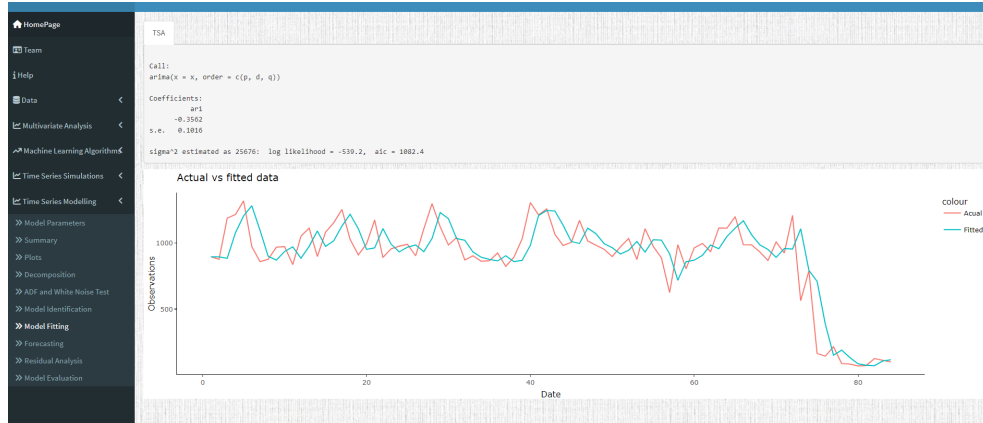


Figure 9: The original series values versus the fitted values of the monthly windspeed time series data.

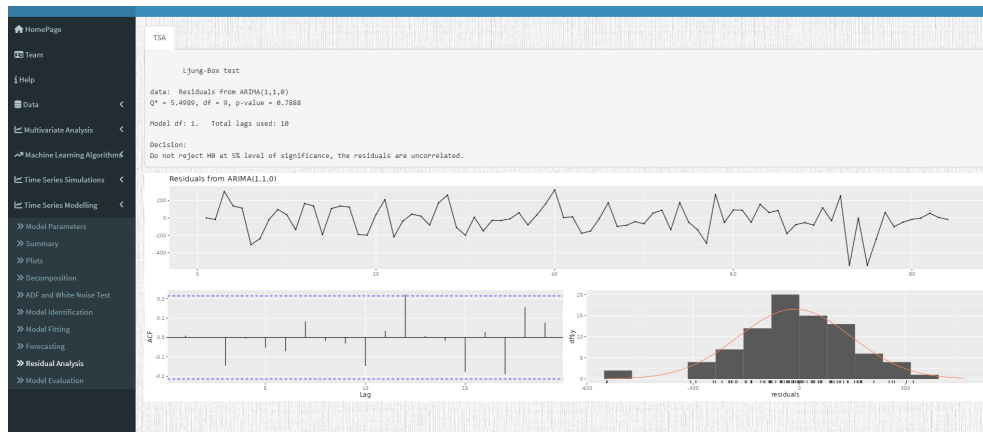


Figure 10: The Residual Analysis of the monthly windspeed time series data.

5 Conclusion

In conclusion, the development the time series component in the Advanced Modelling R Shiny Application is a significant achievement in building an interactive and user-friendly platform for learning time series analysis. Using the shiny framework, this web application takes advantage of R shiny, making it easy for users to access visualizations, simulations, and time series modelling through a friendly web interface. Demonstrating the ARIMA model on the windspeed data from the air population dataset of Graz, Austria, highlights the practical utility of the AM application in real-world scenarios. The AM application not only improves the overall user experience but also removes the need for users to install R on their local devices. As a result, this web-based model offers increased flexibility, allowing users to conveniently access the AM application via tablets or smartphones.

Acknowledgment

This work was based upon research supported in part by the National Research Foundation (NRF) of South Africa (SA), grant RA201125576565 & RA211204653274, nr 145681 & 151035; NRF ref. SRUG2204203865, the Centre of Excellence in Mathematical and Statistical Sciences,

based at the University of the Witwatersrand (SA). The opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF.

References

- Beeley C. (2016), *Web application development with R using Shiny*, Packt Publishing Ltd.
- Cochrane J. H. (1997), *Time series for macroeconomics and finance*, U. of Chicago.
- Cruz T., Jimenez F. G., Bravo A. R. Q., and Ander E. (2023), DataXploreFines: Generalized Data for Informed Decision Making, An Interactive Shiny Application for Data Analysis and Visualization, *arXiv preprint arXiv:2307.11056*.
- Hipel K. W. and McLeod A. I. (1994), *Time series modelling of water resources and environmental systems*, Elsevier.
- Lovrić M., Petrić V., Pavlović K., Schopper A., and Vuckovic M. (2023), Hourly air pollution data for Graz, Austria, [Data set]. Zenodo.
- Mustafa R. Y. (1996), The challenge of teaching statistics to non-specialists, *Journal of statistics education*, vol. 4, no. 1, Taylor & Francis.
- Peng S. and Weng Y. (2022), Development and Application of Experimental Platform for Time Series Analysis Course, In *2022 IEEE 13th International Conference on Software Engineering and Service Science (ICSESS)*, pp. 267–270, IEEE.
- Raicharoen T., Lursinsap C., and Sanguanbhokai P. (2003), Application of critical support vector machine to time series prediction, In *Proceedings of the 2003 International Symposium on Circuits and Systems, 2003. ISCAS'03.*, vol. 5, pp. V–V, IEEE.
- Robert H. and others (2006), *Time Series Analysis and Its Applications With R Examples Second Edition*, Springer.
- Salehi M., Arashi M., Bekker A., Ferreira J., Chen D.-G., Esmaili F., and Frances M. (2021), A Synergetic R-Shiny Portal for Modeling and Tracking of COVID-19 Data, *Frontiers in Public Health*, vol. 8, 623624.
- Sievert C. (2020), *Interactive web-based data visualization with R, plotly, and shiny*, CRC Press.
- Velicer W. F. and Fava J. L. (2003), Time series analysis, *Research methods in psychology*, vol. 2.
- Wickham H. (2021), *Mastering shiny*, O'Reilly Media, Inc.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



**Seminar On Probability
and Stochastic Processes**
August 30-31, 2025
University of Kurdistan



Milstein scheme for Numerical solution of first-order uncertain stochastic differential equations

Azadeh Ghasemifard *

Faculty of Science, University of Mazandaran

Abstract

Uncertain stochastic calculus is an emerging branch of mathematics that develops models integrating both aleatory (random) and epistemic (knowledge-based) uncertainties in dynamic systems. This field acknowledges two types of uncertainty in dynamical systems: randomness and belief degree. The uncertain stochastic differential equation (USDE) framework captures these systems by simultaneously accounting for randomness and human uncertainty, represented by belief degree. This evolving area has led to the introduction of a new class of equations known as USDEs. Since obtaining exact or analytical solutions for these equations is often challenging, numerical methods offer a valuable alternative for approximating solutions in such cases. This paper explores the application of the Milstein method for solving USDEs. The Milstein scheme is applied to solve a stock pricing problem, with its results compared to those obtained using the fourth-order Runge-Kutta and Euler methods. The findings show that the Milstein method produces more accurate stock price estimates.

*a.ghasemi@umz.ac.ir



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Estimation of the Intensity Function of Point Processes Using Soap Film Bases

Mohammad Ghorbani^{*1}

¹Department of Engineering Sciences and Mathematics, Luleå University of Technology, Sweden

Abstract

Kernel intensity estimation can sometimes yield unsatisfactory results when used to estimate the intensity of a spatial point process observed on the entire observation window or on a window with complex boundaries, as it inappropriately smooths the boundary features. In this work, we use the soap film smoothing method for intensity estimation, which is suitable for smoothing over complicated regions. The idea is to minimize an objective function inspired by the Campbell formula penalized by one or two quadratic penalties and to represent the optimization problem in terms of a generalized additive model to allow efficient computation with a k-fold cross-validation method. We investigate the performance of the new estimator on simulated data and apply it to estimate the intensity function of real datasets.

Keywords: Generalized additive model; Kernel smoothing; Quadratic penalty; Soap film smoothing; Spatial point process.

Mathematics Subject Classification (2020): 62M30, 62H11, 62J12.

^{*}Corresponding author, mohammad.ghorbani@ltu.se



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Kolmogorov-Arnold Networks (KAN): Revolutionizing Prediction in Machine Learning

Hossein Haghbin*

Department of Statistics, Faculty of Intelligent Systems Engineering and Data Sciences, Persian Gulf University, Iran

Abstract

Kolmogorov–Arnold Networks (KANs) are a novel class of neural networks inspired by the Kolmogorov–Arnold representation theorem. By replacing traditional weight matrices with learnable univariate function modules, KANs provide a fundamentally different approach to modeling multivariate functions. This paper introduces the core concepts behind KANs, compares their structure and performance to traditional multilayer perceptrons (MLPs), and demonstrates their advantages in both prediction tasks.

Keywords: Kolmogorov-Arnold Networks, Nonlinear Regression, Function Approximation.

Mathematics Subject Classification (2020): 93E35, 60J20, 68W20.

1 Introduction

In recent years, deep learning has achieved remarkable success across a wide range of applications, including computer vision, natural language processing, and scientific computing. Classical neural architectures—such as multilayer perceptrons (MLPs), convolutional neural networks (CNNs), and transformers—typically rely on compositions of affine transformations and pointwise nonlinear activation functions. While effective, these architectures often act as black boxes and require large parameter counts and extensive data to perform well.

A foundational theoretical result underlying modern neural networks is the *Universal Approximation Theorem*, which states that a feedforward neural network with a single hidden layer containing a finite number of neurons can approximate any continuous function on a compact domain, given a suitable activation function [Cybenko \(1989\)](#); [Hornik \(1991\)](#). While this theorem establishes the expressive power of MLPs in theory, in practice, the required number of neurons may be very large, and the learned representations often lack interpretability.

On the other hand, the *Kolmogorov–Arnold Representation Theorem* provides a constructive way to represent any multivariate continuous function as a finite sum of univariate functions composed with addition [Kolmogorov \(1957\)](#). This formulation has inspired alternative architectures to MLPs, such as *Kolmogorov–Arnold Networks (KANs)*, recently introduced by

*Corresponding author, haghbin@pgu.ac.ir

Liu et al. [Liu et al. \(2024\)](#). Unlike MLPs, which use scalar weights and fixed activation functions, KANs leverage adaptive univariate functions—typically modeled as splines—to build each layer.

KANs bring multiple advantages: they align with a solid mathematical foundation, enhance interpretability, and often require fewer parameters for competitive performance, especially in low-data regimes. Related studies have also explored spline-based models and adaptive activations, but KANs offer a unified and principled approach that integrates these ideas into the core of the neural architecture.

This paper aims to:

- Provide a formal introduction to Kolmogorov–Arnold Networks (KANs),
- Contrast their structure and training mechanisms with traditional MLPs,
- Demonstrate their empirical performance in regression and classification tasks, and
- Discuss the interpretability and theoretical implications of such architectures.

2 Foundations and Motivation

One of the foundational results supporting the use of neural networks for function approximation is the Universal Approximation Theorem (UAT). This theorem formalizes the expressive power of feedforward neural networks, particularly multilayer perceptrons (MLPs).

Theorem 2.1 (Universal Approximation Theorem [Hornik \(1991\)](#)). *Let $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ be a nonconstant, bounded, and continuous activation function. Then, for any continuous function $f : [0, 1]^n \rightarrow \mathbb{R}$ and any $\epsilon > 0$, there exists a feedforward neural network with a single hidden layer of the form*

$$F(\mathbf{x}) = \sum_{i=1}^N \alpha_i \sigma(\mathbf{w}_i^\top \mathbf{x} + b_i),$$

such that

$$\sup_{\mathbf{x} \in [0, 1]^n} |F(\mathbf{x}) - f(\mathbf{x})| < \epsilon.$$

This theorem established that MLPs, even with just one hidden layer, are capable of approximating any continuous function on compact subsets of \mathbb{R}^n , provided that the width of the hidden layer is sufficiently large. However, in practice, deeper architectures with multiple hidden layers are often used to improve learning efficiency and generalization. Despite their universal expressiveness, MLPs typically rely on a composition of affine transformations and nonlinear activations, which may lead to difficulties in capturing localized or highly structured features in the data.

Although MLPs are theoretically universal approximators, their practical application faces challenges related to interpretability, training stability, and sample efficiency. Their reliance on global basis functions (like ReLU or sigmoid) can lead to inefficient parameter usage and poor local generalization. These limitations motivate the exploration of alternative architectures grounded in more structured mathematical representations of functions.

3 Kolmogorov–Arnold Networks (KANs)

The Kolmogorov–Arnold Representation Theorem provides a constructive guarantee that any multivariate continuous function can be decomposed into a finite composition of continuous univariate functions and additions. Formally, for any continuous function $f : [0, 1]^n \rightarrow \mathbb{R}$, there exist continuous univariate functions ϕ_q and continuous functions ψ_p , such that:

$$f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} \phi_q \left(\sum_{p=1}^n \psi_p^{(q)}(x_p) \right).$$

This result suggests that complex multivariate mappings do not inherently require multivariate nonlinearities, but instead can be constructed through composition and summation of univariate transformations. While the original proof is non-constructive and impractical for numerical purposes, it forms the theoretical backbone of Kolmogorov–Arnold Networks (KANs).

KANs operationalize this representation in a learnable framework by replacing traditional weight matrices and fixed activation functions used in Multilayer Perceptrons (MLPs) with learnable univariate functions. Each unit in a KAN layer performs a sum of such univariate functions applied to the individual coordinates of the input vector. Specifically, the j -th output of layer ℓ is computed as:

$$\Phi_\ell(\mathbf{x})_j = \sum_{i=1}^{d_{\ell-1}} \phi_{\ell,j,i}(x_i),$$

where each $\phi_{\ell,j,i} : \mathbb{R} \rightarrow \mathbb{R}$ is a learnable function, commonly parameterized using B-splines or Fourier series. The entire network is constructed as a composition of such layers:

$$\text{KAN}(\mathbf{x}) = (\Phi_L \circ \Phi_{L-1} \circ \cdots \circ \Phi_1)(\mathbf{x}),$$

where L is the depth of the network.

To contrast this with traditional MLPs, consider a deep L -layer MLP defined by:

$$\text{MLP}(\mathbf{x}) = (\mathbf{W}_L \circ \sigma_{L-1} \circ \mathbf{W}_{L-1} \circ \cdots \circ \sigma_1 \circ \mathbf{W}_1)(\mathbf{x}),$$

where each \mathbf{W}_ℓ is a learnable weight matrix and σ_ℓ is a fixed elementwise activation function (e.g., ReLU or GELU).

In MLPs, expressive power arises primarily through the combination of linear transformations and fixed nonlinearities. In contrast, KANs achieve expressivity through flexible, data-driven univariate functions and additive combinations. This not only enables improved approximation fidelity but also offers enhanced interpretability by isolating nonlinear transformations at a per-dimension level.

4 Simulation Study: Nonlinear Regression

To evaluate the functional approximation capabilities of Kolmogorov–Arnold Networks (KANs) in comparison with conventional Multi-Layer Perceptrons (MLPs), we conducted a comprehensive simulation study on a synthetic nonlinear regression problem. The ground truth function is defined as a nonlinear interaction between two input variables:

$$y = \sin(3x_1) \cdot \cos(2x_2) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, 0.05^2),$$

where $x_1, x_2 \sim \text{Uniform}(-1, 1)$. A total of 5,000 samples were generated, with 80% allocated for training and 20% for testing. All input features were standardized to zero mean and unit variance prior to model fitting.

Two model families KANs and MLPs were considered. Each was evaluated in both shallow and deep configurations. Shallow architectures consisted of a single hidden layer, whereas deep models incorporated two or more hidden layers. The width of each hidden layer was systematically varied to examine the trade-offs between model capacity and performance.

Model performance was assessed on the test set using Mean Squared Error (MSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2) and the total number of trainable parameters (# Par.) metrics. The numerical results of the simulation are presented in Table 1. Across both shallow and deep configurations, KANs consistently outperform MLPs in terms of MSE and R^2 , often with comparable or fewer trainable parameters. Notably, the shallow KAN model with structure $[2, 3, 1]$ achieves a remarkably low MSE of 0.0082 and an R^2 of

Table 1: Comparison of KAN and MLP on Nonlinear Regression Task

Type	Model	Structure	# Par.	MSE	MAE	R^2
Shallow	KAN	[2,2,1]	84	0.0238	0.1194	0.9727
		[2,3,1]	126	0.0082	0.0715	0.9906
		[2,4,1]	168	0.0107	0.0764	0.9878
	MLP	[2,16,1]	65	0.2212	0.3770	0.7460
		[2,32,1]	129	0.0484	0.1642	0.9444
		[2,64,1]	257	0.0405	0.1448	0.9535
Deep	KAN	[2,3,2,1]	196	0.0220	0.1133	0.9748
		[2,4,2,1]	193	0.0230	0.1100	0.9735
		[2,3,4,1]	252	0.1409	0.2864	0.8382
	MLP	[2,16,8,1]	335	0.0113	0.0822	0.9870
		[2,16,16,1]	308	0.0177	0.1001	0.9797
		[2,32,16,1]	369	0.0133	0.0812	0.9847

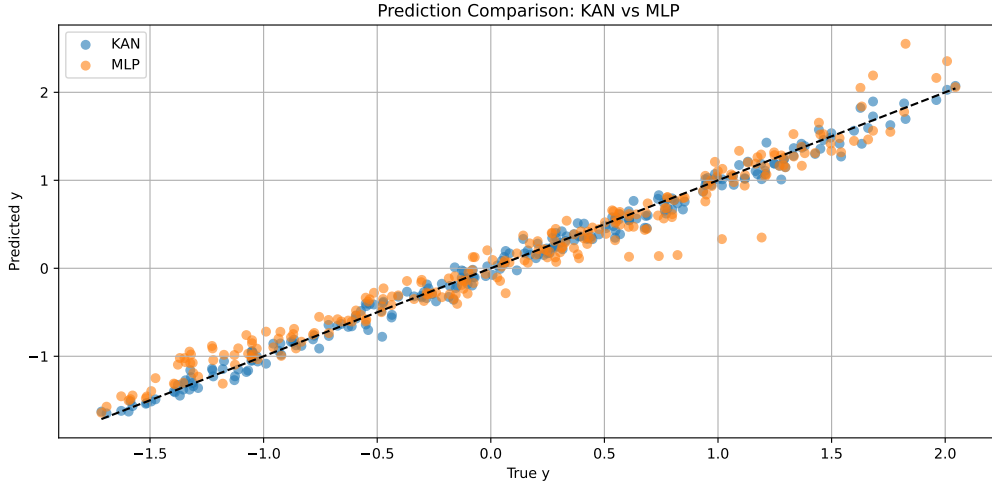


Figure 1: Prediction Comparison: KAN[2,3,1] vs MLP[2,64,1]

0.9906, despite having fewer parameters than its MLP counterpart [2, 64, 1], which attains an MSE of 0.0405 and R^2 of 0.9535.

To complement the quantitative analysis, Figure 1 shows scatter plots comparing the predicted and actual values for the best-performing shallow models: KAN [2, 3, 1] and MLP [2, 64, 1]. The predictions from the KAN model align more tightly with the diagonal reference line, demonstrating better adherence to the true output values. This result corroborates the superior regression accuracy of KAN, even with a significantly smaller model size.

In addition, Figure 2 presents surface plots of the true function alongside the outputs produced by the same two models across a uniform grid of input values. The KAN surface accurately reconstructs the complex nonlinear patterns of the target function, maintaining smooth transitions and capturing fine-grained structure. In contrast, the MLP surface exhibits oversmoothing and visible approximation errors, particularly in boundary regions. This visual evidence reinforces the claim that KANs can achieve high-fidelity approximations with fewer parameters and better generalization.

The findings of this study underscore the expressive efficiency of Kolmogorov–Arnold Networks for smooth nonlinear regression. KANs consistently outperform MLPs in shallow regimes, offering superior accuracy with lower model complexity. While MLPs begin to

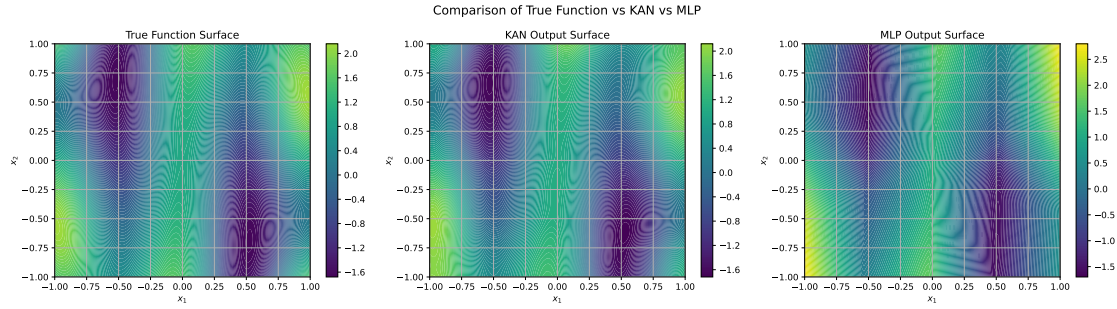


Figure 2: Surface Comparison of True Function, KAN[2,3,1], and MLP[2,64,1] Outputs

close the performance gap in deeper and wider configurations, they require significantly more parameters to do so. These results highlight the potential of KANs as a compact and interpretable alternative to traditional feedforward architectures, particularly in applications where parameter efficiency and functional interpretability are critical.

References

- Cybenko, G. (1989), Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, **2**, 303–314.
- Hornik, K. (1991), Approximation capabilities of multilayer feedforward networks. *Neural Networks*, **4**(2), 251–257.
- Kolmogorov, A. N. (1957), On the representation of continuous functions of many variables by superpositions of continuous functions of one variable and addition. *Doklady Akademii Nauk SSSR*, **114**, 953–956.
- Liu, X., et al. (2024), Kolmogorov–Arnold Networks. *arXiv preprint*, arXiv:2402.12782. <https://arxiv.org/abs/2402.12782>



Goodness of Fit Test Based on Extropy and Dirichlet Process

Shamim Haghshenas Jazi^{*1}, Zahra Saberi^{†1}

¹Department of Mathematical Sciences, Isfahan University of Technology, Isfahan, Iran

Abstract

In this paper, we propose a novel goodness of fit test based on extropy from a Bayesian nonparametric perspective. Extropy, introduced as a complement to entropy, focuses on the structured concentration within a probability distribution rather than on uncertainty. To enhance estimation accuracy, especially with small samples, we employ a Bayesian estimator built upon the Dirichlet process prior. We define a relative extropy as a measure of distance between the empirical distribution and a hypothesized model, and construct a testing framework based on the Relative Belief Ratio (RB) and Strength indices to quantify evidence for or against the null hypothesis. Some simulation studies have been done in order to investigate the proposed approach.

Keywords: Bayesian Nonparametric Methods, Dirichlet Process, Extropy, Relative Belief Ratio, Strength.

Mathematics Subject Classification (2020): 62F15, 62G10, 94A17.

1 Introduction

Assessing the fit between observed data and theoretical models is a cornerstone of statistical analysis. Traditional measures such as entropy have long served to quantify uncertainty. [Grzegorzewski and Wieczorkowski \(1999\)](#) introduced an entropy-based goodness-of-fit test, which has inspired further exploration of uncertainty measures in model assessment.

In 2015, [Lad and Sanfilippo and Agro \(2015\)](#) introduced extropy, a complementary concept to entropy that emphasizes the structured concentration within probability distributions. Following this, [Qiu and Jia \(2018\)](#) developed nonparametric extropy estimators using order statistics.

Although entropy-based goodness-of-fit methods have been well explored, extropy, as its complementary concept, has received comparatively less attention in the context of Bayesian inference. [Lad and Sanfilippo and Agro \(2015\)](#) proposed foundational ideas for its application. [Qiu and Jia \(2018\)](#) expanded on this by developing robust nonparametric estimators. [Evans \(2015\)](#) and [Al-Labadi and Evans \(2018\)](#) presented Bayesian model assessment techniques, which can be extended to incorporate extropy through the Relative Belief framework.

^{*}M.S in Statistics, shamim.haghshenas78@gmail.com

[†]Assistant Professor of Statistics, z_saberi@iut.ac.ir

Meanwhile, classical goodness-of-fit tests like Kolmogorov-Smirnov and Anderson-Darling remain widely used but suffer from limited sensitivity to various types of deviations [D’Agostino and Stephens \(1986\)](#). This has motivated the search for more flexible, robust alternatives.

To provide a Bayesian nonparametric framework for goodness-of-fit testing, one can combine extropy estimation with the Dirichlet process prior [Ferguson \(1973\)](#). This approach uses relative extropy as a distance metric and incorporates the Relative Belief Ratio and its calibration measure, Strength, as discussed in the relative belief framework [Evans \(2015\)](#); [Al-Labadi and Evans \(2018\)](#).

The remainder of the paper is organized as follows. Section 2 introduces the Bayesian nonparametric estimator of extropy based on the Dirichlet process. Section 3 presents the goodness-of-fit testing framework using the proposed estimator and the Relative Belief approach. Section 4 includes simulation studies that evaluate the performance of the method. In Section 5, we examine the sensitivity of the results to prior parameters, and Section 6 discusses computational aspects. Finally, Section 7 concludes the paper.

2 Extropy and Bayesian Nonparametric Estimation

Extropy is a measure of uncertainty that serves as a complementary of entropy [Lad and Sanfilippo and Agro \(2015\)](#). For a random variable with density function $f(x)$, it is defined as:

$$J(f) = -\frac{1}{2} \int f^2(x) dx. \quad (1)$$

Unlike entropy, which focuses on uncertainty, extropy highlights how a distribution organizes its probability mass. Several estimators [Qiu and Jia \(2018\)](#) have been proposed for estimating extropy, and equation (2) can be rewritten as:

$$J(f) = -\frac{1}{2} \int_0^1 \left[\frac{d}{dt} F^{-1}(t) \right]^{-1} dt. \quad (2)$$

Now, if $\mathbf{X} = (X_1, \dots, X_n)$ is a random sample from F and $X_{(i)}$ ’s, $i = 1, \dots, n$, are the order statistics according to \mathbf{X} , then (2) can be approximated as follow:

$$JQ = -\frac{1}{2n} \sum_{i=1}^n \frac{c_i \cdot m/n}{X_{(i+m)} - X_{(i-m)}}, \quad (3)$$

where m is the window size (typically proportional to \sqrt{n}), and c_i is the boundary correction coefficient defined as:

$$c_i = \begin{cases} 1 + \frac{i-1}{m}, & \text{if } 1 \leq i \leq m, \\ 2, & \text{if } m+1 \leq i \leq n-m, \\ 1 + \frac{n-i}{m}, & \text{if } n-m+1 \leq i \leq n. \end{cases}$$

For other non-Bayesian estimation of entropy, refer to [Noughabi and Jarrahiferiz \(2022\)](#) and [Källberg and Seleznev \(2016\)](#).

2.1 Nonparametric Bayesian Estimation for Extropy

For Bayesian nonparametric estimation of extropy, we use the Dirichlet prior. Let $DP(\alpha, G)$ be the Dirichlet process with a positive real number α and a fixed probability measure G [Ferguson \(1973\)](#).

The Dirichlet process has a conjugacy property: if the prior on F is $\text{DP}(\alpha, G)$, then after observing a sample \mathbf{x} , the posterior becomes $\text{DP}(\alpha + n, G_{\alpha, n})$, where $G_{\alpha, n}$ is a weighted average of G and the empirical distribution of the data.

By the conjugacy property of the Dirichlet process, the posterior distribution of F given an observed sample \mathbf{x} is $\text{DP}(\alpha + n, G_{\alpha, n})$.

Now, let $F \sim \text{DP}(\alpha, G)$. The Dirichlet process has a conjugacy property: if the prior on F is $\text{DP}(\alpha, G)$, then after observing a sample \mathbf{x} , the posterior becomes $\text{DP}(\alpha + n, G_{\alpha, n})$, where $G_{\alpha, n}$ is a weighted average of G and the empirical distribution of the data, where

$$G_{\alpha, n} = \frac{\alpha}{\alpha + n} G + \frac{n}{\alpha + n} F_n$$

and F_n is the empirical cumulative distribution function (ECDF) of the observed sample x . By using the properties of the Dirichlet process, if $P \sim \text{DP}(\alpha, G)$, then we have the following representation:

$$P = \frac{\sum_{i=1}^{\infty} L^{-1}(\Gamma_i) \delta_{Y_i}}{\sum_{i=1}^{\infty} L^{-1}(\Gamma_i)},$$

where $\Gamma_i = E_1 + \dots + E_i$; $E_i \stackrel{i.i.d.}{\sim} \text{Exp}(1)$, $Y_i \stackrel{i.i.d.}{\sim} G$, independent of Γ_i . Define the function:

$$L(x) = a \int_x^{\infty} \frac{1}{t} e^{-t} dt, \quad x > 0,$$

with its inverse given by:

$$L^{-1}(y) = \inf\{x > 0 \mid L(x) \leq y\}.$$

Now, by this property, we can estimate $(\frac{d}{dt} F^{-1}(t))^{-1}$ by the inverse of the slope of the straight line connecting the following points:

$$\left(P_N(Y_{(i-m)}) = \sum_{k=1}^{i-m} P_{k,N}, Y_{(i-m)} \right) \quad \text{and} \quad \left(P_N(Y_{(i+m)}) = \sum_{k=1}^{i+m} P_{k,N}, Y_{(i+m)} \right).$$

This equal to:

$$\frac{C_{i,a}}{Y_{(i+m)} - Y_{(i-m)}},$$

where

$$c_{i,a} = \begin{cases} \sum_{k=2}^{i+m} p_{k,N} & 1 \leq i \leq m, \\ \sum_{k=i-m+1}^{i+m} p_{k,N} & m+1 \leq i \leq N-m, \\ \sum_{k=i-m+1}^N p_{k,N} & N-m+1 \leq i \leq N. \end{cases}$$

Note that $E(C_{i,a}) = \frac{m}{N} c_i$, where c_i is defined in (3).

3 Goodness of Fit Testing

Let $\mathbf{x} = (x_1, \dots, x_n)$ be a random sample from distribution F . The main goal is to test the hypothesis $H_0 : F = F_0$, where F_0 is a specified distribution. Now let $F \sim \text{DP}(\alpha, G)$ be the prior on F , for some choice of α and base measure G . Then the posterior distribution is:

$$F \mid \mathbf{x} \sim \text{DP}(\alpha + n, G_{\mathbf{x}})$$

If F_0 is true, then the posterior distribution of the distance d between F and F_0 must be more concentrated around 0 than the prior distribution of this distance.

As a distance for this purpose, we consider relative entropy distance [Ferguson \(1973\)](#), which is defined as follows:

$$d(f||g) = \frac{1}{2} \int (f - g)^2 dx.$$

A Bayesian nonparametric estimation for d is given by:

$$\hat{d} = \hat{d}(f||g) = -J_{m,N,a}(f) - J(g) - \sum_{i=1}^N p_{i,N} g(Y_i),$$

where $J_{m,N,a}(f)$ is the Bayesian estimator of the entropy of f , and $J(g) = -\frac{1}{2} \int g^2(x) dx$ is the entropy of the hypothesized model g .

We can show that if $N \rightarrow \infty$, $m \rightarrow \infty$, $\frac{m}{N} \rightarrow 0$, and $a \rightarrow \infty$, then $\hat{d} \xrightarrow{P} d(f||g)$.

3.1 Relative Belief Ratio

Let $\mathbf{x} = (x_1, \dots, x_n)$ be a sample of $f(x|\theta)$, where $\theta \sim \pi(\theta)$. For a hypothesis testing $H_0 : \theta = \theta_0$, the relative belief ratio is defined as:

$$RB(\theta_0 | \mathbf{x}) = \frac{\pi(\theta_0 | \mathbf{x})}{\pi(\theta_0)}.$$

This is measuring how our belief changed that θ_0 is the true value from a prior to a posterior and If $RB > 1$, there is evidence in favor of θ_0 ; if $RB < 1$, there is evidence against θ_0 and if $RB = 1$, there is no evidence either way.

For calibrating this measure, [Evans \(2015\)](#) suggested to computing the posterior probability as bellow and named it *strength*;

$$\text{Strength} = \Pi(RB(\theta|x) \leq RB(\theta_0|x) | x),$$

Interpretation of the Strength

We distinguish two main cases:

Case 1: $RB_\theta(\theta_0 | x) > 1$ (data support H_0)

- If $Str_\theta(\theta_0)$ is large, it means that there is a small chance that the true value of θ corresponds to a higher relative belief ratio. Thus, there is strong evidence in favor of H_0 .
- If $Str_\theta(\theta_0)$ is small, it indicates a high probability that the true θ value has a larger relative belief ratio than θ_0 . Consequently, $RB_\theta(\theta_0 | x)$ only provides weak evidence in favor of H_0 .

Case 2: $RB_\theta(\theta_0 | x) < 1$ (data are against H_0)

- If $Str_\theta(\theta_0)$ is small, there is strong evidence against H_0 , as the true value of θ likely has a higher relative belief ratio than θ_0 .
- If $Str_\theta(\theta_0)$ is large, only weak evidence against H_0 is provided by $RB_\theta(\theta_0 | x)$.

Classification of the Strength of $RB_\theta(\theta_0 | x)$

Based on the value of $Str_\theta(\theta_0)$, we categorize the strength into three levels:

- If $Str_\theta(\theta_0) \leq 0.25$, the evidence is considered weak.
- If $0.25 < Str_\theta(\theta_0) \leq 0.75$, the evidence is considered moderate.
- If $0.75 < Str_\theta(\theta_0) \leq 1$, the evidence is considered strong.

It is important to note that although $Str_\theta(\theta_0)$ shares some similarities with a p-value, it is not intended for making direct acceptance or rejection decisions for H_0 . Instead, it measures the strength of the relative belief ratio for evaluating H_0 .

4 Simulation Studies

4.1 Simulation Study 1

To evaluate the performance of the proposed estimator, we conducted several simulation studies and practical examples.

In the simulation studies, both Bayesian and non-Bayesian estimators are compared for different sample sizes $n \in \{10, 20, 50\}$ and window sizes $m \in \{2, 3, 4, 7\}$. The distributions considered are:

- Exponential $\text{Exp}(1)$ with true entropy -0.25 ,
- Normal $N(0, 1)$ with true entropy $-\frac{1}{2\sqrt{\pi}} \approx -0.141$.

For each case, 5000 samples were generated and the entropy was estimated for each sample. The average estimate (Est) and the mean squared error (MSE) were then calculated as:

$$MSE = \frac{1}{5000} \sum_{i=1}^{5000} (\text{Estimate}_i - \text{True Value})^2$$

All computations were carried out in R. The results are reported in Tables 1–2.

According to the simulation results, we conclude that the new approximation is more accurate in estimating entropy.

Table 1: Results for Exponential Distribution with Mean 1

$JQ_{m,n}$	$J_{m,N,a} X$	m	n
Est(MSE)	Est(MSE)		
−0.2117(0.0097)	−0.5548(13.8784)	3	10
−0.1608(0.0101)	−0.3703(0.7433)	4	20
−0.1868(0.0050)	−0.2331(0.2397)	7	50

True value: -0.25

Table 2: Results for Normal Distribution $N(0, 1)$

$JQ_{m,n}$	$J_{m,N,a} X$	m	n
Est(MSE)	Est(MSE)		
−0.1519(0.0023)	−0.2942(10.6124)	3	10
−0.1098(0.0015)	−0.1997(0.3497)	4	20
−0.1234(0.0005)	−0.1197(0.0694)	7	50

True value: ≈ -0.1410

Analysis of Table 3 Table 3 shows entropy estimates for different base distributions H and two values of the concentration parameter a .

When $a = 0.05$, the estimates are similar across different H , suggesting low sensitivity to the prior. However, with $a = 5$, the estimates vary more, indicating that the prior has a stronger influence.

Thus, to obtain estimates that rely more on the observed data, a smaller value of a is recommended.

Table 3: Effect of Parameters a and H on the Estimation

Estimate: $a = 5$	Estimate: $a = 0.05$	Base Distribution H
-0.3520	-0.1706	$N(0, 1)$
-0.2138	-0.1688	$N(3, 9)$
-0.2886	-0.1654	t_1
-0.3715	-0.1816	$\text{Exp}(1)$

4.2 Simulation Study 2

For computing relative belief ratio, we must compute prior and posterior densities of $D = d(f||f_0)$, which is not available. So, we need to approximate them via simulations. A detailed computational algorithm Al-Labadi and Evans (2018) for assessing $H_0 : F = F_0$ as follows:

Algorithm A (Computing Relative Belief Ratio Algorithm for Goodness of Fit Tests):

1. Generate a sample of size r_1 from P_N , where P_N is an approximate for $DP(\alpha, F_0)$, then compute $d(\text{pr}) = \hat{d}$ as described in the text.
2. Generate a sample of size r_2 from the approximate of $P_{N|x}$ ($DP(a + n, H_x)$), then compute $d(\text{po}) = \hat{d}|x$
3. For a fixed positive number M , let \hat{F}_D denote the empirical cdf of D from the prior sample. For $i = 0, \dots, M$, $\hat{d}_{i/M(\text{pr})}$ is the estimate of the i/M -th prior quantile of D , where $\hat{d}_0(\text{pr}) = 0$ and $\hat{d}_1(\text{pr})$ is the largest value of $d(\text{pr})$.
4. Let $\hat{F}_D(\cdot|x)$ denote the empirical cdf of D based on the posterior sample. For $d \in [\hat{d}_{i/M(\text{pr})}, \hat{d}_{(i+1)/M(\text{pr})}]$, estimate $RBD(d|x)$ as the ratio of the posterior and prior cdf differences in this interval:

$$R\hat{B}D(d|x) = M \left(\hat{F}_D(\hat{d}_{(i+1)/M(\text{pr})}|x) - \hat{F}_D(\hat{d}_{i/M(\text{pr})}|x) \right).$$

5. Estimate $RBD(0|x)$ as:

$$R\hat{B}D(0|x) = M\hat{F}_D(\hat{d}_{i_0/M(\text{pr})}|x),$$

where i_0 is chosen so that i_0/M is not too small (typically $i_0/M \approx 0.05$).

6. Finally, estimate the strength $DPD(RBD(d|x) \leq RBD(0|x)|x)$ using the following summation:

$$\sum_{i \geq i_0: R\hat{B}D(\hat{d}_{i/M(\text{pr})}|x) \leq R\hat{B}D(0|x)} \left(\hat{F}_D(\hat{d}_{(i+1)/M(\text{pr})}|x) - \hat{F}_D(\hat{d}_{i/M(\text{pr})}|x) \right).$$

In the following, the proposed method is evaluated using an example with parameters set in Algorithm A: $r = 5000$, $N = 500$, $M = 20$, and $i_0 = 1$. The sensitivity to the parameter a is examined, with selected values of a reported in the table. A sample of size $n = 25$ is drawn from the standard normal distribution, and the method is used to test the null hypothesis $H_0 : F = F_0$, as detailed in Table 4. The results for the Relative Belief Ratio (RB) and Strength are shown. When H_0 is true, we expect $RB > 1$ and Strength near 1; when H_0 is false, $RB < 1$ and Strength near 0. The proposed method performs well in this example.

Table 4: Relative Belief Ratios and Strength for different values of a and F_0 in Example

Strength	RB	a	F_0
0.7623	7.5284	1	$N(0, 1)$
0.8670	6.4202	5	$N(0, 1)$
0.7027	9.2031	10	$N(0, 1)$
0	0	1	$N(2, 1)$
0	0	5	$N(2, 1)$
0	0	10	$N(2, 1)$
0	3.3320	1	Exp(1)
0	3.1230	5	Exp(1)
0.0142	3.0000	10	Exp(1)

5 Sensitivity to Prior Parameters

The performance of the proposed Bayesian nonparametric estimator and the resulting goodness-of-fit test heavily depends on the choice of the Dirichlet Process (DP) parameters: the concentration parameter α and the base distribution G .

The concentration parameter α determines the balance between prior belief (encoded in G) and observed data. A small α leads to a more data-driven posterior, whereas a large α causes the posterior to resemble G more strongly, potentially masking the effect of the data. Theoretically, as shown in [Ferguson \(1973\)](#), when $\alpha \rightarrow 0$, the DP becomes fully data-dominated, and when $\alpha \rightarrow \infty$, it collapses to the base distribution G .

The choice of the base distribution G is also critical. If G is close to the true data-generating process, the prior will assist estimation; however, a misspecified G can bias the inference, especially when α is large. Therefore, in practice, one should either use a weakly informative G or conduct robustness checks against multiple base distributions.

Our simulation studies (see [Tables 3 and 4](#)) empirically confirm these effects: lower values of α result in estimates that are more aligned with the observed data, whereas higher values of α lead to increased prior influence. These findings emphasize the importance of thoughtful prior specification in Bayesian nonparametric settings.

6 Computational Considerations

The proposed method involves a relatively high computational cost, particularly when dealing with large sample sizes or repeated simulations. In each run, the algorithm must sort the data and compute the extropy value, both of which require a substantial number of numerical operations.

As the sample size or the number of DP simulations increases, the execution time grows significantly. To handle this, it is recommended to use parallel processing or efficient programming techniques (e.g., vectorized code in R or Python). In addition, applying appropriate data structures or simplifying calculations where possible can lead to faster implementation [Hjort et al. \(2010\)](#).

Despite these computational demands, the proposed method provides valuable flexibility and accuracy, especially in scenarios where classical goodness-of-fit tests are less effective.

7 Conclusion

In this paper, we introduced a Bayesian nonparametric goodness of fit test based on extropy. The proposed method demonstrated superior flexibility, accuracy, and interpretability compared to traditional tests, particularly for small samples and complex deviations.

References

- Al-Labadi, L., & Evans, M. (2018). Prior-based model checking. *Canadian Journal of Statistics*, 46, 380–398.
- D’Agostino, R. B., Stephens, M. A. (1986). *Goodness-of-fit techniques*. Marcel Dekker.
- Evans, M. (2015). *Measuring Statistical Evidence Using Relative Belief*. Chapman and Hall/CRC.
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2), 209–230.
- Grzegorzewski, P., & Wieczorkowski, R. (1999). Entropy-based goodness-of-fit test for exponentiality. *Communications in Statistics - Theory and Methods*, 28, 1183–1202.
- Hjort, N. L., Holmes, C., Müller, P., Walker, S. G. (Eds.). (2010). **Bayesian Nonparametrics**. Cambridge University Press.
- Källberg, D., and Seleznev, O. (2016). Estimation of entropy-type integral functionals. *Communications in Statistics - Theory and Methods*, 45(4), 887–905.
- Lad, F., Sanfilippo, G., & Agro, G. (2015). Extropy: Complementary dual of entropy. *Statistical Science*, 30(1), 40–58.
- Noughabi, H. A., and JarrahiFeriz, J. (2022). Extropy of order statistics applied to testing symmetry. *Communications in Statistics - Simulation and Computation*, 51(10), 3389–3399.
- Qiu, G., & Jia, K. (2018). Extropy estimators with applications in testing uniformity. *Journal of Nonparametric Statistics*, 30, 182–196.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



The PRGM Prediction for Finite Population Parameters Using a Parametric Model Under SE and RN Loss Functions

Razieh Jafaraghaie^{*1}, Ferdous Mohammadi Basatini¹

¹Department of Mathematics and Statistics, Sho.C., Islamic Azad University, Shoushtar, Iran

Abstract

Predicting finite population parameters is a fundamental issue in statistical sampling. This paper investigates the prediction of a general linear combination of finite population values under a normal superpopulation model using Bayesian and robust Bayesian approaches. The primary focus is on the Posterior Regret Gamma-Minimax (PRGM) predictor under two loss functions: Squared Error (SE) and Reflected Normal (RN). The differences between these loss functions and their impacts on prediction accuracy are examined. Using real data and simulations, the performance of PRGM, Bayesian, and frequentist predictors in predicting the finite population mean is compared. Results indicate that the PRGM predictor performs well under both loss functions, particularly when the prior distribution's variance is small.

Keywords: Bayesian approach, PRGM predictor, Finite population, SE loss function, RN loss function, Superpopulation model.

1 Introduction

In finite population sampling analysis, predicting parameters such as the population mean or total is of great importance. Classical methods have been developed based on unbiased estimators (e.g., Godambe (1955); Godambe and Joshi (1965)), but these approaches rely on strong assumptions regarding the data model and underlying parameters. In contrast, model-based Bayesian (superpopulation) approaches consider the finite population as a realization from a parametric model (e.g., the normal distribution) and provide predictions for population parameters based on that model. This framework allows greater flexibility in modeling uncertainty and incorporating prior knowledge.

However, classical Bayesian approaches require the specification of a particular prior distribution, which may lead to unstable results if there is uncertainty about the hyperparameters. To address this issue, robust Bayesian methods have been proposed. These methods define a class Γ of plausible prior distributions and aim to limit the impact of prior uncertainty on the resulting inferences. For instance, Berger (1984) emphasized the advantage of using a broad

^{*}Corresponding author, Razieh.Jafaraghae @iau.ac.ir

class of priors instead of a single prior distribution. In this framework, a set of Bayesian estimators is obtained corresponding to the members of Γ , and the goal is to select an estimator that yields optimal stability in prediction.

One prominent approach in this context is the Posterior Regret Gamma-minimax (PRGM) method, which optimizes predictions by minimizing the maximum "posterior regret." In other words, the PRGM predictor is the estimator that minimizes the worst-case difference between its posterior loss and that of the best possible predictor. The PRGM concept was first introduced by Rios Insua et al. (1995) and has since been applied to various Bayesian prediction problems. In the present study, the PRGM method is employed to predict parameters of a finite population.

The analysis assumes that the finite population arises from a superpopulation normal model. Using this model, PRGM predictors are derived for population parameters such as the mean or total. To evaluate prediction error, two common loss functions are considered: the Squared Error (SE) loss and the Reflected Normal (RN) loss. The SE loss is a standard criterion for assessing prediction error variance, while the RN loss is used to measure relative errors.

The main innovation of this study lies in the analytical derivation of PRGM predictors for finite population parameters under a heterogeneous normal superpopulation model and in the empirical comparison of this method with conventional Bayesian predictors. Through extensive real data and simulation studies, the performance of the PRGM predictor is compared with that of standard Bayesian predictors. The findings indicate that under prior uncertainty, the robust PRGM predictor generally yields lower prediction errors (under SE or RN loss) compared to traditional Bayesian methods.

In the related literature, several key studies have laid the theoretical foundation for this topic. For example, Godambe (1955) and Godambe and Joshi (1965) established the theoretical basis for Bayesian estimation in finite populations. Berger (1984) introduced the robust Bayesian framework and emphasized the importance of considering a class Γ of prior distributions. Bolfarine (1990) focused on linear Bayesian prediction in finite populations, and Ghosh and Kim (1993) proposed a robust predictor based on posterior regret considerations. Finally, Rios Insua et al. (1995) developed the PRGM method for Bayesian estimation. The following references represent the most important contributions to this area of study. Further developments by Jafaraghaie and Nematollahi (2016, 2018), Ghosh (2008), Si et al. (2015), and Kaminska and Porosinski (2009) have contributed to advancing both theoretical understanding and practical application of Bayesian prediction in finite population settings.

The paper is organized as follows: Section 2 defines the superpopulation model and the SE and RN loss functions. Section 3 presents the Bayesian and frequentist predictors, setting the stage for robust methods. Section 4 derives the PRGM predictor under SE and RN loss functions, emphasizing its robustness properties. Section 5 analyzes real data and simulation results to compare predictor performance, and Section 6 discusses the implications and future directions of this work.

2 Concepts and Definitions

2.1 Superpopulation Model

Consider a finite population $\mathcal{U} = \{1, 2, \dots, N\}$ with unknown values represented by the vector $\mathbf{y} = (y_1, \dots, y_N)^T$. A sample of size n , denoted $\mathbf{y}_s = \{y_i : i \in s\}$, is drawn from this population. The goal is to predict

$$\gamma(\mathbf{y}) = \sum_{i=1}^N p_i y_i,$$

where $p_i > 0$ are known weights. The superpopulation model is defined as:

$$y_i = \beta x_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2 x_i), \quad i = 1, \dots, N, \quad (1)$$

where x_i is an auxiliary variable, β is an unknown parameter, and σ^2 is a known variance. This model assumes that the target variable is, on average, proportional to the auxiliary variable.

2.2 Loss Functions

The loss function evaluates the accuracy of the predictor $\delta(\mathbf{y}_s)$ relative to $\gamma(\mathbf{y})$. This paper considers two loss functions:

1. **Squared Error (SE) Loss:** This is an unbounded symmetric loss function defined as:

$$L(\Delta) = \Delta^2, \quad \Delta = \delta(\mathbf{y}_s) - \gamma(\mathbf{y}).$$

Due to its simplicity and mathematical properties, SE is widely used but may be unsuitable in financial applications where large losses are undesirable.

2. **Reflected Normal (RN) Loss:** This is a bounded symmetric loss function defined as:

$$L(\Delta) = 1 - e^{-\frac{\Delta^2}{2\gamma^2}},$$

where γ is a shape parameter and $k = 1$ is the maximum loss. The bounded nature of RN makes it more suitable for financial and decision-making contexts where large losses are unacceptable.

Differences between SE and RN: The SE loss grows quadratically with error, heavily penalizing large errors, which is useful when both small and large errors are equally important. In contrast, the RN loss caps the penalty for large errors, making it suitable for scenarios where prediction stability is prioritized over absolute accuracy. However, RN may be less sensitive in cases where high precision is critical.

3 Bayesian and Frequentist Predictors

3.1 Frequentist Predictor

In the frequentist approach, the prediction of $\gamma(\mathbf{y})$ is based on the model (1) and the minimum variance unbiased estimator $\hat{\beta} = \frac{\bar{y}_s}{\bar{x}_s}$. The frequentist predictor is:

$$\hat{\gamma}(\mathbf{y}) = \sum_{i \in s} p_i y_i + \sum_{i \notin s} p_i x_i \hat{\beta}.$$

Due to its lack of prior information, this predictor may exhibit lower accuracy in small samples.

3.2 Bayesian Predictor

In the Bayesian approach, β is assumed to follow a normal prior distribution $N(\mu, \tau_0^2)$. The conditional posterior distribution of $\gamma(\mathbf{y})$ given the observed data $\mathbf{y}_s, \mathbf{x}_s$ is:

$$N \left(\sum_{i \in s} p_i y_i + \left(\frac{\bar{y}_s}{\bar{x}_s} + B_0 \left(\mu - \frac{\bar{y}_s}{\bar{x}_s} \right) \right) h_s, \sigma^2 \left(\sum_{i \notin s} p_i^2 x_i + (1 - B_0) \frac{h_s^2}{n \bar{x}_s} \right) \right),$$

where $B_0 = \frac{\sigma^2}{\sigma^2 + \tau_0^2 n \bar{x}_s}$ and $h_s = \sum_{i \notin s} p_i x_i$. The Bayesian predictor under both SE and RN loss functions is:

$$\delta^\pi = E(\gamma(\mathbf{y}) \mid \mathbf{y}_s, \mathbf{x}_s) = \sum_{i \in s} p_i y_i + \left(\frac{\bar{y}_s}{\bar{x}_s} + B_0 \left(\mu - \frac{\bar{y}_s}{\bar{x}_s} \right) \right) h_s.$$

The posterior risk under SE and RN is computed as:

$$\rho(\pi, \delta^\pi) = \text{Var}(\gamma(\mathbf{y}) \mid \mathbf{x}_s, \mathbf{y}_s), \quad \rho(\pi, \delta^\pi) = 1 - \frac{\gamma}{\sqrt{\gamma^2 + w}},$$

where $w = \text{Var}(\gamma(\mathbf{y}) \mid \mathbf{x}_s, \mathbf{y}_s)$.

4 PRGM Predictor under SE and RN Loss Functions

In the robust Bayesian approach, the prior distribution (π) belongs to a class $\Gamma_\mu = \{\pi : \pi \sim N(\mu, \tau_0^2), \mu_1 \leq \mu \leq \mu_2\}$. The PRGM predictor is defined as:

$$\delta_\Gamma^{\text{PRGM}} = \frac{\delta^{\pi\mu_1} + \delta^{\pi\mu_2}}{2},$$

where $\delta^{\pi\mu_i}$ is the Bayesian predictor under the prior $N(\mu_i, \tau_0^2)$. Under model (1) and the class Γ_μ , the PRGM predictor is:

$$\delta_\Gamma^{\text{PRGM}} = \sum_{i \in s} p_i y_i + \left(\frac{\bar{y}_s}{\bar{x}_s} + B_0 \left(\mu^* - \frac{\bar{y}_s}{\bar{x}_s} \right) \right) h_s, \quad \mu^* = \frac{\mu_1 + \mu_2}{2}.$$

Under both SE and RN loss functions, the PRGM predictor remains identical, as these loss functions depend on minimizing the conditional expected value.

5 Real Data and Simulation

5.1 Real Data

To evaluate predictor performance, real data from the Household Survey Statistics (HSS) were used, examining the relationship between income (auxiliary variable x_i) and net worth (target variable y_i). Analysis of variance indicated that income significantly affects net worth ($p\text{-value} < 0.05$). Model coefficients are presented in Table 1. Maximum likelihood estimates of the parameters are shown in Table 2.

Table 1: Test of Model Coefficients

	Estimate	Standard Error	T-Value	p-value
Intercept	-0.7726	0.1286	-6.007	0.250
X	4.5432	0.1996	22.765	0.000

Table 2: Maximum Likelihood Parameter Estimates

Parameters	Estimate	Standard Error
β	3.4363	0.066
σ^2	0.5320	0.0359

Table 3: The analysis of variance

Source of Variation	Sum of Squares	df	Mean Squares	F	p-value
Model	1437.2	1	1437.2	1797	0.000
Error	233.6	292	0.8	—	—
Total	1670.8	293	—	—	—

From Table 3, income is significant in the model and the coefficient of determination ($R^2 = 0.8602$) indicates that approximately 86% of the variation in net worth is explained by income.

5.2 Simulation

The simulation study presented in this section is not purely theoretical or synthetic. It is based on the real data analyzed in Section 5.1 (Household Survey Statistics - HSS). Specifically, the

model parameters (such as β and σ^2) were estimated from the real data using maximum likelihood estimation. These estimated parameters were then used to generate synthetic populations under the fitted superpopulation model. Random samples of sizes $n = 20, 30, 50$ were drawn repeatedly (with the process repeated 10,000 times) from these populations to assess and compare the performance of different predictors (frequentist, Bayesian, and PRGM) under both SE and RN loss functions.

Hence, the simulation is directly grounded in the empirical structure of real data and aims to evaluate the practical prediction performance of the proposed methods. The Bayesian predictor used a prior $N(6, 0.01)$, the PRGM predictor used the class $\Gamma_\mu = \{N(\mu, \tau_0^2) : \mu \in [0, 6]\}$, and the frequentist predictor was computed.

Table 4: Predicted Finite Population Mean under SE and RN Loss Functions

Predictor	Predicted Value
δ^{π_0} (Bayes predictor)	1.66046
δ^{PRGM}	1.41057
$\delta^{\text{frequentist}}$	1.4518

Table 4 shows that the PRGM predictor is closest to the true mean (1.427406).

5.2.1 Performance Evaluation:

The Estimated Mean Squared Error (EMSE) and Estimated Absolute Bias (EAB) under SE, and the Estimated Risk Function (ERF) and Estimated Absolute Risk Bias (EARB) under RN, were computed. Results are presented in Tables 5 and 6.

Table 5: Simulated MSE and absolute bias of predictors for different τ_0^2 , μ_0 , and $\mu \in [1, 5]$ over Γ_μ (SE loss function)

	n	$\delta^\pi(\mu_0 = 1)$	$\delta^\pi(\mu_0 = 2)$	$\delta^\pi(\mu_0 = 4)$	$\delta^\pi(\mu_0 = 5)$	δ^{robust}	$\delta^{\text{frequentist}}$
$\sigma_0^2 = 0.001$							
EMSE	20	0.01008	0.01003	0.01004	0.01009	0.01002	0.01020
	30	0.00650	0.00640	0.00640	0.00650	0.00637	0.00654
	50	0.00380	0.00352	0.00341	0.00356	0.00339	0.00352
EAB	20	0.07890	0.07880	0.07902	0.07935	0.07884	0.07961
	30	0.06363	0.06321	0.06336	0.06396	0.06311	0.06401
	50	0.04889	0.04707	0.04643	0.04765	0.04624	0.04714
$\sigma_0^2 = 0.01$							
EMSE	20	0.03076	0.01232	0.00912	0.01406	0.00888	0.01020
	30	0.01915	0.01126	0.00602	0.01199	0.00556	0.00654
	50	0.01627	0.00994	0.00397	0.01131	0.00339	0.00352
EAB	20	0.16806	0.10044	0.07584	0.10719	0.07367	0.07961
	30	0.12174	0.08379	0.06217	0.08922	0.05867	0.06401
	50	0.10423	0.08186	0.05147	0.08916	0.04653	0.04714
$\sigma_0^2 = 0.1$							
EMSE	20	0.35610	0.12434	0.01943	0.14629	0.01211	0.01020
	30	0.29285	0.10291	0.01695	0.12092	0.01095	0.00654
	50	0.22269	0.07993	0.01455	0.09193	0.01055	0.00352
EAB	20	0.59601	0.35168	0.13697	0.38129	0.10736	0.07961
	30	0.53853	0.31803	0.12345	0.34349	0.09765	0.06401
	50	0.46581	0.27622	0.10845	0.29300	0.08918	0.04714

Table 6: Simulated risk function (EMSE) and absolute risk bias (EAB) for different λ_0 and σ_0^2 over Γ_μ (RN loss function)

	n	$\delta^\pi(\mu_0 = 1)$	$\delta^\pi(\mu_0 = 2)$	$\delta^\pi(\mu_0 = 4)$	$\delta^\pi(\mu_0 = 5)$	δ^{robust}	$\delta^{frequentist}$
$\sigma_0^2 = 0.001$							
EMSE	20	0.01948	0.01940	0.01942	0.01952	0.01938	0.01972
	30	0.01273	0.01255	0.01255	0.01274	0.01249	0.01282
	50	0.00750	0.00697	0.00674	0.00705	0.00671	0.00696
EAB	20	0.00532	0.00327	0.00083	0.00289	0.00128	0.00088
	30	0.00347	0.00219	0.00051	0.00183	0.00122	0.00032
	50	0.00310	0.00214	0.00037	0.00054	0.00082	0.00023
$\sigma_0^2 = 0.01$							
EMSE	20	0.05912	0.02414	0.01772	0.02746	0.01721	0.01972
	30	0.03694	0.02172	0.01183	0.02322	0.01092	0.01282
	50	0.03117	0.01937	0.00786	0.02205	0.00671	0.00696
EAB	20	0.03914	0.02407	0.00915	0.02541	0.00767	0.00088
	30	0.02808	0.01698	0.00630	0.01777	0.00541	0.00032
	50	0.02102	0.01289	0.00378	0.01203	0.00458	0.00023
$\sigma_0^2 = 0.1$							
EMSE	20	0.50823	0.21966	0.03803	0.25291	0.02388	0.01972
	30	0.43973	0.18489	0.03314	0.21285	0.02153	0.01282
	50	0.35339	0.14593	0.02840	0.16496	0.02063	0.00696
EAB	20	0.07459	0.06843	0.03285	0.07096	0.02614	0.00088
	30	0.07363	0.06436	0.02953	0.06684	0.02372	0.00032
	50	0.07302	0.05815	0.02473	0.05973	0.02093	0.00023

5.3 Analysis of Results

The PRGM predictor outperforms Bayesian and frequentist predictors under both SE and RN loss functions, particularly when τ_0^2 is small. The RN loss, due to its bounded nature, performs better in scenarios where prediction stability is critical, but it may be slightly less accurate in terms of absolute precision. As sample size increases, EMSE, EAB, ERF, and EARB decrease.

6 Discussion and Conclusion

This paper investigated the prediction of finite population parameters using the PRGM predictor under SE and RN loss functions. Simulation results demonstrated that the PRGM predictor outperforms Bayesian and frequentist predictors, especially with small prior variances. The RN loss function, due to its bounded nature, is more suitable for applications where large losses are unacceptable, while the SE loss excels in scenarios prioritizing high accuracy. Future research should explore these predictors under more complex models or non-normal data.

7 References

1. Berger, J. O. (1984), The robust Bayesian viewpoint (with discussion). In *Robustness of Bayesian analysis*, ed. J. Kadane, 63–124. Amsterdam: North-Holland.
2. Bolfarine, H. (1990), Bayesian linear prediction in finite populations. *Annals of the Institute of Statistical Mathematics*, 42(3), 435–444.

3. Ghosh, M. (2008), Robust estimation in finite population sampling. *Institute of Mathematical Statistics*, 1, 116–122.
4. Ghosh, M., & Kim, D. H. (1993), Robust Bayes estimation of the finite population mean. *The Indian Journal of Statistics*, 55, 322–342.
5. Godambe, V. P., & Joshi, V. M. (1965), Admissibility and Bayes estimation in sampling finite populations I. *The Annals of Mathematical Statistics*, 36(6), 1707–1722.
6. Jafaraghaie, R., & Nematollahi, N. (2016), Admissibility of linear predictors of finite population parameters under reflected normal loss function. *Journal of Sciences, Islamic Republic of Iran*, 27(4).
7. Jafaraghaie, R., & Nematollahi, N. (2018), Robust Bayesian prediction under a general linear-exponential posterior risk function and its application in finite population. *Communications in Statistics: Theory and Methods*, 47(13), 3269–3292.
8. Kaminska, A., & Porosinski, Z. (2009), On robust Bayesian estimation under some asymmetric and bounded loss function. *Statistics*, 43, 253–265.
9. Kiapour, A., & Nematollahi, N. (2011), Robust Bayesian prediction and estimation under a squared log error loss function. *Statistics & Probability Letters*, 81(11), 1717–1724.
10. Rios Insua, D., Ruggeri, F., & Vidakovic, B. (1995), Some results on posterior regret (Γ)-minimax estimation. *Statistics and Decisions*, 13, 315–331.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Statistical Inference Based on Composite Likelihoods for Spatio-temporal Point Processes

Abdollah Jalilian^{*12}

¹Department of Statistics, Razi University, Kermanshah, Iran

²Lancaster Ecology and Epidemiology Group, Lancaster University, Lancaster, UK

Abstract

Understanding the dynamics of tropical rain forests requires modeling the complex interplay of tree births (recruitment), deaths, and spatial interactions with the environment. In this talk, we present a statistical framework for analyzing spatio-temporal patterns of tree recruitment and mortality observed across successive forest censuses. We specify regression models for the conditional intensity of recruitment and the conditional probability of death, incorporating spatial covariates and the current tree configuration. To estimate model parameters, we employ conditional composite likelihood methods based on first-order properties of the data. This enables robust inference while avoiding strong assumptions about temporal dependence. We construct assumption-lean estimators of the parameter covariance matrix under weak spatial dependence, using a central limit theorem for increasing spatial domain asymptotic, an approach particularly suited to ecological data where time series are short but spatial coverage is extensive. The methodology also accommodates stochastic covariates derived from past point patterns by assuming weak dependence in the space-time process. We illustrate the utility of our approach through simulation studies and an application to tropical rain forest data, demonstrating its effectiveness in capturing the spatial structure and temporal evolution of forest dynamics.

Keywords: Central limit theorem; Composite likelihood; Conditional centering; Estimating function; Spatio-temporal processes.

Mathematics Subject Classification (2020): 62M30, 62F12, 62P12.

*Corresponding author, stat4aj@gmail.com



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Complete convergence for randomly weighted sums of WOD sequences

R.Javanmard*, A.Nezakati

Faculty of Mathematical Sciences, Shahrood University of Technology, Shahrood, Iran

Abstract

In this paper, we will study the complete convergence for double-indexed randomly weighted sums of WOD random variables. Several sufficient conditions to prove the complete convergence for randomly weighted sums of WOD random variables are presented.

Keywords: WOD random variables; randomly weighted sums; complete convergence; stochastically dominated.

Mathematics Subject Classification (2020): 60F15, 62G05.

1 Introduction

In Hsu P. L. and Robbins H. (1947) if $\{X, X_n, n \geq 1\}$ be a sequence of independent and identically distributed (i.i.d., for short) random variables and set $S_n = \sum_{i=1}^n X_i$ such that $EX^2 < \infty$ and $EX = \mu$, then for all $\epsilon > 0$,

$$\sum_{n=1}^{\infty} P(|n^{-1}S_n - \mu| > \epsilon) < \infty.$$

The aforementioned result is the well-known **complete convergence**, which implies that $S_n/n \rightarrow 0$ almost surely (a.s., for short) by the Borel-Cantelli lemma.

In Baum L. E. and Katz M. (1965) has been proven For a sequence $\{X, X_n, n \geq 1\}$ of i.i.d. random variables with zero mean and for all $\epsilon > 0$,

$$\sum_{n=1}^{\infty} n^{r-2} P\left(\left|\sum_{i=1}^n X_i\right| > \epsilon n^{1/p}\right) < \infty$$

if, and only if, $E|X|^{rp} < \infty$, where $r > 1$ and $1 \leq p < 2$. Since its first appearance in 1947, many authors have made enormous achievements in this field. In many stochastic models, the assumption that random variables are independent is not plausible. So it is interesting to extend the concept of independence to dependence cases. One of these dependence structures is widely orthant dependent (WOD), which has important applications in time series and financial mathematics. Many scholars studied the convergence properties for weighted sums of WOD random variables, and obtained many interesting results. In this work, we will further study the strong convergence properties for randomly weighted sums of WOD random variables.

*email: Rezvan.Javanmard@shahroodut.ac.ir

Now let us recall the concept of WOD random variables, which was raised by Wang and et al. (2013) as follows.

Definition 1.1. A sequence $\{X_n, n \geq 1\}$ of random variables is said to be widely upper orthant dependent (WUOD) if there exists a finite sequence $\{g_U(n), n \geq 1\}$ of positive numbers such that for each $n \geq 1$ and for all $x_i \in (-\infty, \infty)$, $1 \leq i \leq n$,

$$P(X_1 > x_1, X_2 > x_2, \dots, X_n > x_n) \leq g_U(n) \prod_{i=1}^n P(X_i > x_i).$$

A sequence $\{X_n, n \geq 1\}$ is widely lower orthant dependent (WLOD) if there exists a finite sequence $\{g_L(n), n \geq 1\}$ of positive numbers such that for each $n \geq 1$ and for all $x_i \in (-\infty, \infty)$, $1 \leq i \leq n$,

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \leq g_L(n) \prod_{i=1}^n P(X_i \leq x_i).$$

If $\{X_n, n \geq 1\}$ is both WUOD and WLOD, then it is said to be WOD with dominating coefficients $g(n) = \max\{g_U(n), g_L(n)\}$.

The definitions of stochastic domination and slowly varying function below will be used throughout the paper.

Definition 1.2. A triangular array $\{X_{nk}, 1 \leq k \leq n, n \geq 1\}$ of random variables is said to be **stochastically dominated** by a random variable X , if there exists a positive constant C such that

$$P(|X_{nk}| > x) \leq CP(|X| > x), \quad \text{for any } x \geq 0, 1 \leq k \leq n, \text{ and } n \geq 1.$$

Definition 1.3. Let $l(x)$ be a real-valued function, positive and measurable on $(0, \infty)$. It is **slowly varying at infinity** if for any $t > 0$,

$$\lim_{x \rightarrow \infty} \frac{l(tx)}{l(x)} = 1.$$

Baum L. E. and Katz M. (1965) studied the complete convergence for a sequence of independent identical distributed random variables. Then the theorems for independent random variables were extended to the cases of dependent random variables. Recently, Wang and et al. (2014) and Wang and et al. (2017) have studied the complete convergence for weighted sums of WOD random variables. Lu D. and Wang J. (2021) investigate the complete convergence and the complete moment convergence for maximal randomly weighted sums of a triangular array of row-wise WOD random variables. Now many people pay attention to the randomly weighted sums of random variables. Wu Y. and Zhao D. (2024) further investigate the single logarithm rule in the full sense of convergence for randomly weighted sums of WOD, where the weights are also assumed to be WOD.

In this paper, we will study the complete convergence for double-indexed randomly weighted sums of WOD where the weights are independent and nonnegative random variables. Several sufficient conditions to prove the complete convergence for randomly weighted sums of WOD random variables are presented. The result improves and generalizes some corresponding ones in the literature and the conditions considered in our paper are better and easier than Wu Y. and Zhao D. (2024).

2 Preliminary lemmas

In this section, we provide some lemmas to prove our main results. The first one is a basic property for WOD random variables, which can be found in Dawei Lu and Jialu Wang. To prove the main results, we list three assumptions (H1–H3) and present some important lemmas in the following. To prove the main results, we list three assumptions (H1–H3) and present some important lemmas in the following.

- (H1) Let $\{X_n, n \geq 1\}$ be a sequence of WOD random variables with dominating coefficients $g_u(m), g_l(m), m \geq 1$. Denote $g(m) = \max\{g_u(m), g_l(m)\}$.
- (H2) Let $\{X_{ni}, 1 \leq i \leq n, n \geq 1\}$ be a triangular array of row-wise WOD random variables with dominating coefficients $g_u^{(n)}(m), g_l^{(n)}(m), 1 \leq m \leq n, n \geq 1$ in each row. Denote $g^{(n)}(m) = \max\{g_u^{(n)}(m), g_l^{(n)}(m)\}$.
- (H3) Let $\{A_{ni}, 1 \leq i \leq n, n \geq 1\}$ be a triangular array of row-wise independent random variables, which are independent of $\{X_{ni}, 1 \leq i \leq n, n \geq 1\}$.

Here a triangular array of random variables $\{X_{ni}, 1 \leq i \leq n, n \geq 1\}$ are called row-wise WOD random variables if $\{X_{ni}, 1 \leq i \leq n\}$ is a sequence of WOD random variables for every $n \geq 1$. Then we show a basic property of WOD random variables in the following lemma. See [Wang and et al. \(2014\)](#) for more details.

Lemma 2.1. Let $\{X_n, n \geq 1\}$ be a sequence of WOD random variables with dominating coefficients $g(m), m \geq 1$:

- (i) $\{f_n(X_n), n \geq 1\}$ remains a sequence of WOD random variables with the same dominating coefficients $g(m), m \geq 1$, provided $\{f_n(\cdot), n \geq 1\}$ are all nondecreasing (or all nonincreasing) functions.
- (ii) For each $m \geq 1$ and all $s \in \mathbb{R}$:

$$E \exp \left\{ s \sum_{i=1}^m X_i \right\} \leq g(m) \prod_{i=1}^m E \exp \{s X_i\}$$

According to Lemma 2.1 and the definition of WOD random variables, we can obtain the following corollaries easily.

Corollary 2.1. Let $\{X_n, n \geq 1\}$ be a sequence of WOD random variables and $\{Y_n, n \geq 1\}$ be a sequence of nonnegative and independent random variables. If the two sequences are independent with each other, then $\{X_n Y_n, n \geq 1\}$ is still a sequence of WOD random variables.

In the following, we show some pivotal inequalities for WOD random variables. The most significant inequalities are Marcinkiewicz-Zygmund type maximal inequality and Rosenthal type maximal inequality given in the lemma below. See [Ding and et al. \(2017\)](#) for more details.

Lemma 2.2. Let $p \geq 1$ and $\{X_n, n \geq 1\}$ be a sequence of WOD random variables with $\mathbb{E}X_n = 0$ and $\mathbb{E}|X_n|^p < \infty$ for each $n \geq 1$. Then, for $m \geq 1$, there exist positive constants $C_1(p)$ and $C_2(p)$ depending only on p such that:

$$\mathbb{E} \left(\max_{1 \leq k \leq m} \left| \sum_{i=1}^k X_i \right|^p \right) \leq [C_1(p) + C_2(p)g(m)] (\log m)^p \sum_{i=1}^m \mathbb{E}|X_i|^p$$

for $1 \leq p \leq 2$, and

$$\mathbb{E} \left(\max_{1 \leq k \leq m} \left| \sum_{i=1}^k X_i \right|^p \right) \leq C_1(p)(\log m)^p \sum_{i=1}^m \mathbb{E}|X_i|^p + C_2(p)g(m)(\log m)^p \left(\sum_{i=1}^m \mathbb{E}|X_i|^2 \right)^{p/2}$$

for $p > 2$.

The following lemma can be found in [Wu Y. and Zhao D. \(2024\)](#).

Lemma 2.3. Let $\{Y_i, 1 \leq i \leq n\}$ and $\{Z_i, 1 \leq i \leq n\}$ be two sequences of random variables. Then for any $t > q > 0$, $\varepsilon > 0$, and $a > 0$:

$$\begin{aligned} \mathbb{E} \left[\left(\max_{1 \leq m \leq n} \left| \sum_{i=1}^m (Y_i + Z_i) \right| - \varepsilon a \right)_+^q \right] &\leq C_q \left(\varepsilon^{-t} + \frac{t-q}{q} \right) a^{q-t} \mathbb{E} \left[\max_{1 \leq m \leq n} \left| \sum_{i=1}^m Y_i \right|^t \right] \\ &\quad + C_q \mathbb{E} \left[\max_{1 \leq m \leq n} \left| \sum_{i=1}^m Z_i \right|^q \right], \end{aligned}$$

where $C_q = 1$ if $0 < q \leq 1$, or $C_q = 2^{q-1}$ if $q > 1$

The following lemma is an important property for stochastic domination, which can be found in [Wu and et al. \(2017\)](#) or [Shen and et al. \(2015\)](#).

Lemma 2.4. Let $\{X_{nk}, 1 \leq k \leq n, n \geq 1\}$ be a triangular array of random variables stochastically dominated by a random variable X . For any $a > 0$ and $b > 0$:

$$\begin{aligned}\mathbb{E}|X_{nk}|^a \mathbf{I}(|X_{nk}| \leq b) &\leq C_1 [\mathbb{E}|X|^a \mathbf{I}(|X| \leq b) + b^a \mathbb{P}(|X| > b)] \\ \mathbb{E}|X_{nk}|^a \mathbf{I}(|X_{nk}| > b) &\leq C_2 \mathbb{E}|X|^a \mathbf{I}(|X| > b)\end{aligned}$$

where C_1 and C_2 are positive constants. Consequently:

$$\mathbb{E}|X_{nk}|^a \leq C \mathbb{E}|X|^a$$

for some positive constant C .

The next one is important to prove the main result from [Lu and et al. \(2020\)](#)

Lemma 2.5. Let $r \geq 1$, $0 < p \leq 2$, $\alpha > 0$, $\beta > 0$ with $1/\alpha + 1/\beta = 1/p$. Let $\{X_{nk}, 1 \leq k \leq n, n \geq 1\}$ be an array of random variables stochastically dominated by X , and $\{A_{nk}\}$ a triangular array of random variables satisfying $\sum_{k=1}^n |A_{nk}|^\alpha \leq Cn$ a.s., independent of $\{X_{nk}\}$. Then:

$$\sum_{n=1}^{\infty} n^{r-2} l(n) \sum_{k=1}^n \mathbb{P}(|A_{nk} X_{nk}| > n^{1/p}) \leq \begin{cases} C \mathbb{E}|X|^{(r-1)\beta} l(|X|^\beta), & \alpha < rp \\ C \mathbb{E}|X|^{(r-1)\beta} \log(1 + |X|) l(|X|^\beta), & \alpha = rp \\ C \mathbb{E}|X|^{rp} l(|X|^p), & \alpha > rp \end{cases}$$

We give One remark on the condition $\sum_{k=1}^n |A_{nk}|^\alpha \leq Cn$ a.s.

Remark 2.2. By $\sum_{k=1}^n |A_{nk}|^\alpha \leq Cn$ a.s., we have

$$\sum_{k=1}^n |A_{nk}|^s \leq \begin{cases} Cn & \text{a.s., for any } 0 < s \leq \alpha, \\ Cn^{s/\alpha} & \text{a.s., for any } s > \alpha, \end{cases}$$

and thus,

$$\sum_{k=1}^n \mathbb{E}|A_{nk}|^s \leq \begin{cases} Cn & \text{for any } 0 < s \leq \alpha, \\ Cn^{s/\alpha} & \text{for any } s > \alpha. \end{cases}$$

Lemma 2.6. Let $r \geq 1$, $0 < p \leq 2$, $\alpha > 0$, $\beta > 0$ with $1/\alpha + 1/\beta = 1/p$. Let $\{X_{nk}, 1 \leq k \leq n, n \geq 1\}$ be a triangular array of random variables stochastically dominated by X , and $\{A_{nk}\}$ a triangular array of random variables satisfying $\sum_{k=1}^n |A_{nk}|^\alpha \leq Cn$ a.s., independent of $\{X_{nk}\}$. Then for any $s > \max\{\alpha, (r-1)\beta\}$:

$$\begin{aligned}\sum_{n=1}^{\infty} n^{r-2-s/p} l(n) \sum_{k=1}^n \mathbb{E}[|A_{nk} X_{nk}|^s \mathbf{I}(|A_{nk} X_{nk}| \leq n^{1/p})] \\ \leq \begin{cases} C \mathbb{E}|X|^{(r-1)\beta} l(|X|^\beta), & \alpha < rp \\ C \mathbb{E}|X|^{(r-1)\beta} \log(1 + |X|) l(|X|^\beta), & \alpha = rp \\ C \mathbb{E}|X|^{rp} l(|X|^p), & \alpha > rp \end{cases}\end{aligned}$$

The proof is in [Lu and et al. \(2020\)](#) The next one is the randomly weighted version of Lemma 2.4 in [Chen P. Y. and Sung S. H. \(2018\)](#).

Lemma 2.7. Let $q > 0$, $r \geq 1$, $0 < p \leq 2$, $\alpha > 0$, $\beta > 0$ with $\frac{1}{\alpha} + \frac{1}{\beta} = \frac{1}{p}$. Let $\{X_{nk}, 1 \leq k \leq n, n \geq 1\}$ be a triangular array of random variables stochastically dominated by X , and $\{A_{nk}\}$ a triangular array of random variables with $\sum_{k=1}^n |A_{nk}|^\alpha \leq Cn$ a.s., independent of $\{X_{nk}\}$. Then:

1. If $\alpha < rp$:

$$\begin{aligned} & \sum_{n=1}^{\infty} n^{r-2-q/p} l(n) \sum_{k=1}^n \mathbb{E} \left[|A_{nk} X_{nk}|^q I \left(|A_{nk} X_{nk}| > n^{1/p} \right) \right] \\ & \leq \begin{cases} C \mathbb{E} |X|^{(r-1)\beta} l(|X|^\beta), & q < (r-1)\beta \\ C \mathbb{E} |X|^{(r-1)\beta} \log(1+|X|) l(|X|^\beta), & q = (r-1)\beta \\ C \mathbb{E} |X|^q l(|X|^\beta), & q > (r-1)\beta \end{cases} \end{aligned}$$

2. If $\alpha = rp$:

$$\begin{aligned} & \sum_{n=1}^{\infty} n^{r-2-q/p} l(n) \sum_{k=1}^n \mathbb{E} \left[|A_{nk} X_{nk}|^q I \left(|A_{nk} X_{nk}| > n^{1/p} \right) \right] \\ & \leq \begin{cases} C \mathbb{E} |X|^{(r-1)\beta} \log(1+|X|) l(|X|^\beta), & q \leq \alpha = rp \\ C \mathbb{E} |X|^q l(|X|^\beta), & q > \alpha = rp \end{cases} \end{aligned}$$

3. If $\alpha > rp$:

$$\begin{aligned} & \sum_{n=1}^{\infty} n^{r-2-q/p} l(n) \sum_{k=1}^n \mathbb{E} \left[|A_{nk} X_{nk}|^q I \left(|A_{nk} X_{nk}| > n^{1/p} \right) \right] \\ & \leq \begin{cases} C \mathbb{E} |X|^{rp} l(|X|^p), & q < rp \\ C \mathbb{E} |X|^{rp} \log(1+|X|) l(|X|^p), & q = rp \\ C \mathbb{E} |X|^q l(|X|^p), & q > rp \end{cases} \end{aligned}$$

Lemma 2.8. Let $1 \leq p \leq 2$, $\alpha > 0$, $\beta > 0$ with $\frac{1}{\alpha} + \frac{1}{\beta} = \frac{1}{p}$. Let $\{X_{nk}, 1 \leq k \leq n, n \geq 1\}$ be a triangular array of random variables stochastically dominated by X , and $\{A_{nk}\}$ a triangular array of random variables with $\sum_{k=1}^n |A_{nk}|^\alpha \leq Cn$ a.s., independent of $\{X_{nk}\}$. If $\mathbb{E}|X|^p < \infty$, then:

$$n^{-1/p} \sum_{k=1}^n \mathbb{E} \left[|A_{nk} X_{nk}| I \left(|A_{nk} X_{nk}| > n^{1/p} \right) \right] \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Additionally, if $\mathbb{E}X_{nk} = 0$ for all n, k , then:

$$n^{-1/p} \max_{1 \leq m \leq n} \left| \sum_{k=1}^m \mathbb{E} \left[A_{nk} X_{nk} I \left(|A_{nk} X_{nk}| \leq n^{1/p} \right) \right] \right| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

The proof is in [Lu and et al. \(2020\)](#).

3 Main result and its proof

With preliminaries in Section 2 accounted for, we can give our main results and their proofs.

Theorem 3.1. Let $r \geq 1$, $1 \leq p < 2$, $\alpha > 0$, $\beta > 0$ with $\frac{1}{\alpha} + \frac{1}{\beta} = \frac{1}{p}$. Let $\{X_{nk}, 1 \leq k \leq n, n \geq 1\}$ be a triangular array of rowwise WOD random variables with dominating coefficient $g(n) = O(n^\tau)$ for some $\tau \geq 0$, stochastically dominated by a random variable X with $\mathbb{E}X_{nk} = 0$ and $\mathbb{E}|X|^p < \infty$. Let $\{A_{nk}, 1 \leq k \leq n, n \geq 1\}$ be a triangular array of rowwise independent and nonnegative random variables satisfying

$$\sum_{k=1}^n |A_{nk}|^\alpha \leq Cn \quad \text{a.s.}$$

for some constant $C > 0$, and independent of $\{X_{nk}, 1 \leq k \leq n, n \geq 1\}$. If

$$\begin{aligned} & \mathbb{E} \left[|X|^{(r-1)\beta+2\tau} l(|X|^\beta) \right] < \infty, & \text{if } \alpha < rp, \\ & \mathbb{E} \left[|X|^{(r-1)\beta+2\tau} \log(1+|X|) l(|X|^\beta) \right] < \infty, & \text{if } \alpha = rp, \\ & \mathbb{E} \left[|X|^{rp+2\tau} l(|X|^p) \right] < \infty, & \text{if } \alpha > rp, \end{aligned}$$

then for any $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} n^{r-2} l(n) \mathbb{P} \left(\max_{1 \leq m \leq n} \left| \sum_{k=1}^m A_{nk} X_{nk} \right| > \varepsilon n^{1/p} \right) < \infty. \quad (1)$$

where $l(x)$ is slowly varying at infinity.

Proof. For $1 \leq k \leq n$ and $n \geq 1$, denote

$$Y_{nk} = -n^{1/p} \mathbf{I} \left(A_{nk} X_{nk} < -n^{1/p} \right) + A_{nk} X_{nk} \mathbf{I} \left(|A_{nk} X_{nk}| \leq n^{1/p} \right) + n^{1/p} \mathbf{I} \left(A_{nk} X_{nk} > n^{1/p} \right),$$

$$\begin{aligned} Z_{nk} &= A_{nk} X_{nk} - Y_{nk} \\ &= \left(A_{nk} X_{nk} + n^{1/p} \right) \mathbf{I} \left(A_{nk} X_{nk} < -n^{1/p} \right) + \left(A_{nk} X_{nk} - n^{1/p} \right) \mathbf{I} \left(A_{nk} X_{nk} > n^{1/p} \right). \end{aligned}$$

We obtain that

$$\begin{aligned} &\sum_{n=1}^{\infty} n^{r-2} l(n) \mathbb{P} \left(\max_{1 \leq m \leq n} \left| \sum_{k=1}^m A_{nk} X_{nk} \right| > \varepsilon n^{1/p} \right) \\ &= C \sum_{n=1}^{\infty} n^{r-2} l(n) \mathbb{P} \left(\max_{1 \leq m \leq n} \left| \sum_{k=1}^m (Y_{nk} - \mathbb{E}[Y_{nk}] + Z_{nk} - \mathbb{E}[Z_{nk}]) \right| > \varepsilon n^{1/p} \right) \\ &\leq C \sum_{n=1}^{\infty} n^{r-2} l(n) \mathbb{P} \left(\max_{1 \leq m \leq n} \left| \sum_{k=1}^m (Y_{nk} - \mathbb{E}[Y_{nk}]) \right| > \frac{1}{2} \varepsilon n^{1/p} \right) \\ &\quad + C \sum_{n=1}^{\infty} n^{r-2} l(n) \mathbb{P} \left(\max_{1 \leq m \leq n} \left| \sum_{k=1}^m (Z_{nk} - \mathbb{E}[Z_{nk}]) \right| > \frac{1}{2} \varepsilon n^{1/p} \right) \\ &\triangleq I_1 + I_2. \end{aligned}$$

At first, we consider the case of $r > 1$. For I_2 , noting that when $\alpha \leq rp$, we have by $1/\alpha + 1/\beta = 1/p$ that $\alpha = p\beta/(\beta - p) \leq rp$, and thus, $(r-1)\beta \geq rp > 1$. Similarly, we also get that $rp > \max\{(r-1)\beta, 1\}$ when $\alpha > rp$. Thus, we know that there exist some real number q such that $1 \leq q < (r-1)\beta$ if $\alpha \leq rp$, and $1 \leq q < rp$ if $\alpha > rp$. Taking $q = 1$ in lemma 2.7, we have by Markov's inequality that

$$\begin{aligned} I_2 &= \sum_{n=1}^{\infty} n^{r-2} l(n) \mathbb{P} \left(\max_{1 \leq m \leq n} \left| \sum_{k=1}^m (Z_{nk} - \mathbb{E}[Z_{nk}]) \right| > \frac{1}{2} \varepsilon n^{1/p} \right) \\ &\leq C \sum_{n=1}^{\infty} n^{r-2-1/p} l(n) \sum_{k=1}^n \mathbb{E}[|A_{nk} X_{nk}|] \mathbf{I} \left(|A_{nk} X_{nk}| > n^{1/p} \right) < \infty. \end{aligned}$$

Note that

$$\begin{aligned} \mathbb{E} \left[\max_{1 \leq m \leq n} \left| \sum_{k=1}^m (Z_{nk} - \mathbb{E}[Z_{nk}]) \right| \right] &\leq \mathbb{E} \left[\left| \sum_{k=1}^n (Z_{nk} - \mathbb{E}[Z_{nk}]) \right| \right] \\ &= \mathbb{E} \left[\sum_{k=1}^n |Z_{nk} - \mathbb{E}[Z_{nk}]| \right] = \sum_{k=1}^n \mathbb{E}[|Z_{nk} - \mathbb{E}[Z_{nk}]|] \\ \mathbb{E}[|Z_{nk} - \mathbb{E}[Z_{nk}]|] &\leq 2 \mathbb{E}[|Z_{nk}|] \quad \text{and} \quad \mathbb{E}[|Z_{nk}|] \leq \mathbb{E}[|A_{nk} X_{nk}| \cdot \mathbf{I}_{\{|A_{nk} X_{nk}| > n^{1/p}\}}] \end{aligned}$$

For I_1 , without loss of generality, we assume that $\max_{1 \leq k \leq n} |A_{nk}| \leq n^{1/\alpha}$ holds almost surely. Noting that $r > 1$, we set $p < s < \min \left\{ \frac{p(r+2\tau)}{\tau+1}, 2, \alpha \right\}$ if $\alpha \leq rp$, and $p < s < \min \left\{ \frac{p(r+2\tau)}{\tau+1}, 2, rp \right\}$ if $\alpha > rp$. Taking $q > \max \left\{ \alpha, (r-1)\beta, 2\tau+2, p(r-1+\tau), \frac{2p(r-1+\tau)}{s-p} \right\}$, we obtain by Markov's inequality and Lemma 2.2 that:

$$\begin{aligned}
I_1 &= \sum_{n=1}^{\infty} n^{r-2} l(n) \mathbb{P} \left(\max_{1 \leq m \leq n} \left| \sum_{k=1}^m (Y_{nk} - \mathbb{E} Y_{nk}) \right| > \frac{1}{2} \varepsilon n^{1/p} \right) \\
&\leq C \sum_{n=1}^{\infty} n^{r-2-q/p} l(n) \mathbb{E} \left(\max_{1 \leq m \leq n} \left| \sum_{k=1}^m (Y_{nk} - \mathbb{E} Y_{nk}) \right| \right)^q \\
&\leq C_1 \sum_{n=1}^{\infty} n^{r-2-q/p+\tau} l^*(n) \left(\sum_{k=1}^n \mathbb{E} |Y_{nk} - \mathbb{E} Y_{nk}|^2 \right)^{q/2} \\
&\quad + C_2 \sum_{n=1}^{\infty} n^{r-2-q/p} l(n) (\log n)^q \sum_{k=1}^n \mathbb{E} |Y_{nk} - \mathbb{E} Y_{nk}|^q \\
&\triangleq I_3 + I_4.
\end{aligned}$$

where $l^*(n) = l(n)(\log n)^q$ is a slowly varying function at infinity.

For I_4 , by Lemmas 2.5 and 2.6, we obtain

$$\begin{aligned}
&\sum_{n=1}^{\infty} n^{r-2-q/p} l(n) \sum_{k=1}^n \mathbb{E} |Y_{nk} - \mathbb{E} Y_{nk}|^q \leq C \sum_{n=1}^{\infty} n^{r-2-q/p} l(n) \times \\
&\sum_{k=1}^n \left[\mathbb{E} |A_{nk} X_{nk}|^q I(|A_{nk} X_{nk}| \leq n^{1/p}) + n^{q/p} P(|A_{nk} X_{nk}| > n^{1/p}) \right] < \infty.
\end{aligned}$$

For I_3 , we have by C_r -inequality and Markov's that

$$\begin{aligned}
&\sum_{n=1}^{\infty} n^{r-2-q/p+\tau} l^*(n) \left(\sum_{k=1}^n \mathbb{E} (Y_{nk} - \mathbb{E} Y_{nk})^2 \right)^{q/2} \\
&\leq \sum_{n=1}^{\infty} n^{r-2-q/p+\tau} l^*(n) \\
&\quad \times \left(\sum_{k=1}^n \left[\mathbb{E} |A_{nk} X_{nk}|^2 I(|A_{nk} X_{nk}| \leq n^{1/p}) + n^{2/p} P(|A_{nk} X_{nk}| > n^{1/p}) \right] \right)^{q/2} \\
&\leq C \sum_{n=1}^{\infty} n^{r-2-q/p+\tau} l^*(n) \left(\sum_{k=1}^n \mathbb{E} |A_{nk} X_{nk}|^2 I(|A_{nk} X_{nk}| \leq n^{1/p}) \right)^{q/2} \\
&\quad + C \sum_{n=1}^{\infty} n^{r-2-q/p+\tau} l^*(n) \left(\sum_{k=1}^n n^{2/p} P(|A_{nk} X_{nk}| > n^{1/p}) \right)^{q/2} \\
&\leq C \sum_{n=1}^{\infty} n^{r-2-q/p+\tau} l^*(n) \left(n^{(2-s)/p} \sum_{k=1}^n \mathbb{E} |A_{nk} X_{nk}|^s I(|A_{nk} X_{nk}| \leq n^{1/p}) \right)^{q/2} \\
&\quad + C \sum_{n=1}^{\infty} n^{r-2-q/p+\tau} l^*(n) \left(n^{(2-s)/p} \sum_{k=1}^n \mathbb{E} |A_{nk} X_{nk}|^s \right)^{q/2} \\
&\leq C \sum_{n=1}^{\infty} n^{r-2-q/p+\tau} l^*(n) \left(n^{(2-s)/p} \sum_{k=1}^n \mathbb{E} |A_{nk}|^s \mathbb{E} |X|^s \right)^{q/2} < \infty
\end{aligned}$$

Note that

$$\begin{aligned}
&\mathbb{E} \left[|A_{nk} X_{nk}|^s I(|A_{nk} X_{nk}| \leq n^{1/p}) \right] \leq \mathbb{E} [|A_{nk} X_{nk}|^s] \\
&\sum_{n=1}^{\infty} n^{r-2-q/p+\tau} l^*(n) \left(n^{(2-s)/p} \sum_{k=1}^n \mathbb{E} |A_{nk} X_{nk}|^s \right)^{q/2} \\
&\leq C \sum_{n=1}^{\infty} n^{r-2-q/p+\tau} l^*(n) \left(n^{(2-s)/p} \mathbb{E} |X|^s \sum_{k=1}^n |A_{nk}|^s \right)^{q/2}
\end{aligned}$$

and with use from Remark 2.2 and Since

$$n^{r-2-q/p+\tau} \cdot \left(n^{(2-s)/p} \cdot n\right)^{q/2} = n^{r-2+\tau+(q/2)(1-s/p)}$$

and

$$r-2+\tau+\frac{q}{2}\left(1-\frac{s}{p}\right) < -1$$

if

$$q > \frac{2p(r-1+\tau)}{s-p}$$

then

$$\sum_{n=1}^{\infty} n^{r-2+\tau+\frac{q}{2}(1-s/p)} \cdot l^*(n) < \infty$$

Next, we will consider the case of $r = 1$. For any $1 \leq k \leq n$

and $n \geq 1$, set

$$\begin{aligned} \left\{ \max_{1 \leq m \leq n} \left| \sum_{k=1}^m A_{nk} X_{nk} \right| > \varepsilon n^{1/p} \right\} &\subseteq \bigcup_{k=1}^n \left\{ |A_{nk} X_{nk}| > n^{1/p} \right\} \\ &\cup \left\{ \max_{1 \leq m \leq n} \left| \sum_{k=1}^m T_{nk} \right| > \varepsilon n^{1/p} \right\}. \end{aligned}$$

Thus, we have

$$\begin{aligned} &\sum_{n=1}^{\infty} n^{r-2} l(n) \mathbb{P} \left(\max_{1 \leq m \leq n} \left| \sum_{k=1}^m A_{nk} X_{nk} \right| > \varepsilon n^{1/p} \right) \\ &\leq \sum_{n=1}^{\infty} n^{r-2} l(n) \mathbb{P} \left(\bigcup_{k=1}^n \left\{ |A_{nk} X_{nk}| > n^{1/p} \right\} \right) + \sum_{n=1}^{\infty} n^{r-2} l(n) \mathbb{P} \left(\max_{1 \leq m \leq n} \left| \sum_{k=1}^m T_{nk} \right| > \varepsilon n^{1/p} \right) \\ &\leq \sum_{n=1}^{\infty} n^{r-2} l(n) \sum_{k=1}^n \mathbb{P} \left(|A_{nk} X_{nk}| > n^{1/p} \right) + \sum_{n=1}^{\infty} n^{r-2} l(n) \mathbb{P} \left(\max_{1 \leq m \leq n} \left| \sum_{k=1}^m T_{nk} \right| > \varepsilon n^{1/p} \right). \end{aligned}$$

By Lemmas 2.5 and 2.6, to prove 1, it suffices to prove that

$$\sum_{n=1}^{\infty} n^{r-2} l(n) P \left(\max_{1 \leq m \leq n} \left| \sum_{k=1}^m (T_{nk} - \mathbb{E} T_{nk}) \right| > \varepsilon n^{1/p} \right) < \infty.$$

In fact, by $r = 1$, we have $\alpha > p = rp$. If $\sum_{k=1}^n |A_{nk}|^\alpha \leq Cn$ a.s. for some $\alpha > 0$, then we can get that $\sum_{k=1}^n |A_{nk}|^\alpha \leq Cn$ a.s. for any $0 < \alpha \leq \alpha$ by Remark 2.2 Thus we may assume that $p < \alpha < 2$. By Markov's inequality and Lemmas 2.2 and 2.6, we have

$$\begin{aligned} &\sum_{n=1}^{\infty} n^{r-2} l(n) P \left(\max_{1 \leq m \leq n} \left| \sum_{k=1}^m (T_{nk} - \mathbb{E} T_{nk}) \right| > \varepsilon n^{1/p} \right) \\ &\leq C \sum_{n=1}^{\infty} n^{r-2} \cdot n^{-2/p} l^*(n) \sum_{k=1}^n \mathbb{E} \left[|A_{nk} X_{nk}|^2 \mathbf{I} \left(|A_{nk} X_{nk}| \leq n^{1/p} \right) \right] < \infty. \end{aligned}$$

□

References

- Baum L. E. and Katz M. (1965), *Convergence rates in the law of large numbers*. Transactions of the American Mathematical Society, 120, 10823. DOI:10.2307/1994170.
- Chen P. Y. and Sung S. H. (2018), *On complete convergence and complete moment convergence for weighted sums of ρ^* -mixing random variables*. J Inequ Appl., 2018, Article ID 121, 16 pages.
- Ding Y., Wu Y., Ma S., Tao X. and Wang X. (2017), *Complete convergence and complete moment convergence for widely orthant-dependent random variables*, Communication in Statistics-Theory and Methods, 46, 8278-8294. DOI:10.1080/03610926.2016.1177085.

- Hsu P. L. and Robbins H. (1947), *Complete convergence and the law of large numbers*, *P. National Acad. Sci. USA*, 33, 25-31. DOI:10.1073/pnas.33.2.25.
- Lu C., Li X., Wang R. and Wang X. (2020), *Complete and complete moment convergence for randomly weighted sums of ρ^* -mixing random variables and its applications*. *Statistics*, 54, 205-237. DOI:10.1080/02331888.2019.1707201.
- Lu D. and Wang J. (2021), *Complete convergence and complete moment convergence for maximal randomly weighted sums of widely orthant-dependent random variables with applications*. *Communications in Statistics - Theory and Methods*, 50, 763-791. DOI:10.1080/03610926.2019.1640879.
- Shen A. T., Zhang Y. and Volodin A. (2015), *Applications of the Rosenthal-type inequality for negatively superadditive dependent random variables*. *Metrika.*, 78, 295311.
- Wang K. Y., Wang Y. B. and Gao Q. W. (2013), *Uniform asymptotics for the finite-time ruin probability of a new dependent risk model with a constant interest rate*. *Methodol. Comput. Appl. Probab.*, 15, 109124. DOI:10.1007/s11009-011-9226-y.
- Wang X., Xu C., Hu T.-C., Volodin A. and Hu S. (2014), *On complete convergence for widely orthant-dependent random variables and its applications in nonparametric regression models*. *Test*, 23, 60729. DOI:10.1007/s11749-014-0365-7.
- Wang X. J., Wu Y. and Rosalsky A. (2017), *Complete convergence for arrays of rowwise widely orthant dependent random variables and its applications*. *Stochastics - An International Journal of Probability and Stochastic Processes*, 89, 122852. DOI:10.1080/17442508.2017.1297813.
- Wu Y., Wang X. J. and Hu S. H. (2017), *Complete moment convergence for weighted sums of weakly dependent random variables and its application in nonparametric regression model*. *Statist. Probab. Lett.*, 127, 5666.
- Wu Y. and Zhao D. (2024), *Law of the single logarithm for randomly weighted sums of dependent sequences and an application*. *AIMS Mathematics*, 9(4), 1014110156.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Reliability Measures in Semi-Markov Repairable Systems

K. Khorshidian^{*1}, M. Fathizadeh²

¹Department of statistics, Shiraz University, Shiraz, Iran.

²Department of statistics, Vali-Asr University, Rafsanjan, Iran.

Abstract

The existence of a large number of studies related to the various aspects of reliability systems, where almost all of them emphasis on the mean residual life (MRL), reliability and availability functions, motivated us to investigate the state-dependent version of these measures for semi-Markov repairable systems. The first hitting time to a set of certain events has been considered by the assumption of having some information about the current status in semi-Markov processes (SMPs). The distribution function and expected value for the purposed hitting time have been separately formulated based on the conditional distribution of the excess life, as two reliability measures for semi-Markov models. The relation between these state-dependent measures has also been investigated. Moreover, the conditional probability of standing in a specified set at a certain time s has been considered as another important measure in semi-Markov models. As an application, a two unit cold standby system with one repairman has been considered as a semi-Markov model and the associated state-dependent metrics which called the state-dependent mean residual life (SDM), state-dependent reliability function (SDR) and state-dependent availability function (SDA), have been studied in details. To clarify and investigate the validity of results, a graphical study has been performed.

Keywords: Hitting Time, Mean Residual Life, Repairable systems, Semi-Markov Processes.

Mathematics Subject Classification (2020): Primary 60K15, Secondary 60K20, 90B25.

1 Introduction

In the present study we will focus on some certain time intervals and moments of times in an operating mechanism, according to the performance, failure or successes which may occur together with their probabilities. Time to occurrence of some significant events in stochastic models is of important and challenging issues in various sciences. The considered event may be an earthquake with a certain magnitude, occurrence of a certain period of hydrological droughts, the beginning of bull market in stock price market, etc. The quantity that specifies how much longer a mechanism or mechanical system may be alive and works after a certain

^{*}Corresponding author, khorshidian@shirazu.ac.ir

instant of time is known as mean residual life (MRL) and has a special position in planning the repair and replacement strategies in the reliability theory. Another function that is closely related to the last measure is the reliability function. This function gives the probability that a natural phenomenon, device, or other objects of interest may be active over a certain period of time.

SMPs are amongst the powerful tools in modeling stochastic events. In recent decades, a wide range of real problems in the various branches of sciences have been studied through Markov renewal theory. In 1999, Altinok and Kolcak have employed SMP for modeling the earthquake occurrence in North Anatolia, Turkey. They estimated earthquake occurrence probabilities with considering different states for magnitude of earthquakes. Nnaji et al. (2016) have considered drought periods as a four state semi-Markov model and have estimated drought indices in Apalachicola–Chattahoochee–Flint River Basin.

Reliability analysis has been done by using different approaches. One method is through the theory of stochastic processes, which mainly is done by establishing some Markov or semi-Markov structure for the purposed system. SMPs have been employed more in modeling and further analysis of systems rather than Markov processes, because of assumption of non-exponential lifetimes and its generality.

In recent decades many authors have employed SMPs to evaluate the reliability measures of repairable systems such as Bieth, Hong, and Sarkar (2009), Fathizadeh and Khorshidian (2016), Knopik and Migawa (2019), Wu et al. (2018). Analyzing the reliability and availability functions have been a common subject in many of these researches. Some reliability measures have been developed and successfully formulated for large wide class of semi-Markov systems by Limnios (2012) and (2014). In fact, Limnios arrived at a formulation for the reliability parameters in the systems with general state spaces from a theoretical perspective. The results of his studies include both continuous and discrete time SMPs and may be employed in a wide range of reliability systems. Although most of practical systems in engineering sciences are repairable, the study of reliability measures especially MRL function in non-repairable structures have received more attention (see e.g., Asadi and Bayramoglu (2006), Asadi and Goliforushani (2008), Eryilmaz (2013), Navarro and Eryilmaz (2007) and Raqab (2010)). Almost all of these studies have been done with the assumption of having some information only about the starting points of the system.

In practice, in many situations there are some useful information about the state of the system at a given time t and predicting the expected remaining time to visit other future states, may be of interest. For example, in most of manufacturing disciplines an operator or regular repairman is available who can provide some information about current or earlier working state of the system. In such cases, introducing some state-dependent measures may help us to make better decisions or improve the residual life of the mechanism.

Recently, Do and Be´renguer (2021) have studied the degradation-based failure models and introduced two novel conditional reliability measures to increase the residual life of the system. They have considered a coherent structured system consisting of n non-identical components and obtained the MRL and reliability function, with the assumption of having some information about current state of the system. They have indicated that these metrics can be employed to identify those degraded components which urgently must be replaced for improving the residual life of the system.

Prediction for reward processes has been brought to the literature of SMPs by Soltani, Khorshidian, and Ghafaripour (2010). In this regard, they have considered the past history of the triple process (semi-Markov, age and reward processes) up to time t , and introduced a prediction formula for the expected reward at a future instant $t + s$, by using the conditional expectation techniques.

1.1 Aims and Approach

In this paper, we will consider $B(t) = (X(t), U(t))$, where $\{X(t); t \geq 0\}$ is a SMP with the finite state space $\mathbb{E} = \mathbb{E}_1 \cup \mathbb{E}_2$, and $U(t)$ as the corresponding age process. Also, we define $S_{\mathbb{E}_1}^t$ to be the elapsed time for visit \mathbb{E}_1 after the current time t , i.e., the first hitting time to \mathbb{E}_1 after t ;

$$S_{\mathbb{E}_1}^t := \inf \{l \geq 0 : X(t+l) \in \mathbb{E}_1\}.$$

A formula for $E(S_{\mathbb{E}_1}^t | B(t))$ will have been presented based on the conditional distribution of the excess life for SMPs. In a same manner, we will also formulize $P(S_{\mathbb{E}_1}^t > s | B(t))$ and $P(X(t+s) = j, j \in \mathbb{E}_2 | B(t))$. These measures may be employed to verify the time to occurrence of important events, the probability that a device or other object be active over a certain time interval of length s and the probability that a device or other object be in a specified state at a certain time s , respectively.

There are great amounts of studies which seriously employ and benefit the MRL and other reliability functions for different types of systems, coherent and non-coherent, repairable and non-repairable. These studies and applications encouraged us to try state-dependent (information-based) versions of MRL, reliability function and availability function in repairable systems with semi-Markov structures. In order to do this, a two-unit cold standby system with one repairman has been considered as a semi-Markov model and the associated SDM, SDR and SDA have been studied in details.

The outline of this paper is as follows. In Section 2, some basic notation and definitions concerning SMPs have been given. Some formulas for the expected value and distribution function of conditional hitting time to some specified state of the considered SMP has been proposed in Section 3. Section 4 is devoted to calculating SDM, SDR and SDA in SMPs besides the applications of the obtained results in a well-known standby reliability system.

2 Notation and Preliminaries

In this section we provide some brief preliminaries on SMPs. For more details on SMPs and their applications we refer the readers to Cinlar (1975), Janssen and Manca (2006), Limnios and Oprisan (2001). A Markov renewal process may be defined as follows:

Let \mathbb{E} be a countable set and $(\mathbf{J}, \mathbf{T}) = \{(J_n, T_n), n = 0, 1, 2, \dots\}$ denote a two-dimensional stochastic process with values in $\mathbb{E} \times [0, \infty)$, then (\mathbf{J}, \mathbf{T}) is a Markov renewal process if for all $j \in \mathbb{E}$,

$$P(J_{n+1}=j, T_{n+1}-T_n \leq t | J_n, T_n, \dots, J_0, T_0) = P(J_{n+1}=j, T_{n+1}-T_n \leq t | J_n),$$

with probability one. We assume that the state space \mathbb{E} is finite, the process is time homogeneous with transition probabilities:

$$\mathcal{Q}_{ij}(t) = P(J_{n+1}=j, T_{n+1}-T_n \leq t | J_n = i), \quad i, j \in \mathbb{E}, \quad n \geq 0,$$

and $\mathcal{Q}(t) = [\mathcal{Q}_{ij}(t)]_{i,j \in \mathbb{E}}$, $t \in \mathbb{R}^+$, as the matrix of renewal kernels. Note that $\{J_n, n = 0, 1, \dots\}$ and $\{T_n, n = 0, 1, \dots\}$ are the state and time instant of the n^{th} transition.

The associated counting process representing the total number of transitions within $[0, t]$ is denoted by $\{N(t) : t \geq 0\}$. A stochastic process $\{X(t) : t \geq 0\}$ is called a SMP on \mathbb{E} , generated by the Markov renewal process (\mathbf{J}, \mathbf{T}) , with semi-Markov kernel $\mathcal{Q}(t)$, $t \geq 0$, if $X(t) = J_{N(t)}$. Since the trajectory of the SMPs have constant values on the half-interval $[T_n, T_{n+1})$ and are continuous from the right, $X(t) = X(T_n) = J_n$, for $t \in [T_n, T_{n+1})$. Moreover,

$$p_{ij} = P(J_n = j | J_{n-1} = i) = \lim_{t \rightarrow \infty} \mathcal{Q}_{ij}(t), \quad i, j \in \mathbb{E}, \quad n \geq 0$$

and $\mathbf{P} = [p_{ij}]$ is the transition matrix of the embedded Markov chain $\{J_n\}$. Furthermore, the sojourn time cumulative distribution in state i , is denoted as:

$$H_i(t) = P(T_{n+1} - T_n \leq t | J_n = i) = \sum_{j \in \mathbb{E}} \mathcal{Q}_{ij}(t). \quad (1)$$

The conditional distribution function of the waiting time in each state i , given that the next occupied state is j , is known as:

$$F_{ij}(t) = P(T_{n+1} - T_n \leq t \mid J_n = i, J_{n+1} = j) = \begin{cases} \frac{Q_{ij}(t)}{p_{ij}} & \text{if } p_{ij} \neq 0 \\ U_1(t) & \text{if } p_{ij} = 0, \end{cases}$$

where $U_1(\cdot)$ is the unitary step function. We denote the Markov renewal matrix by $\mathbf{M}(t) = [M_{ij}(t)]_{i,j \in \mathbb{E}}$, where $M_{ij}(t) = E[N_{ij}(t)]$, with $N_{ij}(t)$ as the number of visits to the state j within $[0, t]$ given that $X(0) = i$. The Laplace-Stieltjes of a measure $A(\cdot)$ will be denoted by $\bar{A}(s) = \int_0^\infty e^{-st} dA(t)$.

In the case of finite state space \mathbb{E} , from Janssen and Manca (2006):

$$\bar{\mathbf{M}}(s) = (I - \bar{\mathbf{Q}}(s))^{-1} \quad (2)$$

By the random variable $T_1(j \mid i)$ we denote the first time to enter the state j from the initial state $i \in \mathbb{E}$, with distribution function $G_{ij}(\cdot)$, and the corresponding Laplace-Stieltjes transform:

$$\bar{G}_{ij}(s) = \begin{cases} \frac{\bar{M}_{ij}(s)}{\bar{M}_{jj}(s)}, & i \neq j \\ 1 - \frac{1}{\bar{M}_{jj}(s)}, & i = j \end{cases}, \quad (3)$$

also, $\mu_{ik} = \int_0^\infty t G_{ik}(dt)$. By using the Markov renewal matrix one can obtain transition probabilities of the SMP, $P_{ij}(t) = P(X(t) = j \mid X(0) = i)$, as follows:

$$P_{ij}(t) = \int_0^t M'_{ij}(u) (1 - H_j(t-u)) du, \quad (4)$$

For $t \in [0, \infty)$, the age process or current life is denoted by $U(t) = t - T_{N(t)}$, and $V(t) = T_{N(t)+1} - t$, denotes the corresponding excess life. During the paper we will use the abbreviations $\mathbf{X}, \mathbf{U}, \mathbf{V}$ for the processes $\mathbf{X} = \{X(t), t \geq 0\}$, $\mathbf{U} = \{U(t), t \geq 0\}$ and $\mathbf{V} = \{V(t), t \geq 0\}$, frequently.

The following lemma helps us in deriving the conditional distribution of the excess life $V(t)$, $t \geq 0$, given that $U(t) = u$, $X(t) = j$, (Soltani, Khorshidian, and Ghafaripour (2010)).

Lemma1: Let \mathbf{X} be a SMP with \mathbf{U} and \mathbf{V} as the corresponding current and excess life processes respectively, then:

$$P(V(t) > v \mid U(t) = u, X(t) = j, X(0) = i) = \frac{1 - H_j(u+v)}{1 - H_j(u)}, \quad u \leq t, v > 0,$$

whenever $M_{ij}(s)$ is strictly increasing in $s = t - u$.

From Janssen and Manca (2006, 175), it is known that the joint distribution of current and the excess lives is:

$$\begin{aligned} \bar{F}_{ij, U(t), V(t)}(u, v) &= P(U(t) > u, V(t) > v, X(t) = j \mid X(0) = i) \\ &= \int_0^{t-u} [1 - H_j(t+v-s)] dM_{ij}(s) \end{aligned} \quad (5)$$

Since $M_{ij}(s)$ assumes no mass at zero in $i \neq j$, but for $i = j$ it assumes the mass $M_{ii}(0) = 1$, therefore Equation (5) becomes:

$$\begin{aligned} \bar{F}_{ij, U(t), V(t)}(u, v) &= \int_{0+}^{t-u} (1 - H_j(t+v-s)) dM_{ij}(s) \\ &\quad + \delta_{ij} (1 - H_j(t+v)) \end{aligned} \quad (6)$$

where δ_{ij} is the Kronecker delta.

The joint distribution of the current life, excess life and the next successive state $J_{N(t)+1}$, is:

$$\begin{aligned} \bar{F}_{ijk, U(t), V(t)}(u, v) &= P(U(t) > u, V(t) > v, X(t) = j, J_{N(t)+1} = k \mid X(0) = i) \\ &= \int_0^{t-u} p_{jk} (1 - F_{jk}(t+v-s)) dM_{ij}(s) \\ &= \int_{0+}^{t-u} p_{jk} (1 - F_{jk}(t+v-s)) dM_{ij}(s) \\ &\quad + \delta_{ij} p_{jk} (1 - F_{jk}(t+v)) M_{ij}(0) \end{aligned} \quad (7)$$

The following lemma is an effective tool in calculating the desired state-dependent measures.

Lemma 2: Let \mathbf{X} be a SMP with \mathbf{U} and \mathbf{V} as its current and excess life processes respectively. Also, let $q_{jk}(\cdot)$ and $h_j(\cdot)$ denote the densities of $Q_{jk}(\cdot)$ and $H_j(\cdot)$. Then, for the conditional probability $\psi_{jk}(\cdot, \cdot, \cdot)$ defined as:

$$\psi_{jk}(t, u, v) = P(X(t+v) = k \mid X(t) = j, U(t) = u, V(t) = v, X(0) = i),$$

we have:

$$\psi_{jk}(t, u, v) = \frac{q_{jk}(u+v)}{h_j(u+v)} \quad u \leq t, v > 0,$$

whenever $M_{ij}(s)$ is strictly increasing in $s = t - u$.

Theorem 1: Let $\mathbf{X} = \{X(t), t \geq 0\}$ denotes an irreducible positive recurrent SMP with state space $\mathbb{E} = \mathbb{E}_1 \cup \mathbb{E}_2$. Consider \mathbf{U} and \mathbf{V} as the corresponding current and excess life processes. Then given $B(t)$, the mean of first hitting time to the set \mathbb{E}_1 after the time epoch $t \geq 0$, becomes:

$$\begin{aligned} \mathcal{M}_{\mathbb{E}_1}(X(t), U(t)) &= \int_0^\infty v F_{V(t)|B(t)}(dv) (8) \\ &= \sum_{i \in \mathbb{E}_2} E(S_{\mathbb{E}_1}^0 \mid X(0) = i) \int_0^\infty \psi_{X(t), i}(t, u, v) F_{V(t)|B(t)}(dv). \end{aligned}$$

where, $F_{V(t)|B(t)}(\bullet)$ is the conditional distribution of the excess life $V(t)$, $t \geq 0$, given $B(t)$. It is worth noting that in many of semi-Markov models there is only one state of particularly importance, therefore it is more appropriate to present the above result for the case that the set \mathbb{E}_1 is a singleton.

Remark 1: Under the assumptions of Theorem 1, for the case $\mathbb{E}_1 = \{k\}$, we denote $\mu_{ik} = E(S_{\{k\}}^0 \mid X(0) = i)$, as the expected first hitting time to the state k from state i . Another measure of interest is the probability of not visiting \mathbb{E}_1 until an specified leading time s , by having information about the system at current time t .

Theorem 2: Under the assumptions of Theorem 1, the probability of not visiting the set \mathbb{E}_1 until a certain leading time s given $B(t)$, becomes:

$$\begin{aligned} \phi_{\mathbb{E}_1}(s; X(t), U(t)) &= \bar{F}_{V(t)|B(t)}(s) \\ &+ \sum_{i \in \mathbb{E}_2} \int_0^s P(S_{\mathbb{E}_1}^0 > s-v \mid X(0) = i) \psi_{X(t), i}(t, u, v) F_{V(t)|B(t)}(dv) \end{aligned}$$

where $\bar{F}_{V(t)|B(t)}(s) = 1 - F_{V(t)|B(t)}(s)$.

Theorem 3: Under the assumptions of Theorem 1, the probability that the process be in the set \mathbb{E}_2 at a certain leading time s given $B(t)$, becomes:

$$\begin{aligned} \phi_{\mathbb{E}_2}(s; X(t), U(t)) &= \bar{F}_{V(t)|B(t)}(s) \\ &+ \sum_{i \in \mathbb{E}} \int_0^s \sum_{j \in \mathbb{E}_2} P(X(s-v) = j \mid X(0) = i) \psi_{X(t), i}(t, u, v) F_{V(t)|B(t)}(dv) \end{aligned}$$

where $\bar{F}_{V(t)|B(t)}(s) = 1 - F_{V(t)|B(t)}(s)$.

3 Application in a Repairable System

Reliability function is the most frequently used measure in life data analysis and reliability engineering which is defined as $R(t) = P(T > t)$, where T denotes the lifetime of a system. It gives the probability that the system operates continuously during $[0, t]$ in a qualified condition. Another customary measure in reliability theory and survival analysis is the MRL function. This function has an important role in inspecting the ageing properties of the different mechanisms and at a specified time epoch t is defined as $m(t) = E(T - t \mid T > t)$.

In reliability theory, the probability that a system is operational at a given time is called availability. This measure has a special position in study of continuously operating systems and is defined as $A(t) = P(X(t) = j, j \in U)$, where U is up states.

In practice, sometimes there exist certain information for some small period just before the time epoch t , and we want to know about $m(t)$, $R(t)$ and $A(t)$ by using this information. For example, consider a k out of n system, the inspector arrives at time t and she/he wants to know about some of the system. Moreover suppose that she/he observes that at time t the system is working with exactly $k \leq s \leq n$ components, and also knows the time instant of the last failure. For such situations we will consider the state-dependent version of (t) , $R(t)$ and $A(t)$ which can be defined as follows:

$$SDM(X(t), U(t)) = E(S_{\mathbb{E}_1}^t \mid B(t)) = \mathcal{M}_{\mathbb{E}_1}(X(t), U(t)),$$

$$SDR(s; X(t), U(t)) = P(S_{\mathbb{E}_1}^t > s \mid B(t)) = \phi_{\mathbb{E}_1}(s; X(t), U(t)),$$

$$SDA(s; X(t), U(t)) = P(X(t+s) = j, j \in \mathbb{E}_2 \mid B(t)) = \phi_{\mathbb{E}_2}(s; X(t), U(t))$$

where \mathbb{E}_1 , \mathbb{E}_2 are considered to be down and up states of the semi-Markov system, respectively. In this section a two-unit cold standby system with one repairman will have been considered as a three state SMP and the associated state-dependent measures is studied in details.

4 Semi-Markov Standby Systems

The following model is a special case of the system that have been studied by Fathizadeh and Khorshidian (2016).

1. The system consists of two similar units which operate in cold standby configuration, a switch and a repair facility.
2. Whenever the operating unit fails, the standby unit is switched to operate with a switching device. It is assumed that the switch is perfect.
3. The system fails if the operating unit fails and the repair job has not been finished yet. In this case the whole failed system will be replaced by a new identical one.
4. The lifetime of the operating unit is indicated by $\eta_1 \sim \exp(\theta)$. The repair periods of units is denoted by $\eta_2 \sim \exp(\alpha)$. Replacing time of the failed system is $\eta_3 \sim \exp(\beta)$.
5. All random variables assumed to be independent.

In order to construct a semi-Markov reliability model and derive the associated renewal kernel, we will introduce the following states:

1. The system is failed;
2. The failed unit is under repairing and the other is operating;
3. Both operating unit and the spare one are up.

By the process $\{X(t), t \geq 0\}$, denote the states of the system at time t . It is easy to verify that $\{X(t), t \geq 0\}$ is a semi-Markov process with state space $\mathbb{E} = \{0, 1, 2\}$ and the following kernels, see Fathizadeh and Khorshidian (2016):

$$\begin{aligned} Q_{02}(t) &= 1 - e^{-\beta t}, & Q_{10}(t) &= \frac{\theta}{\theta + \alpha} (1 - e^{-(\theta + \alpha)t}), \\ Q_{11}(t) &= 1 - e^{-\theta t} - \frac{\theta}{\theta + \alpha} (1 - e^{-(\theta + \alpha)t}), & Q_{21}(t) &= 1 - e^{-\theta t}, \end{aligned}$$

MRL of the Standby Systems

For the semi-Markov system, from Equation (1) we have

$$H_1(t) = H_2(t) = 1 - e^{-\theta t}, \quad H_0(t) = 1 - e^{-\beta t},$$

which according to Lemma 1, imply that:

$$f_{(V(t)|X(t),U(t))}(v | 1, u) = f_{(V(t)|X(t),U(t))}(v | 2, u) = \theta e^{-\theta v}$$

From Lemma 2, it yields that:

$$\begin{aligned}\psi_{21}(t, u, v) &= 1, \\ \psi_{10}(t, u, v) &= e^{-\alpha(u+v)}, \\ \psi_{11}(t, u, v) &= 1 - e^{-\alpha(u+v)}.\end{aligned}$$

We will call the Heaviside Expansion Theorem for computing the Laplace inverse of rational functions.

Lemma 3: Let $F(s)$ and $G(s)$ be polynomials where $F(s)$ has degree less than that of $G(s)$. Suppose that $G(s)$ has n distinct zeros α_k , $k = 1, 2, \dots, n$, then

$$\mathcal{L}^{-1} \left\{ \frac{F(s)}{G(s)} \right\} = \sum_{k=1}^n \frac{F(\alpha_k)}{G'(\alpha_k)} e^{\alpha_k t}.$$

Now, by using the results (4.1)-(4.2) in Fathizadeh and Khorshidian (2016), along with Equation (4) we arrive at the desired result. Altinok, Y., and D. Kolcak. 1999. An application of the semi-Markov model for earthquake occurrences in North Anatolia. Turkey. The Journal of the Balkan Geophysical Society. 2(4): 90-99.

References

- Altinok, Y., and D. Kolcak. (1999). An application of the semi-Markov model for earthquake occurrences in North Anatolia. Turkey. *The Journal of the Balkan Geophysical Society*. 2(4): 90-99.
- 2006)]Asadi, M., and I. Bayramoglu:(2006) Asadi, M., and I. Bayramoglu. (2006). The mean residual life function of k-out-of-n structure at the system level. *IEEE Transactions on Reliability*. 55: 314-318.
- Asadi, M., and S. Goliforushani. (2008). On the mean residual life function of coherent systems. *IEEE Transactions on Reliability*. 57:574-580.
- Bieth, B., L. Hong, and J. Sarkar. (2009). A standby system with two types of repair persons. *Applied Stochastic models in Business and Industry*. 26: 577-594.
- Cinlar, E. (1969). Markov Renewal Theory. *Adv. Appl. Prob.* 1, 123-187.
- Cinlar, E. (1975). *Introduction to Stochastic Processes*. Prentice-Hall, Inc.
- Do, P., and Ch. Be' renguer. (2021) Residual life-based importance measures for predictive maintenance decision-making. Proceedings of the Institution of Mechanical Engineers, *Part O: Journal of Risk and Reliability*. 236(1):98-113.

- Eryilmaz, S. (2013). On Mean Residual Life of discrete Time Multi-state systems. *Quality Technology and Quantitative Management*. 10:241-249.
- Fathizadeh, M., and K. Khorshidian. 2016. An alternative approach to reliability analysis of cold standby systems. *Communications in Statistics – theory and methods*. 45(21): 6471-6480.
- Janssen, J., and R. Manca. (2006). *Applied semi-Markov processes*. Springer.
- Limnios, N. 2012. Reliability measures of semi-Markov systems with general state space. *Methodology and Computing in Applied Probability* 14(4): 895-917.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



On the double-log skew-normal distribution and its application

Abbas Mahdavi*, Hossein Negarestani

Department of Statistics, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran.

Abstract

Data confined to the interval $(0, 1)$ are frequently analyzed using various distributions in a range of practical applications. However, many existing studies fail to adequately consider aspects such as skewness, kurtosis, and heavy-tailed behavior in this data. This research introduces an alternative skew-normal type distribution that operates within a bounded interval, created by merging the characteristics of skew-normal distributions with the double-log transformation. The resulting model, with its enhanced capabilities for skewness and boundedness, offers a flexible and effective framework for modeling rates and proportions. We have implemented an EM-type algorithm to reliably estimate the parameters of the proposed model. To demonstrate the efficacy of our method, we carried out an experiment that included an analysis of one actual dataset. The findings emphasize the adaptability and precision of our proposed model when compared to conventional models.

Keywords: EM-type algorithms; Mixture models; Double-log-normal distribution, Skew-normal distribution

Mathematics Subject Classification (2020): 62E10, 62F10.

1 Introduction

The beta distribution is widely used for modeling percentages and proportions within the interval $(0, 1)$. However, various alternative distributions have been suggested in the literature to provide more flexible modeling options than the beta distribution. Among these alternatives are the Topp–Leone distribution, Kumaraswamy distribution, arcsine distribution, generalized beta type distribution, standard two-sided power distribution, McDonald arcsine distribution, exponentiated Topp–Leone distribution, generalized beta distribution, and unit Lindley distribution. The skew-normal (SN) distribution, denoted by $SN(\mu, \sigma, \lambda)$ and introduced by [Azzalini \(1985\)](#), has shown its great flexibility in modeling data with asymmetric behaviors by the use of fewer components in the fitting of mixture models. The probability density function (pdf) of SN distribution is expressed by

$$f_{SN}(x; \mu, \sigma, \lambda) = \frac{2}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) \Phi\left(\lambda \frac{x - \mu}{\sigma}\right), \quad x \in \mathbb{R}, \quad (1)$$

*Corresponding author, a.mahdavi@vru.ac.ir

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the pdf and cumulative density function (cdf) of standard normal distribution, $\mu \in \mathbb{R}$, $\sigma > 0$ and $\lambda \in \mathbb{R}$ represent the location, scale and shape parameters, respectively. Setting $\lambda = 0$, the classical normal distribution with mean μ and variance σ^2 is obtained. It is noteworthy to mention that the SN distribution can be expressed by the following hierarchical stochastic representation (Lin *et al.*, 2007):

$$\begin{aligned} Y|\tau &\sim N(\mu + \delta(\lambda)\tau, (1 - \delta^2(\lambda))\sigma^2), \\ \tau &\sim TN(0, \sigma^2)I_{(0, \infty)}, \end{aligned} \quad (2)$$

where $\delta(\lambda) = \lambda/\sqrt{1 + \lambda^2}$. The notation $TN(\mu, \sigma^2)I_{\mathbb{A}}$ represents a doubly truncated normal distribution confined within the interval $\mathbb{A} = \{a_1 < x < a_2\}$, and $I_{\mathbb{A}}$ is an indicator function of set \mathbb{A} .

The main objective of this study is to introduce a new statistical distribution, called as the double-log skew-normal (DLSN) distribution, which is defined on the interval $(0, 1)$ and generated using the double-log transformation, defined as $Y \stackrel{d}{=} \exp(-\exp(X))$, where $X \sim SN(\mu, \sigma, \lambda)$. The rest of this paper is organized as follows. In Section 2, we introduce with the DLSN distribution. Next, in Section 3, we adopt an EM-type algorithm for the fitting of the DLSN model. A real data analysis is illustrated in Section 4.

2 The model

The random variable Y is said to follow a DLSN distribution, denoted by $DLSN(\mu, \sigma, \lambda)$, with support $(0, 1)$ and shape parameters $\mu \in \mathbb{R}$, $\sigma > 0$ and $\lambda \in \mathbb{R}$, if its pdf is

$$f_{DLSN}(y; \mu, \sigma^2, \lambda) = \frac{-2}{\sigma y \log y} \phi\left(\frac{\log(-\log y) - \mu}{\sigma}\right) \Phi\left(\lambda \frac{\log(-\log y) - \mu}{\sigma}\right), \quad 0 < y < 1. \quad (3)$$

Note that double-log normal (DLN) distribution is recovered from (3) if $\lambda = 0$ as follows:

$$f_{DLN}(y; \mu, \sigma^2) = \frac{-1}{y \log y \sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(\log(-\log y) - \mu)^2\right\}, \quad 0 < y < 1, \quad (4)$$

Figure 1 visually illustrates the shapes of several DLSN densities alongside their corresponding DLN densities, offering insight into the diverse characteristics that this family of distributions can exhibit. The plots vividly demonstrate that the DLSN distribution encompasses a wide range of shapes characterized by distinct directional skewness and kurtosis properties, which are dictated by the specific parameter values chosen for each density.

It is straightforward to demonstrate that the DLSN distribution admits the following stochastic representation:

$$\begin{aligned} Y|\tau &\sim DLN(\mu + \delta(\lambda)\tau, (1 - \delta^2(\lambda))\sigma^2), \\ \tau &\sim TN(0, \sigma^2)I_{(0, \infty)}. \end{aligned} \quad (5)$$

From (5) the joint pdf of Y and τ is given by

$$\begin{aligned} f_{Y, \tau}(y, \tau) &= f_{\tau}(\tau) f_{Y|\tau}(y) \\ &= \frac{-1}{\pi\sigma^2 y \log y \sqrt{1 - \delta^2(\lambda)}} \exp\left\{\frac{-1}{2(1 - \delta^2(\lambda))\sigma^2} \left(\log(-\log y) - \mu - \delta(\lambda)\tau\right)^2\right. \\ &\quad \left. - \frac{1}{2\sigma^2}\tau^2\right\}, \quad 0 < y < 1, \tau > 0. \end{aligned} \quad (6)$$

It follows from (6) that the conditional distribution of τ given Y is

$$\begin{aligned} f_{\tau|Y=y}(\tau) &= \frac{f_{Y, \tau}(y, \tau)}{f_Y(y)} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}\sqrt{1 - \delta^2(\lambda)}} \frac{\exp\left\{\frac{-1}{2(1 - \delta^2(\lambda))\sigma^2} \left(\tau - \delta(\lambda)(\log(-\log y) - \mu)\right)^2\right\}}{\Phi\left(\lambda \frac{\log(-\log y) - \mu}{\sigma}\right)}, \end{aligned}$$

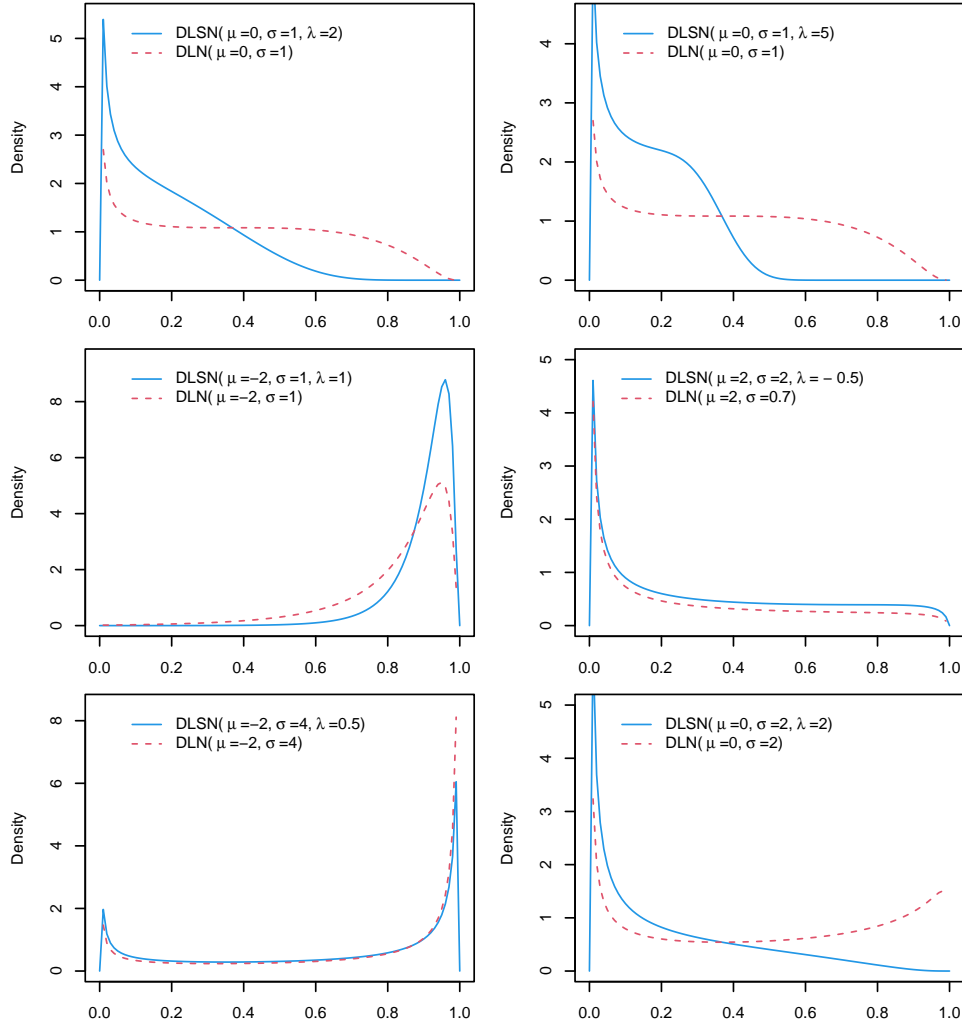


Figure 1: Different density shapes of DLSN distribution.

and using the fact that $\lambda = \delta(\lambda)/\sqrt{1 - \delta^2(\lambda)}$, it yields to

$$\tau \mid Y = y \sim TN\left(\delta(\lambda)(\log(-\log y) - \mu), \sigma\sqrt{1 - \delta^2(\lambda)}\right). \quad (7)$$

In the following subsection, we propose an EM-type algorithm for determining the ML estimates of the parameters of $DLSN(\mu, \sigma, \lambda)$ based on the joint density defined in (6).

3 Parameter estimation

Let $\mathbf{y} = (y_1, \dots, y_n)^\top$ denote the observed data, $\boldsymbol{\tau} = (\tau_1, \dots, \tau_n)^\top$ be the latent variable and $\mathbf{y}_c = (\mathbf{y}^\top, \boldsymbol{\tau}^\top)^\top$ be the complete data. From (6), the log-likelihood function of $\boldsymbol{\theta} = (\mu, \sigma, \lambda)^\top$ corresponding to the complete-data \mathbf{y}_c , excluding additive constants and terms that do not involve the parameters of the model, is given by

$$\begin{aligned} & \ell_c(\boldsymbol{\theta} \mid \mathbf{y}_c) \\ &= -\frac{1}{2} \sum_{i=1}^n \left\{ \frac{1}{\sigma^2(1 - \delta^2(\lambda))} \left(\tau_i^2 - 2\delta(\lambda)\tau_i(\log(-\log y_i) - \mu) + (\log(-\log y_i) - \mu)^2 \right) \right. \\ & \quad \left. + 2\log \sigma^2 + \log(1 - \delta^2(\lambda)) \right\}. \end{aligned} \quad (8)$$

Given the current estimate of θ in the k th iteration, $\hat{\theta}^{(k)} = (\hat{\mu}^{(k)}, \hat{\sigma}^{(k)}, \hat{\lambda}^{(k)})^\top$, the E-step of the EM algorithm involves the evaluation of the Q -function, a conditional expectation of (8) given the observed data. It requires the conditional expectations derived from the conditional distribution as expressed in (7). By Lemma 2 of Lin *et al.* (2007), we obtain

$$\begin{aligned}\hat{\tau}_i^{(k)} &= E(\tau_i \mid y_i, \hat{\theta}^{(k)}) \\ &= \delta(\hat{\lambda}^{(k)}) (\log(-\log y_i) - \hat{\mu}^{(k)}) + \hat{\sigma}^{(k)} \sqrt{1 - \delta^2(\hat{\lambda}^{(k)})} \frac{\phi(\hat{\lambda}^{(k)} \hat{\eta}_i^{(k)})}{\Phi(\hat{\lambda}^{(k)} \hat{\eta}_i^{(k)})},\end{aligned}\quad (9)$$

and

$$\begin{aligned}\hat{s}_i^{(k)} &= E(\tau_i^2 \mid y_i, \hat{\theta}^{(k)}) \\ &= \delta^2(\hat{\lambda}^{(k)}) (\log(-\log y_i) - \hat{\mu}^{(k)})^2 + \hat{\sigma}^{2(k)} (1 - \delta^2(\hat{\lambda}^{(k)})) \\ &\quad + \delta(\hat{\lambda}^{(k)}) (\log(-\log y_i) - \hat{\mu}^{(k)}) \hat{\sigma}^{(k)} \sqrt{1 - \delta^2(\hat{\lambda}^{(k)})} \frac{\phi(\hat{\lambda}^{(k)} \hat{\eta}_i^{(k)})}{\Phi(\hat{\lambda}^{(k)} \hat{\eta}_i^{(k)})},\end{aligned}\quad (10)$$

where $\hat{\eta}_i^{(k)} = (\log(-\log y_i) - \hat{\mu}^{(k)}) / \hat{\sigma}^{(k)}$ and $\delta(\hat{\lambda}^{(k)}) = \hat{\lambda}^{(k)} / \sqrt{1 + \hat{\lambda}^{2(k)}}$.

Then the objective function $Q(\theta \mid \hat{\theta}^{(k)})$ is

$$\begin{aligned}Q(\theta \mid \hat{\theta}^{(k)}) &= -\frac{1}{2} \sum_{i=1}^n \left\{ \frac{1}{\hat{\sigma}^{2(k)} (1 - \delta^2(\hat{\lambda}^{(k)}))} \left(\hat{s}_i^{(k)} - 2\delta(\hat{\lambda}^{(k)}) \hat{\tau}_i^{(k)} (\log(-\log y_i) - \hat{\mu}^{(k)}) \right. \right. \\ &\quad \left. \left. + (\log(-\log y_i) - \hat{\mu}^{(k)})^2 \right) + 2 \log \hat{\sigma}^{2(k)} + \log (1 - \delta^2(\hat{\lambda}^{(k)})) \right\}.\end{aligned}\quad (11)$$

The CM-step then conditionally maximizes $Q(\theta \mid \hat{\theta}^{(k)})$ with respect to θ , resulting a new estimate $\hat{\theta}^{(k+1)}$, as outlined below:

CMQ-step 1: Update $\hat{\mu}^{(k+1)}$ by

$$\hat{\mu}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \left(\log(-\log y_i) - \delta(\hat{\lambda}^{(k)}) \hat{\tau}_i^{(k)} \right); \quad (12)$$

CMQ-step 2: Fixing $\mu = \hat{\mu}^{(k+1)}$, we can then update $\hat{\sigma}^{2(k)}$ by

$$\begin{aligned}\hat{\sigma}^{2(k+1)} &= \frac{1}{2n(1 - \delta^2(\hat{\lambda}))} \left\{ \sum_{i=1}^n \hat{s}_i^{(k)} - 2\delta(\hat{\lambda}) \sum_{i=1}^n (\log(-\log y_i) - \hat{\mu}^{(k+1)}) \hat{\tau}_i^{(k)} \right. \\ &\quad \left. + \sum_{i=1}^n (\log(-\log y_i) - \hat{\mu}^{(k+1)})^2 \right\};\end{aligned}\quad (13)$$

CMQ-step 3: Update $\hat{\lambda}^{(k+1)}$ by obtaining the root of the following equation respect to the λ :

$$\begin{aligned}&n\hat{\sigma}^{2(k+1)}\delta(\lambda)(1 - \delta^2(\lambda)) + (1 + \delta^2(\lambda)) \sum_{i=1}^n \left(\log(-\log y_i) - \hat{\mu}^{(k+1)} \right) \hat{\tau}_i^{(k)} \\ &- \delta(\lambda) \sum_{i=1}^n \left(\hat{s}_i^{(k)} + (\log(-\log y_i) - \hat{\mu}^{(k+1)})^2 \right) = 0.\end{aligned}\quad (14)$$

Updating $\hat{\lambda}^{(k+1)}$ requires performing a one-dimensional search to find the root of λ . An alternative strategy involves maximizing the restricted actual log-likelihood function, leading to the next step in the ECME algorithm known as the CML step.

CML-step: Update $\hat{\lambda}^{(k+1)}$ by optimizing the following constrained log-likelihood function:

$$\hat{\lambda}^{(k+1)} = \arg \max_{\lambda} \sum_{i=1}^n \log f_{DLSN}(y_i; \hat{\mu}^{(k+1)}, \hat{\sigma}^{(k+1)}, \lambda).$$

Table 1: Summary results from fitting various models to the Anxiety data.

Parameter	Beta	KU	GBI	McA	SN	DLN	DLSN
$\hat{\mu}(\hat{\alpha})$	6.01	4.64	3.55	10.40	1.16	1.21	1.28
$\hat{\sigma}(\hat{\beta})$	3.28	4.28	13.24	0.55	1.22	1.53	0.45
$\hat{\delta}$	—	—	15.17	41.11	—	-0.87	0.51
$\hat{\nu}$	—	—	—	—	—	—	1.75
$\hat{\ell}$	94.23	97.32	101.11	101.07	99.91	93.42	102.71
AIC	-184.45	-190.64	-196.21	-196.13	-193.81	-182.84	-199.42
BIC	-178.01	-184.19	-186.55	-186.47	-184.15	-176.40	-189.76
K-S statistic	0.079	0.064	0.038	0.034	0.044	0.040	0.029

The iterative nature of the algorithm used above necessitates the repetition of iterations until a convergence criterion is satisfied. This convergence criterion is typically based on the relative change in the log-likelihood function, such as

$$\left\| \frac{\ell(\hat{\theta}^{(k+1)})}{\ell(\hat{\theta}^{(k)})} - 1 \right\|,$$

where $\ell(\theta) = \sum_{i=1}^n \log f_{DLSN}(y_i; \theta)$. To achieve convergence, it is necessary to set a threshold for this relative change that is considered small enough to indicate convergence. In this case, a threshold of 10^{-6} is chosen as the desired level of relative change.

4 Data analysis

The Better Life Index (BLI) data set, measured in the year 2015, is used to demonstrate the usefulness of DLSN distribution. The data set can be found in <https://stats.oecd.org/index.aspx?DataSetCode=BLI2015>. The BLI data set consists of 11 indicator and 24 variables and it is used to classify the OECD (Organisation for Economic Co-operation and Development) countries as well as Brazil and Russia. Here, we use an indicator that is entitled “Employment rate” as data set.

To facilitate a comprehensive comparison, we fit several statistical distributions to these data sets, including

a) Beta distribution :

$$f(y : \alpha, \beta) = \frac{1}{B(\alpha, \beta)} y^{\alpha-1} (1-y)^{\beta-1}, \quad \alpha > 0, \beta > 0,$$

where $B(\alpha, \beta)$ is the beta function.

b) Kumaraswamy (KU) distribution:

$$f_{KU}(y : \alpha, \beta) = \alpha \beta y^{\alpha-1} (1-y^\alpha)^{\beta-1}, \quad \alpha > 0, \beta > 0.$$

c) Generalized beta of the first kind (GBI):

$$f_{GBI}(y : \alpha, \beta, \delta) = \frac{\delta}{B(\alpha\delta^{-1}, \beta)} y^{\alpha-1} (1-y^\delta)^{\beta-1}, \quad \alpha > 0, \beta > 0, \delta > 0.$$

d) McDonald arcsine (McA) distribution :

$$f_{McA}(y : \alpha, \beta, \delta) = \frac{\delta}{B(\alpha\delta^{-1}, \beta) \pi \sqrt{y-y^2}} \left[\frac{2}{\pi} \arcsin(\sqrt{y}) \right]^{\alpha-1} \times \left(1 - \left[\frac{2}{\pi} \arcsin(\sqrt{y}) \right]^\delta \right)^{\beta-1}, \quad \alpha > 0, \beta > 0, \delta > 0.$$

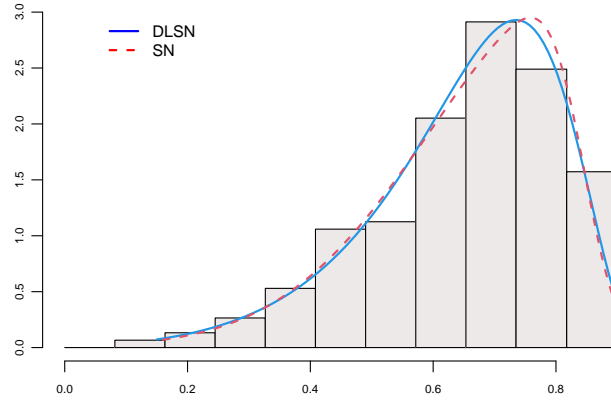


Figure 2: Histogram of the BLI samples. The lines represent the fitted DLSN and SN densities.

e) The SN distribution given in (1).

Furthermore, the reduced DLN distribution is fitted as the sub-models of the proposed DLSN distribution using the EM-algorithm given in Section 3. Given that the competing models vary in the number of parameters they encompass, we employ standard statistical criteria to assess their goodness-of-fit. Specifically, we utilize the Akaike Information Criterion (AIC) (Akaike, 1973) and the Bayesian Information Criterion (BIC) (Schwarz, 1978). Through the model comparison considering the penalized criteria, the Kolmogorov–Smirnov (K-S) test (Smirnov, 1948) is considered to evaluate the goodness of fit of all models.

The summary results of the fitted models are detailed in Table 1. The findings indicate that the proposed DLSN distribution demonstrates a significantly superior fit compared to competing models, as evidenced by its notably lower values of both AIC and BIC. Besides, the smallest K-S test statistic (0.029) suggests that the DLSN model yields the shortest distance between the empirical and parametric distributions of experimental data. To illustrate these findings, Figure 2 presents a histogram of the BLI data overlaid with the fitted DLSN and SN distributions. A closer examination reveals that the DLSN distribution more accurately captures the heavy tail and leptokurtic features inherent in the data when compared to the other models.

Discussion and Results

In this study, we introduced and examined a novel flexible bounded distribution defined on the interval $(0, 1)$, integrating features from the SN distribution and the double-log function. The DLSN distribution we proposed is particularly effective at modeling bounded data with significant skewness, making it a valuable tool for researchers dealing with such datasets. Moreover, we developed an ECME algorithm for estimating the model parameters. To evaluate the model's performance, we conducted an analysis of a real dataset, comparing our findings against several established models in the literature. The results suggest that our proposed model serves as an alternative to traditional bounded distributions, especially for modeling rates and proportions. The methodologies and EM-type algorithms developed in this study have the potential to be adapted for a multivariate version of the DLSN model. We are currently exploring this extension and look forward to sharing our findings in a future publication..

References

- Akaike, H. (1998), Information theory and an extension of the maximum likelihood principle, *In Second International Symposium on Information Theory (Eds., B.N. Petrov and F. Csaki),* BNPBF Csaki Budapest: Academiai Kiado.
- Azzalini, A. (1985), A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, 171-178.
- Lin, T. I., Lee, J. C. and Yen, S. Y. (2007), Finite mixture modelling using the skew normal distribution, *Statistica Sinica*, 909-927.
- Schwarz, G. (1978), Estimating the dimension of a model, *The Annals of Statistics*, 461-464.
- Smirnov, N. (1948), Table for estimating the goodness of fit of empirical distributions, *The Annals of Mathematical Statistics*, 19(2), 279-281.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Modeling Electricity Demand: A Comparative Study of Univariate, Multivariate, and Functional SSA

Rahim Mahmoudvand^{*1,2}

¹Department of Statistics, Faculty of Science, Bu-Ali Sina University, Iran and

²Department of Economic and Business Sciences, University of Cagliari, Italy

Abstract

Accurate modeling of patterns in electricity demand is essential for effective energy forecasting and grid management. This study presents a comparative analysis of Singular Spectrum Analysis (SSA) in its univariate (SSA), multivariate (MSSA), and functional (FSSA) forms, applied to half-hourly electricity demand data from the state of Victoria, Australia. The data, exhibiting pronounced daily and weekly seasonality, is reshaped into univariate, multivariate, and functional time series formats to align with the respective SSA frameworks.

Each method is evaluated across a range of window lengths and grouping strategies, with performance assessed using the Root Mean Square Error (RMSE) of the reconstructed series. Our results show that Univariate SSA consistently yields the most accurate reconstructions, followed by MSSA, while FSSA exhibits higher reconstruction error in this setting. Additionally, we investigate the seasonal components extracted by each method and find that while MSSA and FSSA effectively capture weekly periodicity, only Univariate SSA is capable of detecting both daily and sub-daily (e.g., hourly) patterns—albeit requiring careful selection of window length.

These findings highlight the importance of choosing the appropriate SSA variant based on the frequency characteristics of the data and the specific goals of analysis, especially when dealing with high-resolution time series.

Keywords: Singular Spectrum Analysis (SSA); Functional Time Series; Electricity Demand Forecasting; Seasonality; Time Series Decomposition.

Mathematics Subject Classification (2020): 62M10, 62H12, 62P30.

1 Introduction

Electricity demand exhibits pronounced seasonal patterns driven by climatic conditions, economic activity, and social behavior. Accurately modeling and forecasting this seasonality is

*Corresponding author, r.mahmoudvand@basu.ac.ir

critical for energy planning, infrastructure development, and operational reliability in power systems. Traditional time series models have long been employed for this purpose; however, recent advances in multivariate and functional time series (FTS) methods (Hyndman and Athanasopoulos (2018) and Amato and *et al.* (2025)) offer new avenues for capturing complex temporal dependencies and structural variations in electricity consumption data.

The complex nature of time series, which often violates the assumptions of traditional time series analysis methods, motivates the introduction of new approaches and the development of the literature. Among these, we focus on Singular Spectrum Analysis (SSA), a non-parametric and model-free method for time series analysis that has been developed for univariate (Golyandina and *et al.* (2001)), multivariate (Hassani and Mahmoudvand (2013)), and functional (Haghighin and *et al.* (2021)) time series.

This paper presents a comparative study of univariate, multivariate, and functional SSA for modeling the seasonal dynamics of electricity demand. Each framework provides distinct advantages: univariate models are computationally efficient and interpretable, multivariate models capture interdependencies among related series, and FTS models leverage the continuous nature of daily or intra-daily load profiles to preserve functional characteristics. By applying these techniques to real-world electricity demand data, we assess their relative performance in capturing seasonal structures and producing accurate forecasts.

Our study contributes to the growing body of literature on energy demand forecasting by highlighting the trade-offs and practical considerations associated with different modeling strategies. The findings offer insights for researchers and practitioners seeking robust and scalable tools for seasonal demand analysis in increasingly complex energy systems.

2 Methodology

In this section, we provide a brief introduction to the methodology of model capturing in SSA. As it has mentioned above, we have three different version of SSA that admit different object as input. However, the general algorithm for all version are the same.

2.1 SSA decomposition for univariate time series

Let $\{y_1, \dots, y_N\}$ denote a univariate time series of length N . SSA decomposition begins by transforming the time series into a structured matrix form, followed by matrix decomposition and reconstruction steps that isolate distinct components such as trend, oscillations, and noise. The overall process is outlined below.

1. **Embedding:** Choose a window length parameter L such that $2 \leq L \leq N - 1$, and define $K = N - L + 1$. Construct the trajectory (Hankel) matrix $\mathbf{Y} \in \mathbb{R}^{L \times K}$, where each column is a lagged segment of the original series:

$$\mathbf{Y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_K \\ y_2 & y_3 & \cdots & y_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & \cdots & y_N \end{bmatrix}$$

2. **Singular Value Decomposition (SVD):** Decompose the matrix \mathbf{Y} using SVD into a sum of rank-one matrices: $\mathbf{Y} = \mathbf{Y}_1 + \cdots + \mathbf{Y}_d$.
3. **Grouping:** Partition the set of elementary matrices into m disjoint groups I_1, \dots, I_m , and form grouped matrices by summing the elements within each group. In the simplest case ($m = 2$), group $I_1 = 1, \dots, r$ is associated with the signal (e.g., trend and periodicity), and $I_2 = r + 1, \dots, d$ corresponds to noise.

4. **Diagonal Averaging (Reconstruction):** Transform each grouped matrix \mathbf{Y}_{I_j} back into a time series of length N via diagonal averaging (also known as Hankelization). This step reconstructs the components of the original series.

2.2 SSA decomposition for the Multivariate Time Series

Let $\{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ where $\mathbf{y}_t \in \mathbb{R}^D$, is a D -variate time series of length N . Multivariate SSA (MSSA) follows a similar procedure to univariate SSA, with the primary difference being in the construction of the trajectory matrix. Specifically, the trajectory matrix is formed by horizontally stacking the trajectory matrices of each individual (marginal) time series.

In this case, the combined trajectory matrix \mathbf{Y} is defined as:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}^{(1)} & \mathbf{Y}^{(2)} & \dots & \mathbf{Y}^{(D)} \end{bmatrix},$$

where each $\mathbf{Y}^{(i)}$ is the trajectory matrix corresponding to the i^{th} time series. The rest of the SSA procedure (SVD, grouping, diagonal averaging, etc.) is then applied to this combined matrix.

2.3 SSA decomposition for the Functional Time Series

1. **Embedding:** Given a functional time series $\{X_t(u)\}_{t=1}^N$ in a Hilbert space \mathcal{H} , construct the trajectory matrix:

$$\mathbf{Y} = [X_1(u), X_2(u), \dots, X_L(u)] \in \mathcal{H}^L$$

where L is the window length such that $1 < L < N$.

2. **Decomposition:** Apply singular value decomposition (SVD) to the trajectory matrix:

$$\mathbf{Y} = \sum_{i=1}^L \sqrt{\lambda_i} U_i \otimes V_i$$

where λ_i are the eigenvalues, and U_i, V_i are the corresponding orthonormal singular functions.

3. **Grouping:** Partition the set of indices $\{1, 2, \dots, L\}$ into d disjoint groups I_1, I_2, \dots, I_d , according to the desired signal components (e.g., trend, seasonality, noise).
4. **Reconstruction:** For each group I_k , form the grouped approximation:

$$\mathbf{Y}^{(k)} = \sum_{i \in I_k} \sqrt{\lambda_i} U_i \otimes V_i$$

and apply diagonal averaging (also known as Hankelization) to reconstruct the functional time series component $\{\tilde{X}_t^{(k)}(u)\}_{t=1}^N$.

3 Empirical analysis

3.1 Data description

The Half-hourly electricity demand dataset, available in the `fpp2` package in R, contains electricity demand data for the state of Victoria, Australia. The dataset spans from 1 January 2014 to 31 January 2014, with measurements recorded every 30 minutes, resulting in 48 observations per day. Figure 1 illustrates three perspectives of this dataset: as a univariate time series, a multivariate time series (where each of the 48 half-hour intervals is treated as a separate variable), and as a functional time series (where each day is represented by a continuous function over time).

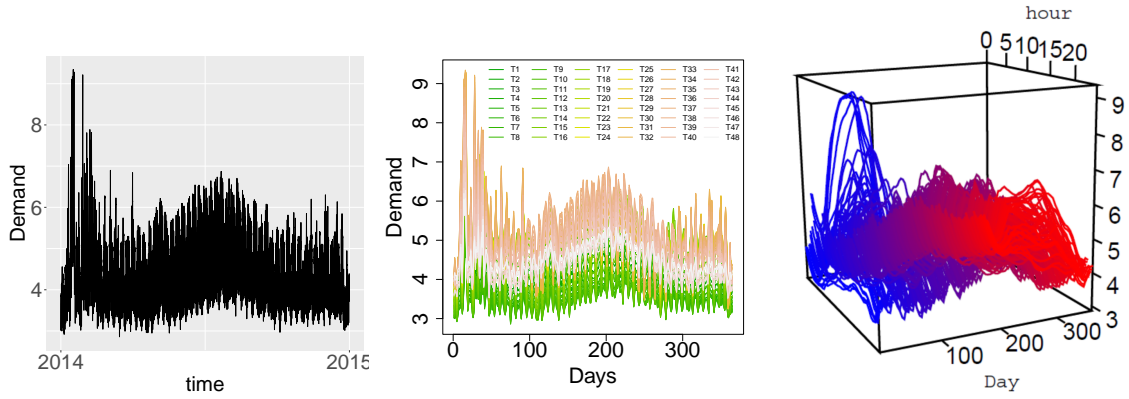


Figure 1: Univariate (left), Multivariate (middle), and Functional (right) representations of the Half-hourly electricity demand dataset.

As evident in Figure 1 and due to the intrinsic daily cycle of electricity usage, the data exhibits a strong seasonal pattern. As discussed in Golyandina and *et al.* (2001) and supported by other studies, the window length in SSA can be chosen as a multiple of the seasonal period, rather than half the series length—especially when dealing with long series where computation becomes intensive.

3.2 Model fitting

To evaluate the performance of different SSA frameworks, we applied Univariate SSA (SSA), Multivariate SSA (MSSA), and Functional SSA (FSSA) on the dataset. We experimented with window lengths ranging from 24 to 60 and grouping choices varying from 3 to 8. The RMSE for the reconstructed signals was calculated and used to compare performance across methods. Figure 2 displays the RMSE results. Among the three approaches, SSA yields the best performance, followed by MSSA. FSSA, in this case, resulted in the highest RMSE, indicating comparatively poorer reconstruction accuracy.

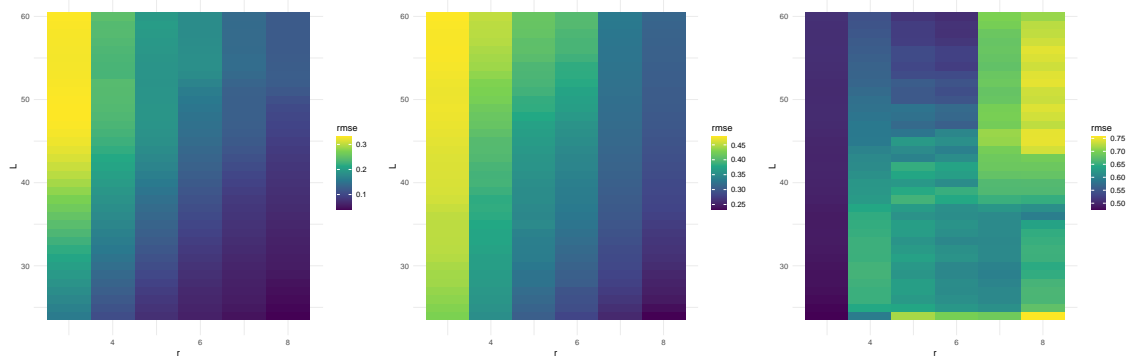


Figure 2: RMSE for fitting Univariate (left), Multivariate (middle), and Functional (right) SSA models to the Half-hourly electricity demand dataset.

3.3 Seasonality analysis

As has been established, the inputs for univariate SSA, multivariate SSA, and functional SSA differ, implying that the objective of the analysis must be clearly defined to enable a fair comparison. In this study, we focus on the fitting performance, which is one of the most

common objectives in time series modeling. In this section, we explore the seasonal components identified by the aforementioned methods.

We fixed the window length at $L = 48$, although we tested other values and obtained similar results. As illustrated in Figures 3 and 4, multivariate SSA and functional SSA reveal a clear weekly seasonality (7-day cycle). Univariate SSA, on the other hand, identifies two dominant seasonal components with periods of approximately 48 and 24, corresponding to daily and half-daily cycles, respectively.

It is important to note that the time unit in the multivariate and functional SSA representations is the day, while in the univariate SSA, it is the half-hour. Therefore, capturing a 7-day periodicity in the univariate SSA requires detecting a periodic component at lag 336. We verified that such periodicity is indeed captured when the window length exceeds 336. However, identifying it via paired eigentriple plots is challenging, so we employed the `parestimate` function from the `Rssa` package for more precise period estimation. From this perspective, multivariate SSA and functional SSA offer more interpretable insights into weekly seasonality.

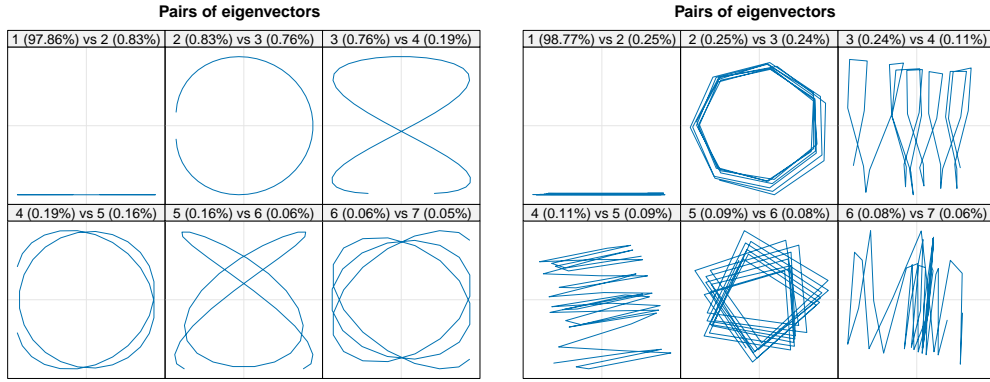


Figure 3: Paired plot of eigenvectors using the univariate SSA (left) and multivariate SSA (right).

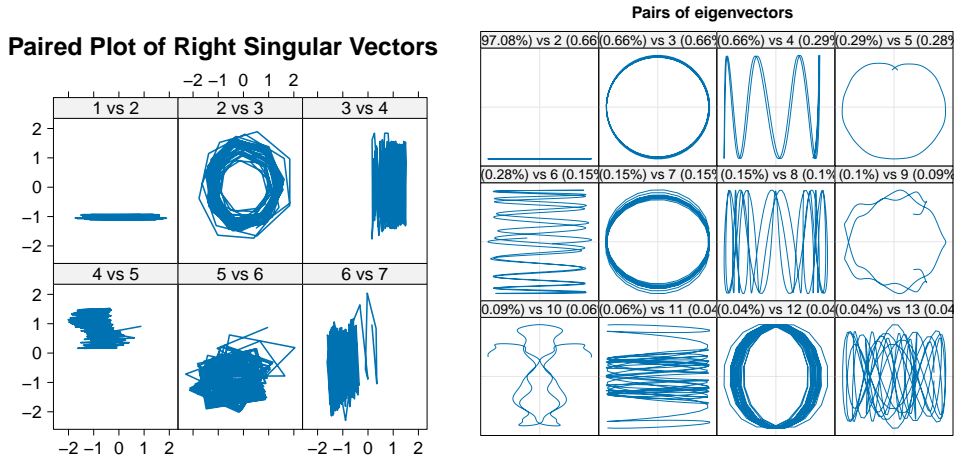


Figure 4: Paired plot of right singular vectors using the Functional SSA (left) and paired plot of eigenvectors using univariate SSA with larger window length (right).

4 Conclusion

In this study, we investigated the performance of Univariate Singular Spectrum Analysis (SSA), Multivariate SSA (MSSA), and Functional SSA (FSSA) on half-hourly electricity demand data. By exploring a range of window lengths and grouping structures, we assessed the reconstruction accuracy of each method using the Root Mean Squared Error (RMSE) as the evaluation metric. The results indicate that Univariate SSA consistently outperforms both MSSA and FSSA in terms of fitting accuracy.

Furthermore, we examined the seasonal components captured by each approach. Our findings show that while multivariate and functional SSA are effective at identifying weekly seasonality (e.g., 7-day periodicity), they are unable to capture sub-daily patterns such as hourly or half-daily periodicities. In contrast, Univariate SSA successfully detects both daily and sub-daily seasonalities. However, identifying such components requires careful tuning of the window length, which may not be straightforward in practice.

These findings underscore the importance of selecting the appropriate SSA variant based on the temporal resolution of the data and the specific goals of the analysis. While MSSA and FSSA offer interpretability for coarser seasonal structures, Univariate SSA remains the most flexible and powerful tool for uncovering fine-grained periodicities in high-resolution time series.

Acknowledgment

This study was funded by the European Union - NextGenerationEU, Mission 4, Component 2, in the framework of the GRINS -Growing Resilient, INclusive and Sustainable project (GRINS PE00000018 – CUP F53C22000760007). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

References

- Amato, U., Antoniadis, A., De Feis, I., and Gijbels, I. (2025). Functional time series forecasting: a systematic review. *Statistical Papers*, **66**(1), 21.
- Fan, S., and Hyndman, R. J. (2011). Short-term load forecasting based on a semi-parametric additive model. *IEEE transactions on power systems*, **27**(1), 134-141.
- Golyandina, N., Nekrutkin, V., and Zhigljavsky, A. (2001). *Analysis of Time Series Structure: SSA and related techniques*, Chapman & Hall/CRC, New York - London.
- Haghibin, H., Najibi, S. M., Mahmoudvand, R., Trinka, J., & Maadooliat, M. (2021). Functional singular spectrum analysis. *Stat*, **10**(1), e330.
- Hassani, H., & Mahmoudvand, R. (2013). Multivariate singular spectrum analysis: A general view and new vector forecasting approach. *International Journal of Energy and Statistics*, **1**(01), 55-83.
- Hyndman, R. J., and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, **54**(8), 799-805.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Improved Generalized Additive Models in Time Series

Ahad Malekzadeh, Fatemeh Rajabi Naraki *

Department of Computer Science and Statistics, K.N. Toosi University of Technology, Tehran, Iran

Abstract

Generalized additive models (GAMs) have become a powerful and well-established tool in the field of statistics, which demonstrating remarkable performance for determining the model governing the changes of a random variable over time. By implementing GAM on time series data, it is possible to examine the behavior of nonlinear data and predict the future. One of the most important issues in these models is determining the appropriate link function because choosing it incorrectly will lead to unbiased results. In this paper, we intend to present an improved interpretation of the GAM that determines the appropriate link function for a time series dataset. Furthermore, the efficiency and performance of this method are illustrated with real-world examples.

Keywords: Generalized additive model, Link function, Improved generalized additive model, Smooth function, Splines, Penalized likelihood.

Mathematics Subject Classification (2020): 62M10, 91B84

1 Introduction

Data collection and analysis has become a principal foundation of decision-making and planning in various fields in the past decade. With the advancement of technology and increased access to data, organizations and individuals can better utilize this information. Data analysis is now used in most parts of industry and science, including engineering, economics and finance, marketing, medicine and health, etc. Ultimately, this approach has been influential not only in the business world but also in the daily lives of people. In general, data collection and analysis is known as a powerful tool in today's world that has a profound impact on all aspects of human life.

In the meanwhile, time series data performs a significant obligation in data analysis and refers to a set of data that is collected over time and arranged in chronological order. This data usually consists of repeated observations of a particular variable at regular intervals. Time series data allows analysts to identify long-term trends and seasonal or periodic patterns. Using various models such as Autoregressive Integrated Moving Average (ARIMA), Segmented Linear Regression Models, neural networks, etc. more accurate predictions can be achieved,

*Corresponding author: Ftm23rjb@gmail.com

nevertheless a more appropriate approach for this type of data is to use Generalized Additive Models (GAMs).

One of the first approaches that proposed to model nonlinear relationships was the additive model (AM), which was presented by [Friedman and Stuetzle \(1981\)](#), and then [Hastie and Tibshirani \(1986\)](#) invented its generalized form is called GAM and presented it to the world of science, which can fit smooth relationships between variables without knowing the non-linear relationship in advance. Consequently, several researchers, including [Wood \(2017\)](#), [Yee \(2015\)](#), [James et al. \(2013\)](#), etc., investigated and developed the infrastructure of GAMs. Initially, the additive model was defined for observations y_i and predictor variables x_1, \dots, x_p as follows,

$$AM : E(y_i|x_j) = \sum_{j=1}^p f_j(x_{ij}), \quad (1)$$

where the terms f_j denote non-parametric, smooth functions for the independent variable that are fitted to the data. As indicated, after defining the additive model, a generalized version of the model was obtained by applying a function to the expected value $E(y_i|x_j)$. The following model determines the structure of a nonparametric GAM in general,

$$GAM : S[E(y_i|x_j)] = S(\mu_i) = \sum_{j=1}^p f_j(x_{ij}), \quad (2)$$

where $\mu = E(Y)$ denotes the expected value, S represents a function that links the expected value to the predictor variables x_1, \dots, x_p , which is known in most contexts as the canonical link function with the symbol $g(\cdot)$ and note that an exponential family distribution for y_i (e.g. normal, binomial, or poisson distributions) is specified.

Mainly, GAMs endure in both parametric and nonparametric methods and are utilized in various data scenarios such as when the variable is independent, time series, or has a lag. GAMs were introduced in time series epidemiology studies by [Schwartz \(1993\)](#) and have since been used and developed by many as a standard approach to solving the control variable problem ([Schwartz \(1994b\)](#), [Bremner et al. \(1999\)](#), [Dominici et al. \(2000\)](#)). To illustrate the importance of using GAMs in time series data, we can cite studies that have been conducted in this field, for instance, the impact of air pollutants on human health has been investigated by using GAM by [Ravindraa \(2019\)](#), or to describe the physiology of heart and lung interaction, using GAM to decomposition time series and waveforms has been discussed by [Enevoldsen et al. \(2023\)](#), and after the COVID-19 pandemic spread around the world, many studies have been conducted on the diagnosis of disease incidence, mortality rates, and vaccination, such as: A extension of the GAM in spatial-temporal way for modeling COVID-19 mortality risk in Toronto was investigated by [Feng \(2022\)](#), and [Eckardt et al. \(2024\)](#) wrote their paper on the use of generalized functional additive mixed models with compositional variables for COVID-19 prevalence curves.

At the present time, we intend to address the issue of selecting the appropriate link function utilizing GAM on time series data. Since GAMs can work with different types of statistical distributions (e.g., normal, binary, Poisson, etc.), the link function is chosen depending on the type of distribution. For example, for binary data (e.g., success/failure), the logit function is usually used, or for count data, the Poisson (log) link function may be used. Using these functions, GAMs provide the ability to model complex, nonlinear relationships, helping analysts to derive more accurate results from the data. The correct choice of link function is of great importance depending on the type of data and the type of distribution of the dependent variable and can have a significant impact on the accuracy and interpretability of the model. In addition, if the link function is misspecified, it will provide biased and uncertain results. In this article, we intend to address an improved version of GAM (IGAMs) on time series data that considering the issue of selecting the appropriate link function.

The rest of the paper is as follows. In Section 2, we explain the theory of IGAMs when the data are time series, and discuss the select of the link function and parameter estimation.

In Section 3, we present an example from the COVID-19 study which we fit these models and measure the model selection criteria for them. Eventually, in Section 4, we conclude and discuss.

2 Model specification

The GAM presented by [Hastie and Tibshirani \(1986\)](#) is the classical addendum of general linear models (GLMs) (introduced by [Nelder and Wedderburn \(1972\)](#)) that replaces the linear predictor $\sum \beta_j x_{ij}$ with the additive predictor term $\sum f_j(x_{ij})$. In this section, we explain the use of GAMs in time series studies and attempt to provide explanations about introducing the model, estimating parameters, and choosing the appropriate link function.

2.1 Time Series by using GAM

GAM has been introduced as an effective and flexible approach for performing nonlinear regression analysis in time series studies. In time series studies, GAM has been the most widely used method because it allows nonparametric adjustments for nonlinear seasonal confounding effects, trends, and other covariate variables. These models have been broadly used in many analyses that are time-dependent, including climate change, health and medical data, product price changes, and inflation, etc. The GAM for a trend in a series of T observations y_t at observation times x_t ($t = 1, 2, \dots, T$) is defined as follows,

$$E(y_t|x_t) = \mu_t = S\left(\sum_{j=1}^p f_j(x_{t-j})\right), \quad (3)$$

where S is true unknown response function that estimated non-parametrically and f_j is smooth function. Solutions for estimating the true response function have been introduced, including the use of kernel functions by [Hardle et al. \(1993\)](#) and the use of P-spline approach by [Muggeo and Ferrara \(2008\)](#). To ensure the uniqueness and identifiability of the model, several constraints need to be imposed on the model. We assume that for a fixed value of $c > 0$ the following constraints are required:

$$\sum_{j=1}^p \int_{\min(x_j)}^{\max(x_j)} f_j(t)^2 dt = c, \quad (4)$$

furthermore, each predictor function is postulated to be (the usual conditions used to obtain identifiability in GAMs):

$$\int_{\min(x_j)}^{\max(x_j)} f_j(t) dt = 0, \quad (5)$$

in addition, we suppose that the response function is monotonically increasing, i.e.

$$\frac{\partial h_T(t)}{\partial t} \geq 0. \quad (6)$$

The main reason for imposing these restrictions is to achieve interpretable results. For non-monotonic response functions, the interpretation of the functions $f_j(\cdot)$ becomes vastly difficult, because an increasing (decreasing) function $f_j(\cdot)$ does not translate into an increasing (decreasing) effect on the response.

2.2 Link function and Estimation procedure

As previously mentioned, within the context of GLMs and GAMs, the inverse of the response function is generally called the link function ($g = S^{-1}(\cdot)$), which is approximated by a combination of functions of the form $S = h_0(h(\cdot))$, where $h_0(\cdot)$ is a constant response function (usually the canonical response function). In this manner, the inner unknown function $h(\cdot)$

must be estimated, which is characterized by expanding the basis into the $h(\eta_i) = \Phi^T(\eta_i)\alpha = \sum_{k=1}^{m_k} \phi_k(\eta_i)\alpha_k$, where η_i is the vector of the m_h B-spline basis functions evaluated at η_i and α is the basis coefficient vector corresponding to the inner function. We also use a basis expansion of B-spline written by $f_j(x) = \psi_j^T(x)\beta_j, j = 1, \dots, p$, to estimate the unknown functions $f_j(\cdot)$ where $\psi_j(x)$ is the vector of the m_j basis functions evaluated at x and β_j is the corresponding coefficient vector. Finally, by substituting the above expressions into equation 3, we obtain the following model estimation as IGAM,

$$IGAM : E(y_t|x_t) = \mu_t = h_0(\Phi^T(\sum_{j=1}^p \psi_{ij}^T(x)\beta_j)\alpha). \quad (7)$$

Various methods have been proposed for estimating GAM parameters, initially [Hastie and Tibshirani \(1986\)](#) used the local scoring algorithm which is Fisher score update using local score estimation and then, [Wood \(2004\)](#) showed that parameters can be estimated using a penalized likelihood approach. Here, for estimating IGAM parameters, based on what was obtained in [Tutz and Petry \(2016\)](#), present an algorithm based on boosting techniques, where each boosting iteration consists of two steps (1) updating the response function with a fixed prediction and (2) updating the prediction with a fixed response function. Both updates are based on penalized and constrained Fisher scoring. (See [Tutz and Petry \(2016\)](#) for more detailed)

3 Application to the data

Following the global pandemic of the coronavirus disease (COVID-19), some scientists and researchers have been able to gather a lot of evidence about the virus and learn more about how it is transmitted and how to treat it more effectively. With the number of hospitalizations for the coronavirus increasing daily, especially in 2021, it was necessary to study this data to detect and predict the peak of the disease in order to maintain global health around the world. It should be considered that vaccination has saved millions of lives during the COVID-19 pandemic. Within a year of the start of the pandemic, several vaccines against the virus were developed. This speed was possible due to previous research on coronavirus vaccines, increased attention and funding for research and development, and a high interest in volunteering for clinical trials. Here, we intend to examine the doses of Covid-19 vaccine administered in several countries such as China, Iran, France, etc., based on the data available on the Global Vaccination website [OurWorldinData](#). Here we are going to examine the doses of COVID-

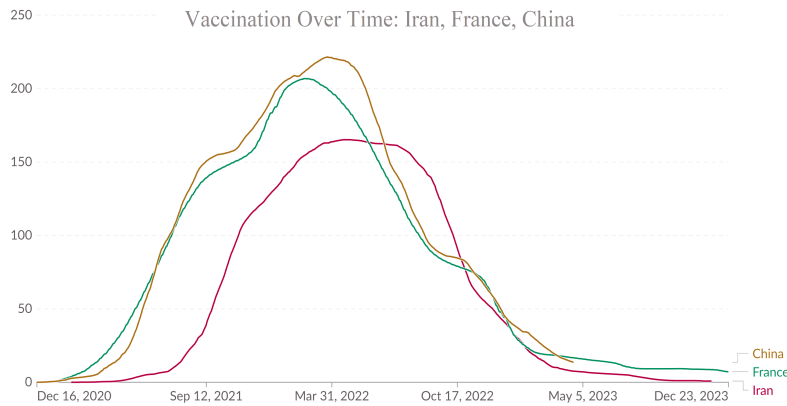


Figure 1: Per 100 population. The value shown for each date is the total number of vaccine doses administered in the 12 months prior to that date for the three countries China, Iran, and France.

19 vaccine administered in each days per 100 people in some countries such as China, Iran,

Italy, etc. in a two-year period from 2021 to 2023 based on the data available on the Global Vaccination Website. The data for these three regions are given in Figure 1. We have fitted the Additive model (AM), the generalized Additive model (GAM) and the improved generalized Additive model (IGAM) on the data and measured the Root Mean Squared Error (RMSE), the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) and the R-squared (R^2) criteria and we construct its values separately for each region in Table 1.

Table 1: Criteria for three processes AM, GAM and IGAM			
	AM	GAM	IGAM
China	RMSE = 74.507463	RMSE = 4.982501	RMSE = 4.480922
	AIC = 9815.482	AIC = 5200.566	AIC = -2500.448
	BIC = 9829.739	BIC = 5252.811	BIC = -2448.233
	$R^2 = 0.01122333$	$R^2 = 0.9955313$	$R^2 = 0.9983965$
Iran	RMSE = 63.250842	RMSE = 4.031698	RMSE = 2.771918
	AIC = 11371.87	AIC = 5766.393	AIC = -1831.283
	BIC = 11386.65	BIC = 5820.581	BIC = -1777.16
	$R^2 = 0.03341473$	$R^2 = 0.9960379$	$R^2 = 0.9967077$
France	RMSE = 64.178143	RMSE = 5.775090	RMSE = 5.155585
	AIC = 12182.87	AIC = 6944.363	AIC = -2134.83
	BIC = 12197.86	BIC = 6999.267	BIC = -2079.966
	$R^2 = 0.1346133$	$R^2 = 0.9929344$	$R^2 = 0.9950166$

Based on what is given in Table 1, for data of all regions, it is clear that the AM is not a good fit and not suitable at all and should not be used for such data. While both GAM and IGAM approaches perform relatively well, their R^2 value is close to 1, which proves this. However, the IGAM is a better fit and the lower values of the RMSE, AIC and BIC confirm it.

Discussion and Results

As has been shown in numerous papers over the past decades, GAM is a significant method for modeling nonlinear process data that can be predicted without the need for an analyst in the form of a nonlinear relationship. Here, we improve GAM (IGAM) to provide a more appropriate approach by properly selecting the link function, which is one of the main pillars of the model that needs to be estimated. We also analyzed a real-world application, data on the dose of the vaccine administered for COVID-19, and we saw the positive performance of IGAM in the example results given.

References

- Bremner S. A., Anderson H. R., Atkinson R. W., McMichael A. S., Strachan D. P., Bland J. M. and Bower J. S. (1999), Short term association between outdoor air pollution and mortality in London 1992–4, *Occupational and Environmental Medicine*, 56, 237–244.
- Dominici F., Samet J. and Zeger S. L. (2000), Combining evidence on air pollution and daily mortality from the largest 20 US cities: a hierarchical modeling strategy, *Journal of the Royal Statistical Society*, 163, 263–84.
- Eckardt M., Mateu J. and Greven S. (2024), Generalized functional additive mixed models with (functional) compositional covariates for areal Covid-19 incidence curves, *Journal of the Royal Statistical Society Series C, Applied Statistics*, 2024, 73, 880–901.

- Enevoldsen J., Simpson G.L. and Vistisen S.T. (2023), Using generalized additive models to decompose time series and waveforms, and dissect heart–lung interaction physiology, *Journal of Clinical Monitoring and Computing*, 37:165–177
- Feng C. (2022), Spatial-temporal generalized additive model for modeling COVID-19 mortality risk in Toronto, Canada, *ScienceDirect*, Spatial Statistics 49 (2022) 100526.
- Friedman J.H. and Stuetzle W. (1981), "Projection Pursuit Regression", *Journal of the American Statistical Association* 76:817–823.
- Hardle W., Hall P. and Ichimura H., (1993), *Optimal smoothing in single-index models*, Ann. Statist. 21, pp. 157–178.
- Hastie T.J. and Tibshirani R.J. (1986), Generalized Additive Models, *Statistical Science*, 3(1), 297–318.
- Hastie T.J. and Tibshirani R.J. (1990), *Generalized Additive Models*, Chapman and Hall, London.
- James G., Witten D., Hastie T. and Tibshirani R. (2013), *An Introduction to Statistical Learning*, springer, New York, 112.
- Muggeo V.M.R. and Ferrara G., (2008), Fitting generalized linear models with unspecified link function: A p-spline approach, *Comput. Statist. Data Anal.* 52 , pp. 2529–2537.
- Nelder J. and Wedderburn R. (1972), Generalized Linear Models, *Journal of the Royal Statistical Society, Series A (General)*, Blackwell Publishing, 135(3), 370–384.
- Ravindraa K., Rattana P., Morb S. and Aggarwal A.N. (2019), Generalized additive models: Building evidence of air pollution, climate change and human health, *ScienceDirect Environment International* 132 104987.
- Schwartz J. (1993), Air pollution and mortality in Birmingham, AL, *American Journal of Epidemiology*, 137, 1136–1147.
- Schwartz J. (1994), Nonparametric smoothing in the analysis of air pollution and respiratory illness, *The Canadian Journal of Statistic*, 22, 471–487.
- Tutz G. and Petry S. (2016), Generalized additive models with unknown link function including variable selection, *Journal of Applied Statistics*
- Wood, S. N. (2004), *Generalized Additive Models: an introduction with R*, first ed, Chapman and Hall/CRC, New York.
- Wood, S. N. (2017), *Generalized Additive Models: an introduction with R*, 2nd ed, CRC Press.
- Yee T.W. (2015), *Vector Generalized Linear and Additive Models with an Implementation in R*, Springer.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Stochastic Modeling with Normal Mean–Variance Mixtures: A Robust Linear Experts Framework for Censored Data in Machine Learning

Tahere Manouchehri^{*1}, Alireza Nematollahi¹

¹Department of Statistics, Shiraz University, Shiraz, Iran.

Abstract

In this paper, we propose MoE-NMVM-CR, a robust extension of the MoE framework that leverages the Normal Mean–Variance Mixture (NMVM) family of distributions for modeling of censored regression (CR) data. Unlike traditional Gaussian-based MoE models, our approach captures complex data characteristics such as skewness, multimodality, and outliers, making it particularly suitable for real-world scenarios where standard assumptions fail. Experiments on synthetic and real-world datasets, show that MoE-NMVM-CR outperforms traditional MoE, particularly in high-noise and censored settings. This work bridges advanced stochastic modeling with practical machine learning needs, offering a scalable, robust, and interpretable framework for modern data challenges.

Keywords: Mixture of linear experts model; Normal mean–variance mixture; Censored data analysis; ECME algorithm; Model-based clustering.

Mathematics Subject Classification (2020): 91C20, 62Nxx.

1 Introduction

Modern datasets are increasingly characterized by complexity and heterogeneity that challenge the assumptions of classical statistical and machine learning models. For instance, in medical studies, patients with ostensibly similar clinical profiles may respond differently to the same treatment due to unobserved genetic or environmental factors. In finance, regime shifts can lead to abrupt changes in asset returns. In environmental monitoring, pollutant concentrations are often censored at assay detection limits. One widely used approach to model such latent heterogeneity is the mixture of regression models (MRM).

Let $\mathbf{x}_i = (1, x_{1i}, \dots, x_{(p-1)i})^\top$ represent a p -dimensional vector of fixed explanatory variables corresponding to a scalar response variable Y_i , for $i = 1, \dots, n$. Additionally, we define latent membership indicators $\mathbf{Z}_i = (Z_{1i}, \dots, Z_{Gi})^\top$, where $Z_{ji} = 1$ indicates that Y_i belongs to the j th group ($j = 1, \dots, G$), and $Z_{ji} = 0$ otherwise. The mixing probabilities $\Pr(Z_{ji} = 1) = \pi_j$ are constrained to satisfy $\sum_{j=1}^G \pi_j = 1$, where G denotes the number of

^{*}Corresponding author, t.manouchehri@shirazu.ac.ir

mixture components. Consequently, the random vector \mathbf{Z}_i follows a multinomial distribution with one trial and cell probabilities (π_1, \dots, π_G) , denoted by $\mathbf{Z}_i \sim M(1; \pi_1, \dots, \pi_G)$. The classical MRM is expressed as:

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta}_j + \epsilon_{ji}, \quad \text{with mixing probability } \pi_j, \quad (1)$$

where $\boldsymbol{\beta}_j = (\beta_{0j}, \beta_{1j}, \dots, \beta_{(p-1)j})^\top$ is a $p \times 1$ vector of regression coefficients for the j th component, including the intercept β_{0j} . The error terms $\epsilon_{ji} \sim N(0, \sigma_j^2)$ are assumed to be independent and identically distributed. The parameter set $\Theta = \{\boldsymbol{\pi}, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_G\}$ includes the mixing proportions $\boldsymbol{\pi} = (\pi_1, \dots, \pi_{G-1})$ and component-specific parameters $\boldsymbol{\theta}_j = \{\boldsymbol{\beta}_j, \sigma_j^2\}$.

Let $\mathbf{r}_i = (1, r_{i1}, \dots, r_{i(q-1)})^\top$ denote a q -dimensional vector of covariates corresponding to the scalar response variable Y_i . In contrast to the constant mixing proportions in MRM, the MoE model employs a dynamic structure wherein the mixing proportion π_j is specified as a multinomial logistic function of the input covariates \mathbf{r}_i . This logistic function, commonly referred to as the gating function, regulates the allocation of observations to the component distributions. Extending the MRM framework, the probability density function (pdf) of Y in a G -component normal-based mixture of experts (MoE-N) model can be expressed as follows:

$$f(y_i; \Theta) = \sum_{j=1}^G \pi_j(\mathbf{r}_i; \boldsymbol{\tau}_j) \phi(y_i; \mathbf{x}_i^\top \boldsymbol{\beta}_j, \sigma_j^2), \quad (2)$$

where for the gating parameters, $\boldsymbol{\tau} = (\boldsymbol{\tau}_1^\top, \dots, \boldsymbol{\tau}_{G-1}^\top)^\top$ with $\boldsymbol{\tau}_j = (\tau_{j0}, \dots, \tau_{j(q-1)})^\top$,

$$\pi_j(\mathbf{r}_i; \boldsymbol{\tau}_j) = \Pr(Z_{ji} = 1 \mid \mathbf{r}_i) = \frac{\exp(\boldsymbol{\tau}_j^\top \mathbf{r}_i)}{1 + \sum_{l=1}^{G-1} \exp(\boldsymbol{\tau}_l^\top \mathbf{r}_i)}, \quad (3)$$

and the full parameter set $\Theta = \{\boldsymbol{\tau}, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_G\}$. Recently, the classical MoE has found appealing applications in many fields, such as business, marketing, and biological studies. It has also been extended to accommodate heavy-tail and/or skew distributed data see [Chamroukhi \(2016\)](#) and [Nguyen and McLachlan \(2016\)](#).

The structure of this paper is outlined as follows: In Section 2, we provide a concise overview of the NMVM distribution, emphasizing its key features. Section 3 introduces the proposed MoE-NMVM-CR model and presents the details of the maximum likelihood estimation (ML) process for this model, implemented through an EM-type algorithm. Section 4 is dedicated to comprehensive simulation studies that assess the efficiency and performance of the proposed approach. In Section 5, we demonstrate the applicability of our methodology by applying it to real-world data examples.

2 Normal mean–variance mixture distributions

The NMVM distribution is a flexible family of distributions formed by scaling the mean and variance of a normal random variable with a positive mixing random variable. Specifically, a random variable Y follows the NMVM distribution, denoted by $Y \sim \text{NMVM}(\mu, \lambda, \sigma^2, \boldsymbol{\nu})$, if it can be expressed as:

$$Y \mid (W = w) \sim N(\mu + w\lambda, w\sigma^2), \quad (4)$$

where $\mu \in \mathbb{R}$, $\lambda \in \mathbb{R}$, and W is a positive mixing random variable with a probability density function (PDF) $h(\cdot; \boldsymbol{\nu})$, indexed by the parameter vector $\boldsymbol{\nu}$. The NMVM distribution can also be represented stochastically as:

$$Y = \mu + \lambda W + \sqrt{W} Z, \quad (5)$$

where Z is a standard normal random variable $Z \sim N(0, \sigma^2)$ and is independent of W . For additional details, see [McNeil et al. \(2015\)](#). The PDF of the NMVM distribution is expressed as:

$$f_{\text{NMVM}}(y; \mu, \lambda, \sigma^2, \boldsymbol{\nu}) = \int_0^\infty \phi(y; \mu + w\lambda, w\sigma^2) h(w; \boldsymbol{\nu}) dw, \quad (6)$$

where $\phi(y; \mu, \sigma^2)$ denotes the normal PDF with mean μ and variance σ^2 .

The NMVM distribution includes several significant special cases, each offering unique advantages for modeling complex data. One prominent example is the generalized hyperbolic (GH) distribution, derived when the mixing random variable W follows a generalized inverse Gaussian (GIG) distribution. Renowned for its ability to model data with heavy tails and skewness, the GH distribution finds applications in areas like finance and risk management. Additionally, other significant members of the NMVM family, include the generalized hyperbolic skew-t (GHST), normal inverse Gaussian (NIG), variance-gamma (VG), skew Laplace (SL), and normal mean-variance mixtures of the Birnbaum-Saunders (NMVBS) and Lindley (NMVL) distributions. Together, these variants enhance the NMVM framework's flexibility, making it suitable for a diverse range of applications.

3 MoE-NMVM-CR model

In the context of model (2), where we assume $\epsilon_{ij} \sim \text{NMVM}(0, \lambda_j, \sigma_j^2, \nu_j)$, the MoE-NMVM-CR model is defined as follows. Let $\mathbf{x} \in \mathbb{R}^p$ represent the vector of explanatory variables, and let $\mathbf{r}_i = (1, r_{i1}, \dots, r_{i(q-1)})^\top \in \mathbb{R}^q$ denote the covariates corresponding to Y_i . The gating parameters are specified by $\boldsymbol{\tau} = (\boldsymbol{\tau}_1^\top, \dots, \boldsymbol{\tau}_{G-1}^\top)^\top$, where each $\boldsymbol{\tau}_j = (\tau_{j0}, \dots, \tau_{j(q-1)})^\top$, and the mixing proportions are defined in (3). The PDF of the response random vector $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$, where the expert components follow the NMVM distribution, is given by:

$$f(y_i; \boldsymbol{\Theta}) = \sum_{j=1}^G \pi_j(\mathbf{r}_i, \boldsymbol{\tau}_j) f_{\text{NMVM-CR}}(y_i; \mathbf{x}_i^\top \boldsymbol{\beta}_j, \lambda_j, \sigma_j^2, \nu_j), \quad i = 1, \dots, n, \quad (7)$$

where $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_G, \boldsymbol{\tau}\}$ represents the model parameters, and $\boldsymbol{\theta}_j = (\boldsymbol{\beta}_j, \lambda_j, \sigma_j^2, \nu_j)$ in which $\nu_j = (\kappa_j, \psi_j)^\top$ for the VG; for the GHST, NIG, NMVBS, and NMVL, $\nu_j = \nu_j$; and for the SL, ν_j is null.

In the MoE-NMVM-CR model, we assume that the response vector \mathbf{Y} is only partially observed. Specifically, some of the response variables are subject to censoring, which may occur in the form of interval-, left-, or right-censoring. Let the available response variable Y_i be represented as the joint variables (W_i, ρ_i) , where W_i represents either the uncensored reading ($W_i = Y_{Oi}$) or interval-censoring ($W_i = (d_{i1}, d_{i2})$, for some fixed threshold points d_{i1}, d_{i2}). The censoring indicator ρ_i is defined as follows: $\rho_i = 1$ if $d_{i1} \leq Y_i \leq d_{i2}$, and $\rho_i = 0$ if $Y_i = Y_{Oi}$. In this setting, left-censoring occurs if $d_{i1} = -\infty$, and right-censoring occurs if $d_{i2} = +\infty$. For cases where $-\infty \neq d_{i1} < d_{i2} \neq +\infty$, the observed realization is interval-censored. We establish our methodology primarily on the interval-censoring scheme; however, the left- and right-censoring schemes are also examined in the simulation and real-data analyses. The log-likelihood of parameters $\boldsymbol{\theta}$ for the observed data $\mathbf{y} = (y_1, \dots, y_n)$ can be written as:

$$\ell(\boldsymbol{\Theta}|\mathbf{y}) = \sum_{i=1}^n \log \left(\sum_{j=1}^G \pi_j(\mathbf{r}_i, \boldsymbol{\tau}_j) f_{\text{NMVM-CR}}(y_i; \mathbf{x}_i^\top \boldsymbol{\beta}_j, \lambda_j, \sigma_j^2, \nu_j) \right), \quad (8)$$

where $f_{\text{NMVM-CR}}(\cdot)$ represents the PDF of the NMVM distribution, as defined in (6) within the censoring framework. The log-likelihood can then be further refined as follows:

$$\begin{aligned} \ell(\boldsymbol{\Theta}|\mathbf{y}) = \sum_{i=1}^n \log \sum_{j=1}^G \pi_j(\mathbf{r}_i, \boldsymbol{\tau}_j) & \left[\sigma_j^{-1} f_{\text{NMVM}} \left(\frac{y_{Oi} - \mathbf{x}_i^\top \boldsymbol{\beta}_j}{\sigma_j}; \lambda_j, \nu_j \right) \right]^{1-\rho_i} \\ & \times \left(F_{\text{NMVM}} \left(\frac{d_{i2} - \mathbf{x}_i^\top \boldsymbol{\beta}_j}{\sigma_j}; \lambda_j, \nu_j \right) - F_{\text{NMVM}} \left(\frac{d_{i1} - \mathbf{x}_i^\top \boldsymbol{\beta}_j}{\sigma_j}; \lambda_j, \nu_j \right) \right)^{\rho_i} \Big]. \end{aligned} \quad (9)$$

Building upon (7), we define the component label vector $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iG})^\top$, where each element Z_{ij} is a binary latent variable that indicates the membership of the i -th observation

in the j -th component. Under this framework, the conditional distribution of the response variable Y_i , given $Z_{ij} = 1$, is described as:

$$Y_i | Z_{ij} = 1 \sim \text{NMVM}(\mathbf{x}_i^\top \boldsymbol{\beta}_j, \lambda_j, \sigma_j^2, \boldsymbol{\nu}_j), \quad i = 1, \dots, n, \quad j = 1, \dots, G. \quad (10)$$

By utilizing the stochastic representation of the NMVM, the hierarchical representation of the MoE-NMVM-CR model is expressed as follows:

$$\begin{aligned} Y_i | (\mathbf{x}_i, U_i = u_i, Z_{ij}) &\sim \mathcal{N}(\mathbf{x}_i^\top \boldsymbol{\beta}_j + u_i \lambda_j, u_i \sigma_j^2), \\ U_i | Z_{ij} = 1 &\sim \mathcal{H}(u_i; \boldsymbol{\nu}_j), \\ Z_i | \mathbf{r}_i &\sim \mathcal{M}(1; \pi_1(\mathbf{r}_i, \boldsymbol{\tau}_1), \dots, \pi_G(\mathbf{r}_i, \boldsymbol{\tau}_G)), \end{aligned} \quad (11)$$

where $\mathcal{M}(1; \cdot)$ denotes the one-trial multinomial distribution.

To compute the ML estimates of the parameters in the MoE-NMVM-CR model, we adopt the expectation conditional maximization either (ECME) algorithm proposed by (Liu and Rubin, 1994). This algorithm is an extension of the classical expectation conditional maximization (ECM) method (Meng and Rubin, 1993), retaining its attractive properties, such as monotonic convergence and ease of implementation. However, the ECME algorithm enhances computational efficiency by replacing certain conditional maximization (CM) steps with conditional maximum likelihood (CML) steps, which directly maximize the constrained log-likelihood function.

To derive the ML function based on the complete data, we begin by calculating the log-likelihood function for the complete dataset. Let $\mathbf{y} = (y_1, \dots, y_n)^\top$ represent the observed data vector, $\boldsymbol{\rho} = (\rho_1, \dots, \rho_n)^\top$ the vector of censoring indicators, and $\mathbf{u} = (u_1, \dots, u_n)^\top$ and $\mathbf{Z} = (\mathbf{Z}_1^\top, \dots, \mathbf{Z}_n^\top)^\top$ the hidden values. The log-likelihood function for $\boldsymbol{\Theta}$, associated with the complete data $\mathbf{y}_c = (\mathbf{w}^\top, \boldsymbol{\rho}^\top, \mathbf{y}^\top, \mathbf{u}^\top, \mathbf{Z}^\top)^\top$, is given by:

$$\log L_c(\boldsymbol{\Theta}) = \log L_{1c}(\boldsymbol{\tau}) + \sum_{j=1}^G [\log L_{2c}(\boldsymbol{\theta}_j) + \log L_{3c}(\boldsymbol{\nu}_j)],$$

where $\boldsymbol{\theta}_j = (\boldsymbol{\beta}_j^\top, \sigma_j^2, \lambda_j, \boldsymbol{\nu}_j)$. The components of the log-likelihood are:

$$\begin{aligned} \log L_{1c}(\boldsymbol{\tau}) &= \sum_{i=1}^n \sum_{j=1}^G Z_{ij} \log \pi_j(\mathbf{r}_i, \boldsymbol{\tau}), \\ \log L_{2c}(\boldsymbol{\theta}_j) &\approx \sum_{i=1}^n Z_{ij} \left[-\frac{1}{2} \log \sigma_j^2 - \frac{1}{2u_i \sigma_j^2} \left(y_i - (\mathbf{x}_i^\top \boldsymbol{\beta}_j + u_i \lambda_j) \right)^2 \right], \\ \log L_{3c}(\boldsymbol{\nu}_j) &= \sum_{i=1}^n Z_{ij} \log h(u_i; \boldsymbol{\nu}_j), \end{aligned}$$

where $h(u_i; \boldsymbol{\nu}_j)$ is the probability density function of $U_i | Z_{ij} = 1$.

4 Numerical study

This section presents two Monte Carlo simulation studies to examine the asymptotic properties of the ML estimators, and evaluate the clustering.

4.1 Finite sample properties

In this section, we use Monte Carlo simulations to evaluate the performance of the ML estimates obtained via the ECME algorithm for the parameters of the MoE-NMVM-CR model. The study investigates how estimation accuracy and variability are influenced by varying sample sizes ($n = 100, 500, \text{ and } 1000$) under different levels of left-censoring. We generated 500 Monte Carlo (MC) samples from the special cases of the MoE-NMVM-CR model, assuming $G = 2$, $p = 4$, and $q = 2$. The true model parameters were set as follows: $(\lambda_1, \lambda_2) = (2, 3)$,

$(\nu_1, \nu_2) = (3, 5)$, $(\sigma_1^2, \sigma_2^2) = (0.5, 2)$, $\tau_1 = (0.7, 1, 2)^\top$, $\beta_1 = (0, 1.5, -2, -3)^\top$, and $\beta_2 = (-1, 1, 1.5, 3)^\top$. Each observation was generated with a covariate vector $\mathbf{x}_i = (1, x_{1i}, x_{2i}, x_{3i})^\top$, where $x_{1i} \sim U(1, 5)$, $x_{2i} \sim U(-2, 2)$, and $x_{3i} \sim U(1, 4)$. The corresponding gating covariates were $\mathbf{r}_i = (1, r_{1i}, r_{2i})^\top$, with $r_{1i} \sim U(-2, 1)$ and $r_{2i} \sim U(-1, 1)$. Left-censoring was then applied at levels of 7.5%, and 15%, followed by parameter estimation using the ECME algorithm.

Figures 1 and 2 display boxplots of the key parameter estimates across all sub-models. For a fixed censoring level, both bias and variability tend to decrease as the sample size increases, consistent with the asymptotic properties of maximum likelihood estimation. Conversely, for a fixed sample size, higher censoring levels lead to increased bias and variability. These patterns confirm the robustness and reliability of the ECME-based approach under diverse data conditions.

4.2 Evaluation of classification performance

This simulation study evaluates the classification and computational performance of the proposed MoE-NMVM-CR sub-models under increasing levels of interval censoring. In particular, each sub-model was evaluated using data generated from its corresponding distributional form. This approach assesses each model's self-consistency and capacity to recover true component allocations under its own generative assumptions. For each model and censoring level (7.5% and 15%), the simulation was replicated 200 times on samples of size 500. The model parameters are set as follows: $\beta_1 = (0, -1, -2, -3)^\top$, $\beta_2 = (-1, 1, 2, 3)^\top$, $\beta_3 = (0, -2, 1, 3)^\top$, $\tau_1 = (0.7, 1, 6)^\top$, $\tau_2 = (1, 0.9, 10)^\top$, $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (1, 2, 4)$, $\theta_1 = (2, 3)$, $\theta_2 = (3, 2)$, $\theta_3 = (2, 2)$, $(\nu_1, \nu_2, \nu_3) = (3, 5, 7)$, and $(\lambda_1, \lambda_2, \lambda_3) = (2, 3, 5)$ (only for the VG distribution). The covariates are defined as $\mathbf{x}_i = (1, x_{i1}, x_{i2}, x_{i3})^\top$ and $\mathbf{r}_i = (1, r_{i1}, r_{i2})^\top$, where $x_{i1}, x_{i2}, x_{i3} \sim U(1, 5), U(-2, 2), U(1, 4)$, respectively, and $r_{i1} \sim U(-2, 1)$, $r_{i2} \sim U(-1, 1)$.

Each fitted model is evaluated based on the mean right allocation (MRA), standard deviation of allocation rate (SDAR), misclassification rate (MCR), and average CPU time (in minutes). These performance metrics collectively capture both the accuracy and computational efficiency of the sub-models. The results are presented in Table 1.

The results highlight the strong classification performance and computational efficiency of the proposed sub-models within the MoE-NMVM-CR framework. As expected, all models exhibit a slight reduction in performance with increasing censoring levels, reflecting the growing difficulty of accurate allocation under greater information loss. However, the sub-models remain remarkably robust, with only moderate changes in MRA and MCR across censoring levels.

Table 1: Average performance metrics (MRA, SDAR, MCR, and CPU T.) computed over 200 replications for each MoE-NMVM-CR sub-model.

Model	MRA		SDAR		MCR		CPU T.	
	7.5%	15%	7.5%	15%	7.5%	15%	7.5%	15%
GHST	460.136	445.687	9.168	13.752	0.083	0.117	6.892	4.679
NIG	450.543	430.239	11.460	16.044	0.106	0.142	1.436	1.392
NMVB	435.2769	415.556	13.752	18.336	0.138	0.170	0.267	0.528
NMVL	400.0455	380.452	16.044	20.628	0.235	0.245	0.215	0.151
SL	425.8015	405.043	18.336	25.212	0.150	0.199	9.488	12.704
VG	390.5825	365.860	19.628	26.504	0.220	0.276	8.979	8.841

5 Real data analysis

This section presents an in-depth analysis of the wage rates dataset, previously analyzed by Mroz (1984) and more recently by Park et al. (2024), for illustrative purposes of the developed novel MoE-NMVM-CR model. This dataset includes 753 observations of annual wage rates

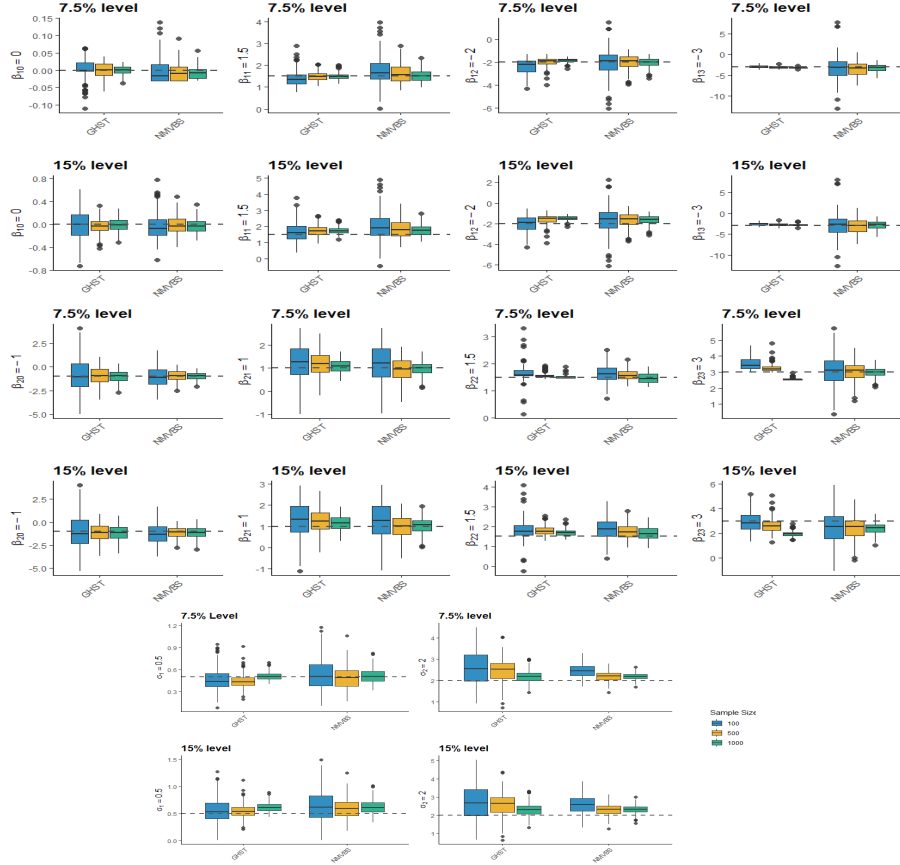


Figure 1: Boxplots of β_1 , β_2 and σ^2 estimates for the MoE-NMVM-CR model under censoring levels of 7.5%, and 15%. The horizontal line indicates the true value of the parameter.

(in hours worked outside the home) for married white women aged 30–60 in 1975. Notably, 325 individuals (43.16%) report zero hours, which are treated as left-censored at zero.

Following prior studies, the response variable y is defined as annual hours worked outside the home divided by 1000. The explanatory variables include x_1 : years of education, x_2 : age, x_3 : previous labor market experience, and $x_4 = x_3^2$: squared experience. The data exhibit a heavily right-skewed and bimodal distribution, as shown in Figure 3, supporting the use of flexible, component-based modeling. Moreover, the Hopkins statistic for the full dataset is $H = 0.1865$, indicating a strong clustering tendency and a significant deviation from spatial randomness. (Banerjee and Davé, 2004).

For the gating network, we specify $\mathbf{r} = (1, r_1, r_2, x_2)^\top$, where r_1 is the local unemployment rate, r_2 is non-wife household income, and x_2 (age) is included again due to its relevance to both group membership and regression structure.

We applied the proposed MoE-NMVM-CR models—including the GHST, NIG, NMVBS, NMVL, SL, and VG sub-models—to the wage rates dataset, considering $G = 1, \dots, 4$ components. In addition, we included the standard MoE-N model based on the normal distribution as a baseline for comparison. Based on all considered model selection criteria (AIC, BIC, EDC, AICC, and ABIC), the two-component structure was consistently preferred across sub-models, with the MoE-NMVL-CR model achieving the best overall fit. Although these results confirm the suitability of a two-cluster solution, they also highlight the benefits of modeling heavy tails and skewness when fitting censored, heterogeneous data.

All models were trained on 70% of the data, while the remaining 30% was held out as a test set for model evaluation. The results reported in the table 2 are based on the training data;

however, the test set was used to assess predictive performance. Notably, the closeness of evaluation metrics—such as RMPE, MCR, RI, and JCI—between the train and test sets indicates that the proposed models generalize well to unseen data, confirming their effectiveness for prediction. Across all MoE-NMVM-CR variants, the estimated coefficient β_{11} indicates that education positively influences work hours in one component and negatively in the other—suggesting heterogeneous effects that would be masked in a single regression. Coefficients associated with labor market experience (x_3 and x_4) suggest increasing returns at lower experience levels, with diminishing marginal effects thereafter, consistent with concave functional behavior. Age (x_2) is negatively associated with annual work hours across both components, indicating a general decline in labor force participation with age. Figure 4 displays a bivariate scatter of non-wife income versus work hours with points colored by MoE cluster on the left and by city status on the right. The clear diagonal division in the MoE-colored panel contrasts with the extensive overlap in the city-colored panel, illustrating that the proposed skewed, heavy-tailed, censored MoE-NMVM-CR model reveals latent labor-market segments that the binary urban/rural indicator cannot resolve.

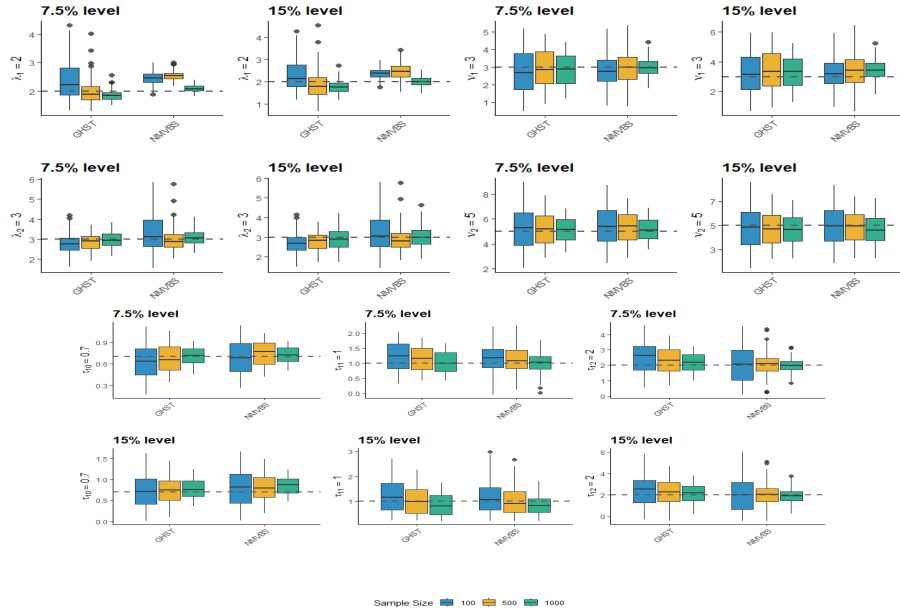


Figure 2: Boxplots of ν , λ and τ estimates for the MoE-NMVM-CR model under censoring levels of 7.5%, and 15%. The horizontal line indicates the true value of the parameter.

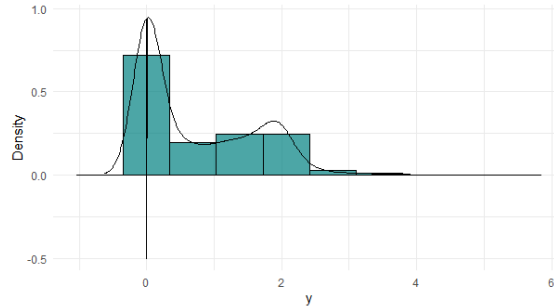


Figure 3: Empirical distribution of y with a kernel density overlay.

Table 2: The AIC, BIC, prediction and clustering performance measures.

	Measure	GHST	NIG	SL	NMVBS	NMVL	VG	N
Train	AIC	31.0155	29.2899	31.3088	31.4468	26.2365	34.9414	52.44673
	BIC	116.3595	114.6339	116.6528	116.7908	103.0461	128.8198	120.72194
	EDC	82.8414	81.1158	83.1347	83.2727	72.8799	91.9499	93.90747
	AICC	32.6756	30.9500	32.9689	33.1068	27.5830	36.9494	53.51340
	ABIC	52.8742	51.1486	53.1675	53.3055	45.9093	58.9860	69.93368
	RMPE	1.2911	0.8389	1.3152	1.7953	0.9345	0.9719	2.3358
	MCR	0.2751	0.2353	0.3397	0.2998	0.3245	0.3017	0.3528
	RI	0.5424	0.5435	0.5380	0.5413	0.5402	0.5380	0.5097
	JCI	0.5903	0.6013	0.5672	0.5826	0.5747	0.5768	0.4241
Test	RMPE	1.3011	0.8481	1.3408	1.9153	0.9805	0.9453	2.5464
	MCR	0.3230	0.2655	0.3540	0.3142	0.3496	0.3274	0.3451
	RI	0.5244	0.5265	0.5224	0.5244	0.5204	0.5224	0.5224
	JCI	0.2130	0.2244	0.2061	0.2146	0.2057	0.2110	0.2077

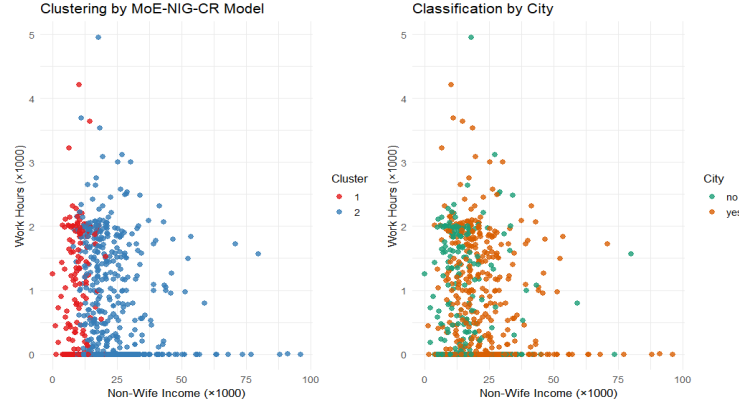


Figure 4: Scatterplots of non-wife income vs. work hours (in thousands). *Left*: points colored by MoE cluster, showing two well-separated groups. *Right*: points colored by the city indicator (metro vs. non-metro), revealing substantial overlap and illustrating that geography alone does not recover the latent clusters.

References

- Banerjee, A. and Davé, R. N. (2004). Validating clusters using the hopkins statistic. In *2004 IEEE International Conference on Fuzzy Systems - Proceedings*, volume 1, pages 149–153, Budapest, Hungary.
- Chamroukhi, F. (2016). Robust mixture of experts modeling using the t distribution. *Neural Networks*, 79:20–36.
- Liu, C. and Rubin, D. B. (1994). The ecme algorithm: a simple extension of em and ecm with faster monotone convergence. *Biometrika*, 81(4):633–648.
- McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press.
- Meng, X.-L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika*, 80(2):267–278.
- Mroz, T. A. (1984). *The sensitivity of an empirical model of married women’s hours of work to economic and statistical assumptions*. Stanford University.
- Nguyen, H. D. and McLachlan, G. J. (2016). Laplace mixture of linear experts. *Computational Statistics & Data Analysis*, 93:177–191.
- Park, J., Dey, D. K., and Lachos, V. H. (2024). Finite mixture of regression models for censored data based on the skew-t distribution. *Computational Statistics*, pages 1–32.



A Comparison of Maximum Likelihood Estimation and Particle Swarm Optimization for Parameter Estimation in Logistic and Probit Regression Models

Seyed Mohsen Mohammadi¹, Reza Etesami², Mohsen Madadi^{3*}

¹M.Sc. Student, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran.

mohsenmohammadi@math.uk.ac.ir

²Ph.D. Student, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran.

rezaetesamii@math.uk.ac.ir

³Associate Professor, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran.

madadi@uk.ac.ir

Abstract

This study compares Maximum Likelihood Estimation (MLE) and Particle Swarm Optimization (PSO) for parameter estimation in logistic and probit regression models, using the Iris and Titanic datasets. After preprocessing and model evaluation, PSO demonstrated superior performance over MLE in both models. The findings demonstrate that metaheuristic algorithms like PSO can enhance the accuracy of classification models, offering valuable insights for researchers choosing optimization methods in statistical modeling.

Keywords: Maximum Likelihood Estimation (MLE), Particle Swarm Optimization (PSO), Logistic Regression, Probit Regression, Parameter Estimation.

Mathematics Subject Classification (2020): 62J12, 62F10, 68T20, 90C59.

1 Introduction

Parameter estimation plays a fundamental role in the development and application of regression models used for classification tasks (Bunke et al, 1977). Traditional statistical methods, such as Maximum Likelihood Estimation (MLE), have long been the standard approach for estimating parameters in logistic and probit regression models due to their theoretical soundness and widespread familiarity (Vasisht, 2007). However, these classical techniques often face limitations when dealing with complex, high-dimensional, or nonlinear problems, which can lead to issues like convergence to local optima or computational inefficiency (Aziz et al., 2016).

*Corresponding author: madadi@uk.ac.ir

In recent years, the advent of artificial intelligence and optimization algorithms has provided alternative methods capable of overcoming some of these challenges (Zga et al., 2025). Among these, metaheuristic algorithms, notably Particle Swarm Optimization (PSO), have gained popularity owing to their ability to efficiently explore complex search spaces and identify optimal or near-optimal solutions. PSO, inspired by the social behavior of flocks of birds or schools of fish, has been successfully applied to various optimization problems in engineering, science, and data science (Sadik et al., 2025).

This study addresses the lack of comparative evaluations between traditional and modern optimization methods by systematically comparing Maximum Likelihood Estimation (MLE) and Particle Swarm Optimization (PSO) for parameter estimation in logistic and probit regression models. Using the Iris and Titanic datasets to represent univariate and multivariate scenarios, the study assesses model performance through mean squared error (MSE) and mean absolute error (MAE). The findings highlight PSO’s potential to outperform MLE, underscoring the value of intelligent optimization techniques in statistical modeling.

2 Literature Review

Binary regression is a crucial branch of regression modeling that is applied when the dependent variable assumes only two possible outcomes, such as "yes/no", "success/failure", "buy/not buy", or "diseased/healthy". The central objective in binary regression is to estimate the probability of the occurrence of one of the two possible outcomes, conditional on a set of explanatory variables (Harris, 2021). These independent variables can be either quantitative or qualitative. Based on the values of these variables, binary regression models compute the likelihood of a particular event occurring. Such models are widely used across various scientific disciplines including medicine, management, social sciences, and marketing (Wilson et al., 2024).

Two of the most prominent binary regression models are *logistic regression* and the *probit model*. Logistic regression transforms a linear combination of input variables into probabilities using the logistic (sigmoid) function, providing interpretability and computational convenience. This model is extensively employed in data mining and machine learning tasks (Williams and Jorgensen, 2023).

A central challenge in fitting binary regression models such as logistic or probit regression lies in accurately and efficiently estimating model parameters. Despite using different distributional assumptions (logistic and normal), both models share similar optimization frameworks and estimation methodologies. In this study, we investigate two widely-used parameter estimation methods: Maximum Likelihood Estimation (MLE) and the Particle Swarm Optimization (PSO) algorithm. These methods are compared and evaluated in a tabular format.

2.1 Maximum Likelihood Estimation (MLE)

Maximum Likelihood Estimation is one of the most widely-used techniques for estimating model parameters in statistical modeling, including logistic and probit regression models (Kasap and Faouri, 2025). The method seeks to maximize the likelihood function, denoted as $L(\beta)$. Given a dataset $\{(y_i, \mathbf{x}_i)\}_{i=1}^n$, where y_i is the binary response variable and \mathbf{x}_i is the vector of predictors, the likelihood function is defined as:

$$L(\beta) = \prod_{i=1}^n f(y_i | \mathbf{x}_i; \beta), \quad (1)$$

where $f(y_i | \mathbf{x}_i; \beta)$ is the conditional probability density (or mass) function of y_i given \mathbf{x}_i and parameter vector β . For computational simplicity, the log-likelihood function is often used:

$$\ell(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta}) = \sum_{i=1}^n \log f(y_i | \mathbf{x}_i; \boldsymbol{\beta}). \quad (2)$$

The goal is to find the parameter vector $\hat{\boldsymbol{\beta}}$ that maximizes the log-likelihood function:

$$\hat{\boldsymbol{\beta}} = \arg \max_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta}). \quad (3)$$

Due to the nonlinearity and non-convexity of the likelihood function in many practical cases, an analytical solution for $\hat{\boldsymbol{\beta}}$ is often intractable. Thus, numerical optimization techniques such as the Newton-Raphson or Fisher Scoring algorithms are employed. These iterative methods update the parameter estimates as follows:

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - \mathbf{H}^{-1}(\boldsymbol{\beta}^{(t)}) \mathbf{s}(\boldsymbol{\beta}^{(t)}), \quad (4)$$

where $\mathbf{s}(\boldsymbol{\beta})$ is the score vector (the gradient of the log-likelihood), and $\mathbf{H}(\boldsymbol{\beta})$ is the Hessian matrix (the matrix of second-order partial derivatives). However, these methods may converge slowly or get trapped in local optima, especially in cases where the likelihood surface has multiple peaks. In such cases, metaheuristic optimization methods like Particle Swarm Optimization can be advantageous.

2.2 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) is a nature-inspired, population-based metaheuristic algorithm originally developed to simulate the social behavior of bird flocking or fish schooling. It is particularly suitable for solving complex, non-linear, and non-convex optimization problems, including those involving multi-modal objective functions where classical techniques face convergence difficulties (Abualigah, 2025).

In PSO, each particle represents a potential solution in the parameter space. Over successive iterations, particles adjust their positions and velocities based on their own experience and that of their neighbors, aiming to converge toward the global optimum. The algorithm is governed by the following equations:

Velocity update equation:

$$\mathbf{v}_i^{(t+1)} = w \mathbf{v}_i^{(t)} + c_1 r_1 (\mathbf{pbest}_i - \mathbf{x}_i^{(t)}) + c_2 r_2 (\mathbf{gbest} - \mathbf{x}_i^{(t)}), \quad (5)$$

Position update equation:

$$\mathbf{x}_i^{(t+1)} = \mathbf{x}_i^{(t)} + \mathbf{v}_i^{(t+1)}, \quad (6)$$

where w is the inertia weight, c_1 and c_2 are cognitive and social acceleration coefficients, respectively, and r_1, r_2 are uniformly distributed random numbers in the interval $[0, 1]$. \mathbf{pbest}_i denotes the best position discovered by the i -th particle, while \mathbf{gbest} represents the best global position found by the entire swarm (Gad, 2022).

The PSO algorithm proceeds until a stopping criterion is met, such as reaching a maximum number of iterations or achieving convergence in the objective function value.

PSO has been successfully applied in estimating parameters of complex statistical models, including differential equations and nonlinear regressions. Its robustness against multimodal likelihood surfaces and multicollinearity makes it a powerful alternative to classical methods, particularly in noisy and high-dimensional datasets (Shami et al., 2022).

3 Research Methodology

This study compares two prominent methods for estimating the parameters of binary regression models, namely logistic and probit regression. The first method is the classical *Maximum Likelihood Estimation* (MLE), a well-established statistical technique for parameter estimation

in probabilistic models. The second is the *Particle Swarm Optimization* (PSO) algorithm which has demonstrated considerable effectiveness in solving complex nonlinear optimization problems. The primary goal is to evaluate the accuracy and efficiency of these methods in estimating the coefficients of binary regression models.

3.1 Binary Regression Models Under Study

3.1.1 Logistic Regression

Logistic regression is a nonlinear model used when the dependent variable is binary, i.e., $y_i \in \{0, 1\}$. The model estimates the conditional probability of the outcome being 1 using the logistic (sigmoid) function:

$$P(Y = 1|X) = \frac{1}{1 + \exp(-(\beta_0 + \beta^T X))} \quad (7)$$

where X is the vector of independent variables, β_0 is the intercept, and β is the vector of model coefficients (Das, 2024).

MLE of Parameters in Logistic Regression For n independent observations, the likelihood function is given by:

$$L(\beta) = \prod_{i=1}^n [P(Y_i = 1|X_i)]^{Y_i} [1 - P(Y_i = 1|X_i)]^{1-Y_i} \quad (8)$$

The log-likelihood is:

$$\ell(\beta) = \sum_{i=1}^n [Y_i \log(P(Y_i = 1|X_i)) + (1 - Y_i) \log(1 - P(Y_i = 1|X_i))] \quad (9)$$

The MLE seeks to maximize $\ell(\beta)$:

$$\hat{\beta}_{MLE} = \arg \max_{\beta} \ell(\beta) \quad (10)$$

Since this optimization problem lacks a closed-form solution, numerical methods such as Newton-Raphson or Fisher Scoring are typically used (Nagarajah and Wijekoon, 2015).

PSO of Parameters in Logistic Regression The PSO algorithm minimizes the negative log-likelihood, which serves as the fitness function:

$$f(\beta) = -\ell(\beta) = -\sum_{i=1}^n \left[Y_i \log \left(\frac{1}{1 + \exp(-(\beta_0 + \beta^T X_i))} \right) + (1 - Y_i) \log \left(1 - \frac{1}{1 + \exp(-(\beta_0 + \beta^T X_i))} \right) \right] \quad (11)$$

PSO attempts to minimize $f(\beta)$ by iteratively updating particles representing potential solutions in the search space.

3.1.2 Probit Regression

Probit regression is another method for modeling binary outcomes, but it uses the standard normal cumulative distribution function $\Phi(\cdot)$ instead of the logistic function. The model is derived from a latent variable formulation (Hsiao, 1996):

$$Y_i^* = \beta_0 + \beta^T X_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, 1) \quad (12)$$

The observed binary response is:

$$Y_i = \begin{cases} 1, & \text{if } Y_i^* > 0 \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

Thus, the probability of success is:

$$P(Y = 1|X) = \Phi(\beta_0 + \beta^T X) = \int_{-\infty}^{\beta_0 + \beta^T X} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz \quad (14)$$

MLE of Parameters in Probit Model The likelihood for n independent observations is:

$$L(\beta) = \prod_{i=1}^n \left[\Phi(\beta_0 + \beta^T X_i) \right]^{Y_i} \left[1 - \Phi(\beta_0 + \beta^T X_i) \right]^{1-Y_i} \quad (15)$$

The log-likelihood is:

$$\ell(\beta) = \sum_{i=1}^n \left[Y_i \log \Phi(\beta_0 + \beta^T X_i) + (1 - Y_i) \log(1 - \Phi(\beta_0 + \beta^T X_i)) \right] \quad (16)$$

The MLE estimate is then:

$$\hat{\beta}_{\text{MLE}} = \arg \max_{\beta} \ell(\beta) \quad (17)$$

PSO of Parameters in Probit Regression Similar to the logistic model, the negative log-likelihood is used as the cost function in PSO:

$$f(\beta) = -\ell(\beta) = -\sum_{i=1}^n \left[Y_i \log \Phi(\beta_0 + \beta^T X_i) + (1 - Y_i) \log(1 - \Phi(\beta_0 + \beta^T X_i)) \right] \quad (18)$$

3.2 Data Collection and Preprocessing

The datasets used in this research include the well-known **Iris** (Mithy et al., 2022) and **Titanic** (Grandhe et al., 2019) datasets. For the Iris dataset, 100 samples were selected from two balanced classes. In the Titanic dataset, 714 valid samples were extracted after removing missing entries and ensuring class balance. Both datasets were split randomly into 70% training and 30% testing sets. Data preprocessing involved categorical-to-numeric conversion, handling of missing values, and randomization. All preprocessing was conducted in R version 4.1.0.

3.3 Performance Evaluation

To comprehensively evaluate the estimation accuracy of MLE and PSO methods, two well-known error metrics were employed:

3.3.1 Mean Squared Error (MSE)

This metric measures the average of the squared differences between the true coefficients (β) and the estimated coefficients ($\hat{\beta}$):

$$\text{MSE} = \frac{1}{K} \sum_{i=1}^K (\beta_i - \hat{\beta}_i)^2 \quad (19)$$

A lower MSE indicates higher estimation precision and sensitivity to large deviations or outliers.

3.3.2 Mean Absolute Error (MAE)

MAE evaluates the average absolute difference between the true and estimated parameters:

$$\text{MAE} = \frac{1}{K} \sum_{i=1}^K |\beta_i - \hat{\beta}_i| \quad (20)$$

Unlike MSE, MAE is more robust to outliers and offers a more stable measure of average estimation error (Ghareib, 1996).

4 Computational Implementation of the PSO Algorithm

Despite the theoretical formulation of Particle Swarm Optimization (PSO) provided earlier, the practical success of PSO in parameter estimation heavily depends on its computational implementation. This section elaborates on the step-by-step structure, parameter settings, and algorithmic flow used for estimating logistic and probit regression parameters.

4.1 PSO Algorithm Steps for Parameter Estimation

The implementation of PSO for logistic and probit regression involves the following steps:

1. **Initialization:** A swarm of N particles is generated randomly in the d -dimensional parameter space, where d is the number of model coefficients (including the intercept). Each particle i is assigned an initial position $\mathbf{x}_i^{(0)}$ and velocity $\mathbf{v}_i^{(0)}$.
2. **Fitness Evaluation:** For each particle, the negative log-likelihood (NLL) is computed as the fitness function:

- Logistic Model:

$$f(\boldsymbol{\beta}) = - \sum_{i=1}^n [Y_i \log(p_i) + (1 - Y_i) \log(1 - p_i)], \quad p_i = \frac{1}{1 + \exp(-\mathbf{x}_i^\top \boldsymbol{\beta})}$$

- Probit Model:

$$f(\boldsymbol{\beta}) = - \sum_{i=1}^n \left[Y_i \log(\Phi(\mathbf{x}_i^\top \boldsymbol{\beta})) + (1 - Y_i) \log(1 - \Phi(\mathbf{x}_i^\top \boldsymbol{\beta})) \right]$$

where $\Phi(\cdot)$ is the standard normal CDF.

3. **Personal and Global Best Update:** For each particle, if its current fitness is better than its historical best (pbest), then update pbest. Then, identify the best-performing particle across the swarm and update the global best (gbest).

4. **Velocity Update:**

$$\mathbf{v}_i^{(t+1)} = w \mathbf{v}_i^{(t)} + c_1 r_1 (\mathbf{pbest}_i - \mathbf{x}_i^{(t)}) + c_2 r_2 (\mathbf{gbest} - \mathbf{x}_i^{(t)})$$

where:

- w is the inertia weight (e.g., 0.7)
- c_1, c_2 are the cognitive and social learning coefficients (e.g., 1.5)
- $r_1, r_2 \sim \mathcal{U}(0, 1)$ are random numbers

5. **Position Update:**

$$\mathbf{x}_i^{(t+1)} = \mathbf{x}_i^{(t)} + \mathbf{v}_i^{(t+1)}$$

6. **Boundary Handling:** Positions and velocities are clamped within predefined bounds to prevent divergence:

$$\mathbf{x}_i^{(t+1)} \in [\mathbf{x}_{\min}, \mathbf{x}_{\max}]$$

7. **Termination:** Repeat steps 2–6 until the maximum number of iterations is reached (e.g., 200), or until the change in the best solution is below a convergence threshold ϵ (e.g., 10^{-6}).

4.2 Parameter Settings Used in This Study

The PSO implementation used in this study adopts the following parameter settings:

- Swarm size (N): 30
- Inertia weight (w): 0.7 (with optional linear decay)
- Cognitive coefficient (c_1): 1.5
- Social coefficient (c_2): 1.5
- Maximum iterations: 200
- Convergence threshold: 10^{-6}
- Bounds: $\beta \in [-100, 100]$ for all dimensions

4.3 Advantages of PSO in Regression Parameter Estimation

Unlike gradient-based methods, PSO does not require derivative information and is less sensitive to initial parameter values. This is particularly beneficial for non-convex likelihood surfaces or models with multiple local optima. Furthermore, PSO's parallel structure allows for efficient implementation and scalability across dimensions and datasets.

4.4 Implementation Note

The PSO algorithm was implemented in the R programming language using custom functions built on top of the `ps` package. The code ensures compatibility with both logistic and probit regression models by allowing for flexible specification of the likelihood function.

5 Discussion and Results

In this study, data analysis was conducted on two benchmark datasets: **Iris** and **Titanic**, using both Maximum Likelihood Estimation (MLE) and Particle Swarm Optimization (PSO) methods. For the **Iris** dataset, logistic and probit regression models were fitted using sepal length as the sole predictor. In contrast, for the **Titanic** dataset, age, passenger class, and gender were used as predictors for survival.

The results indicated that the PSO algorithm consistently yielded more stable parameter estimates than the MLE approach in both regression models. Table 1 presents the detailed findings for the **Iris** dataset, including estimated coefficients and error metrics, while Table 2 summarizes the corresponding results for the **Titanic** dataset.

Table 1: Comparison of MLE and PSO methods on the **Iris** dataset

Model	Intercept β_0	Sepal Length Coefficient	MSE	MAE
Logistic (MLE)	-35.2310	6.4082	0.1636	0.2615
Probit (MLE)	-19.9398	3.6261	0.1662	0.2650
Logistic (PSO)	-33.2379	6.0748	0.1599	0.2621
Probit (PSO)	-32.2310	5.9007	0.1628	0.2251

6 Conclusion

This research aimed to compare the accuracy of parameter estimation in binary classification regression models, specifically logistic and probit models, using both the traditional Maximum Likelihood Estimation (MLE) method and the population-based metaheuristic Particle Swarm Optimization (PSO) algorithm.

Table 2: Comparison of MLE and PSO methods on the **Titanic** dataset

Model	Intercept β_0	Age	Class	Gender	MSE	MAE
Logistic (MLE)	7.1370	-0.0376	-1.1668	-2.4521	0.1297	0.2702
Probit (MLE)	4.1103	-0.0208	-0.6559	-1.4447	0.1305	0.2731
Logistic (PSO)	8.1592	-0.0457	-1.4083	-2.7115	0.1291	0.2538
Probit (PSO)	6.5213	-0.0376	-1.1537	-2.0961	0.1291	0.2209

The findings revealed that PSO provided meaningful improvements in performance metrics compared to MLE. In the **Iris** dataset, the probit model estimated using PSO achieved the lowest Mean Squared Error (MSE) of 0.1628, outperforming its MLE counterpart (MSE = 0.1662). Similarly, in the **Titanic** dataset, the use of PSO led to a 1.3% reduction in MSE for the probit model.

Additionally, across all models and estimation techniques, the gender variable was consistently the most influential predictor of survival. These outcomes suggest that PSO is a more robust and efficient alternative for estimating parameters in binary regression models, especially when model stability and error minimization are of concern.

Given these advantages, future research is encouraged to investigate the effectiveness of PSO in other regression structures and across more diverse and complex datasets.

References

- Bunke, H., Henschke, K., Strüby, R. and Wisotzki, C. (1977), Parameter estimation in non-linear regression models, *Statistics: A Journal of Theoretical and Applied Statistics*, **8**(1), 23–40.
- Vasisht, AK. (2007), Logit and probit analysis, *IASRI, Library Avenue, New Delhi–110*, **12**.
- Aziz, NH Abdul, Ibrahim, Zuwairie, Razali, Saifudin, and Aziz, Nor Azlina Ab. (2016), Estimation-based metaheuristics: a new branch of computational intelligence, *Proc. 3rd National Conf. Postgraduate Research*, pp. 469–476.
- Zga, Adel, Zitouni, Farouq, Harous, Saad, Sallam, Karam, Almazyad, Abdulaziz S., Xiong, Guojiang, and Mohamed, Ali Wagdy. (2025), A comparative study of the performance of ten metaheuristic algorithms for parameter estimation of solar photovoltaic models, *PeerJ Computer Science*, **11**, e2646.
- Sadik, Laith, Samui, Pijush, Keawsawasvong, Suraparb, Al-Jeznawi, Duaa, and Samui, Ritaparna. (2025), Harnessing Metaheuristics and Probabilistic Machine Learning for Uncertainty-Aware Bearing Capacity Estimation of Shallow Foundations, *Transportation Infrastructure Geotechnology*, **12**(1), 1–31.
- Kasap, P., and Faouri, A.O. (2025), Maximum likelihood estimation for the two-parameter Maxwell distribution, *Journal of the National Science Foundation of Sri Lanka*, **52**(4).
- Abualigah, L. (2025), Particle Swarm Optimization: Advances, Applications, and Experimental Insights, *Computers, Materials & Continua*, **82**(2).
- Shami, T.M., El-Saleh, A.A., Alswaitti, M., Al-Tashi, Q., Summakieh, M.A., and Mirjalili, S. (2022), Particle Swarm Optimization: A Comprehensive Survey, *IEEE Access*, **10**, 10031–10061.
- Gad, A.G. (2022), Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review, *Archives of Computational Methods in Engineering*, **29**(5), 2531–2561.
- Harris, Jenine K. (2021), Primer on binary logistic regression, *Family Medicine and Community Health*, **9**(Suppl 1), e001290.

- Wilson, Jeffrey R., Lorenz, Kent A., and Selby, Lori P. (2024), Introduction to binary logistic regression, *Modeling Binary Correlated Responses: Using SAS, SPSS, R and STATA*, Springer, pp. 3–18.
- Williams, Richard and Jorgensen, Abigail. (2023), Comparing logit & probit coefficients between nested models, *Social Science Research*, **109**, 102802.
- Das, Abhik. (2024), Logistic regression, In: *Encyclopedia of Quality of Life and Well-Being Research*, pp. 3985–3986, Springer.
- Nagarajah, Varathan and Wijekoon, Pushpakanthie. (2015), Stochastic restricted maximum likelihood estimator in logistic regression model, *Open Journal of Statistics*, **5**(7), 837–851.
- Hsiao, Cheng. (1996), Logit and probit models, in *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, Springer, pp. 410–428.
- Ghareib, Ahmed Hamdy. (1996), Evaluation of logit and probit models in mode-choice situation, *Journal of Transportation Engineering*, **122**(4), 282–290.
- Grandhe, Padmaja, Damarla, Vishnu Priya, and Mohammad, Shaziya. (2019), Extensive data set analysis & prediction using R, *Journal of Physics: Conference Series*, **1228**(1), 012048.
- Mithy, S. A., Hossain, S., Akter, S., Honey, U., and Sogir, S. B. (2022), Classification of iris flower dataset using different algorithms, *International Journal of Scientific Research In*.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Smooth Transition Markov Switching R-vine copulas

Ferdous Mohammadi Basatini^{*1}, Raziieh Jafaraghaie ²

¹Department of Mathematics and Statistics, Shc.C., Islamic Azad University, Shoushtar, Iran.

²Department of Mathematics and Statistics, Shc.C., Islamic Azad University, Shoushtar, Iran.

Abstract

Multivariate GARCH Models are extensively employed to model financial time series. Significant asymmetries are documented in both the marginal distributions and the dependence structures of financial returns during periods of market distress. Copulas are extensively utilized as a methodological tool to separate asymmetry present in the marginal distributions from that in the dependence structures. Smooth transition hyperbolic GARCH models are regarded as capable of providing a continuum of regimes within the marginal distributions. In order to effectively capture the asymmetry inherent in the dependence structure, we utilize a Markov switching copula distribution. To facilitate the generation of distinct tail dependencies for each variable pair, we implement regular vine (R-vine) copulas. The parameters associated with the marginal distributions are estimated independently from the copula distributions. Our estimation methodology employs a two-step maximization strategy that is fundamentally grounded in the EM algorithm. Empirical analysis of real data serves to further validate the proposed model.

Keywords: Smooth transition, Markov switching, Hyperbolic GARCH, R-Vine copulas.

Mathematics Subject Classification (2020): 91B84, 60J10, 62H05.

1 Introduction

Multivariate GARCH (MGARCH) Models are widely used as a tool to modelling conditional dependence for multivariate financial time series such as the dynamic correlation (DCC)-GARCH models of Engle (2002). The assumptions on the distribution of multivariate return series are often limited to normal distribution, t distribution or other elliptical distributions.

In the literature on financial time series, different behaviours during times of market stress has long been recognized, where return series tend to exhibit asymmetric dependence in these situations. This asymmetry means that in times of crisis, returns tend to be more dependent than they are in good times where the shape and level of correlations change over time. They are typically higher in more volatile periods and are different for positive and negative returns.

^{*}Corresponding author, fe.mohamadi2024@iau.ac.ir

A class of models addressing this phenomenon, are Markov switching (MS), also called regime switching or hidden Markov models [Hamilton \(1989\)](#). There, the behaviour of a financial time series exhibits two or more distinct regimes, which correspond to different states of the economy. The regime is present at a particular point of time is governed by an underlying hidden Markov process [Stöber and Czado \(2014\)](#).

[Patton \(2004\)](#) found significant asymmetry both in the marginal distributions and in the dependence structures of financial returns. Copulas allow to separate asymmetry in the marginals from asymmetry in dependencies. They provide more flexibility in modelling the marginals from the dependence structure.

Apart from the multivariate Gaussian and t copulas, the set of higher-dimensional copulas are available in the literature is rather limited. Further, characteristics of financial data such as the asymmetric dependence and tail dependence must be incorporated appropriately. Choice of copulas is usually thought to be reduced to the Gaussian or the t-Student. Both of these copulas are useful only for capturing linear dependence. The Gaussian copula suffers from the drawback that it lacks tail dependence, and the multivariate t-Student copula is too restrictive in the sense that, while it can generate different tail dependences for each pair of variables, it restricts the upper and lower tail dependences for each pair to be the same. For real life data, the dependence structure among pairs of variables may vary substantially, ranging from independence to complex nonlinear dependence. No existing family of multivariate copulas can handle such a wide range of dependence for each pair of variables.

An alternative way to make multivariate copulas is regular vine (R-vine) copulas [Bedford and Cook \(2002\)](#). The aim is to combine different types of margins and corresponding pair-copulas inside a multivariate distribution. R-vine distributions are built up hierarchically from bivariate copulas as the building blocks called "pair-copula construction" (PCC). The individual pair-copulas may belong to any parametric or non-parametric family. Therefore, all types and strengths of dependence maybe accommodated in the model. [Chollete and *et al.* \(2009\)](#) modelled the international financial returns with regime-switching canonical vine copula and they found that ignoring asymmetric dependence and regime switching selection leads to significant costs for an investor. [Stöber and Czado \(2014\)](#) studied sudden switches in dependence structures using Markov switching regular vine copulas.

Our contribution is to introduce a smooth transition Markov switching R-vine (STMS-RV) copula model. In STMS-RV, the margins follow a hyperbolic GARCH model which has a smooth transition structure [Valizadeh and *et al.* \(2021\)](#). The smooth transition behaviour are imposed based on a logistic function of the last observation and provide a much better fitting for the volatilities. The smooth transition models are an extension of the two-regime models which allow a continuum of regimes between two extreme regimes associated with the extreme value of the transition function [Mohammadi Basatini and Rezakhah \(2019\)](#). In our proposed model the dependence structures are Markov switching R-vine copulas. Parameters are estimated via a two-step estimation procedure.

2 The Smooth Transition Markov Switching Copula

2.1 Regular vine distribution

R-vines as a graph theoretic tool for the construction of multivariate distribution build as follow. An R-vine v on d variables, which consists of a sequence of connected trees T_1, \dots, T_{d-1} , with nodes N_i and edges E_i , $1 \leq i \leq d-1$, satisfies the following properties:

1. T_1 is a tree with nodes $N_1 = \{1, \dots, d\}$ and edges E_1 .
2. For $i \geq 2$, T_i is a tree with nodes $N_i = E_{i-1}$ and edges E_i .
3. If two nodes in T_{i+1} are joined by an edge, the corresponding edges in T_i must share a common node.

The main advantage of PCC is that it is not require that all the associated bivariate pair-copula

belong to the same family. Two main types of PCCs have been proposed in the literature; canonical vines (C-vine) and D-vines. Here we concentrate on D-vines. Let $\mathbf{X} = (X_1, \dots, X_d)$ be a random vector with marginal densities f_1, \dots, f_d , respectively. R-vine structure associates to each edge $u(e), v(e) \mid D(e)$ in E_i , $1 \leq i \leq d-1$ a bivariate copula density $c_{u(e), v(e) \mid D(e)}$. Where $u(e)$ and $v(e)$ are the conditioned set and $D(e)$ is the conditioning set. The joint density of \mathbf{X} is

$$f_{1, \dots, d}(x_1, \dots, x_d) = \prod_{i=1}^{d-1} \prod_{e \in E_i} c_{u(e), v(e) \mid D(e)}(F_{u(e) \mid D(e)}(x_{u(e)} \mid \mathbf{X}_{D(e)}), F_{v(e) \mid D(e)}(x_{v(e)} \mid \mathbf{X}_{D(e)})) \prod_{i=1}^d f_i(x_i). \quad (1)$$

Where $\mathbf{X}_{D(e)}$ denotes the sub-vector of \mathbf{X} determined by the set of indices $D(e)$ and "f" and "F" denote probability density function and cumulative distribution function respectively. The PCCs are a cascade of $d(d-1)/2$ bivariate copula densities, of which the first $d-1$ unconditional and the rest conditional. As an example, consider $d = 3$ where $\mathbf{X} = (X_1, X_2, X_3)$. Following the previous construction, the PCC expansion becomes

$$\begin{aligned} f(x_1, x_2, x_3) &= f(x_1)f(x_2)f(x_3) \\ &\times c_{1,2}(F(x_1), F(x_2))c_{2,3}(F(x_2), F(x_3)) \\ &c_{1,3|2}(F(x_1 \mid x_2), F(x_3 \mid x_2)). \end{aligned}$$

Fig. 1 shows the tree representation of a D-vine, in five dimensions. It consists of four trees arranged in four levels. An edge of a tree corresponds to a pair-copula density denoted by the edge label. For example, the edge 13|2 denotes the pair-copula $c_{13|2}$. The whole structure has $d(d-1)/2$ edges.

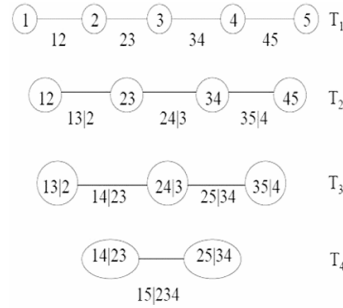


Figure 1: D-vine representation of 5 variables, 4 trees and 10 edges. Each edge is associated with pair-copula.

2.2 Markov switching copula

To model the dependence we use MS copula with two regimes, characterized by different levels or shapes of dependence. Let $\mathbf{X}_t = (X_{1,t}, \dots, X_{d,t})$, $t = 1, \dots, T$ denote a multivariate financial time series. The dependency among \mathbf{X}_t depends on a hidden latent Markov chain Z_t with values $k = 1, 2$. These are called regimes and represent the different states of the economy. It is assumed that Z_t is a homogeneous first order Markov chain with transition probability matrix

$$\mathbf{P} = \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{bmatrix},$$

where $P_{kk} = P(Z_t = k \mid Z_{t-1} = k)$, $k = 1, 2$ and with stationary probabilities $\pi = [\pi_1, \pi_2]$ (Mohammadi Basatini and Rezakhah (2020)). In our model the regime only affects

the dependence structure and the marginal distributions do not depend on regime. So the Markov-switching density of \mathbf{X}_t conditional on being in regime k , is

$$f(\mathbf{X}_t | \mathbf{X}^{t-1}, z_t = k) = \prod_{i=1}^{i=d-1} \prod_{e \in E_i} c_{u(e), v(e) | D(e)}^{(k)}(F_{u(e) | D(e)}(x_{u(e), t} | \mathbf{X}_{D(e)}), F_{v(e) | D(e)}(x_{v(e), t} | \mathbf{X}_{D(e)}); \theta_c^{(k)}) \\ \cdot \prod_{i=1}^d f_i(x_{i, t}; \theta_{m, i}), \quad (2)$$

where $\mathbf{X}^t = (\mathbf{X}'_1, \dots, \mathbf{X}'_t)'$ be the available information up to t . Also $\theta_c^{(k)}$ denotes the parameters of the copula in regime k and $\theta_{m, i}$ denotes the parameters of the marginal distributions.

2.3 Smooth transition hyperbolic GARCH model

In STMS-RV we allow the marginal distribution of each one of the returns follows a smooth transition hyperbolic GARCH (ST-HGARCH) model (Valizadeh and *et al.*, 2021).

Specifically $x_{i, t}$ follows the ST-HGARCH(q, r, p) model as

$$x_{i, t} = \epsilon_{i, t} \sqrt{h_{i, t}} \quad i = 1, \dots, d \\ h_{i, t} = \frac{\gamma_i}{\beta_i(1)} + w_{i, t} [1 - \frac{\delta_i(B)}{\beta_i(B)} (1 - B)^r] x_{i, t}^2, \quad (3)$$

where $\epsilon_{i, t}$ are identically and independently random variables with mean 0 and variance 1, $\gamma_i > 0$, B is the back-shift operator, $\beta_i(y) = 1 - \sum_{j=1}^p \beta_{i, j} y^j$, $\delta_i(y) = 1 - \sum_{j=1}^q \delta_{i, j} y^j$ and p, q are known positive integers; $(1 - B)^r = 1 - \sum_{j=1}^\infty g_j B^j$ where $g_j = \frac{r \Gamma(j - r)}{\Gamma(1 - r) \Gamma(j + 1)}$ in which $0 < r < 1$. Also $h_{i, t}$ is the conditional variance of $x_{i, t}$ as, $\text{Var}(x_{i, t} | x_{i, t}^{t-1}) = h_{i, t}$, where $x_i^t = (x_{i, 1}, \dots, x_{i, t})$ denotes the history of the variable i . The parameter $w_{i, t}$ is called the amplitude parameter that determines the magnitude of variations in the conditional variance.

Here we consider $w_{i, t}$ as a logistic function defined as

$$w_{i, t} = \frac{\exp(\eta_i x_{i, t-1}^2)}{1 + \exp(\eta_i x_{i, t-1}^2)}. \quad (4)$$

It is clear that $w_{i, t}$ bounded between 0 and 1; $\eta_i > 0$ is called the smoothness parameter which determines the speed of transition between high and low volatilities. So the amplitude changes with the size of the last observation and hence the magnitude of the last shock causes smooth changes of the conditional variance.

3 Estimation

The proposed model has the advantage that the marginal parameters and the copula parameters are estimated separately in a two-step procedure. We use EM algorithm to estimate the RS copula parameters.

3.1 Two-step estimation

In STMS-RV model the marginal are not regime-switching, so we use this advantage to separate the estimation into two steps. Let $\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_T)'$ denote all the data, the total loglikelihood is given by

$$L(\mathbf{X}; \theta_m, \theta_c) = \sum_{t=1}^T \log f(\mathbf{X}_t | \mathbf{X}^{t-1}; \theta_m; \theta_c), \quad (5)$$

where θ_m and θ_c denote the parameters of the marginals and of the RS copula, respectively. The likelihood decompose in two part; L_m that contains the marginal densities and L_c which contains the dependence structure:

$$\begin{aligned} L(\mathbf{X}; \theta_m, \theta_c) &= L_m(\mathbf{X}; \theta_m) + L_c(\mathbf{X}; \theta_m, \theta_c), \\ L_m(\mathbf{X}; \theta_m) &= \sum_{t=1}^T \sum_{i=1}^d \log f_i(x_i | x_i^{t-1}; \theta_{m,i}), \\ L_c(\mathbf{X}; \theta_m, \theta_c) &= \sum_{t=1}^T \log c(F_1(x_{1,t} | x_1^{t-1}; \theta_{m,1}), \dots, F_d(x_{d,t} | x_d^{t-1}, \theta_{m,d}); \theta_c), \end{aligned} \quad (6)$$

where $x_i^t = (x_{i,1}, \dots, x_{i,t})$ denotes the history of the variable i . For simplicity we show the dependency structure by $c(F_1(x_{1,t} | x_1^{t-1}; \theta_{m,1}), \dots, F_d(x_{d,t} | x_d^{t-1}, \theta_{m,d}); \theta_c)$ which follows R-vine copulas. The likelihood of the marginal models, L_m is a function of the parameter vector $\theta_m = (\theta_{m,1}, \dots, \theta_{m,d})$, that collects the parameters of each one of the d marginal densities f_i . The copula likelihood, L_c , depends directly on the vector $\theta_c = (\theta_c^{(1)}, \theta_c^{(2)}, p_{11}, p_{22})$. This vector collects the copula parameters over both regimes as well as the parameters of the Markov transition probability matrix. It also depends indirectly on the parameters of the marginal densities, through the distribution function F_i , because F_i transforms observations into uniform $[0, 1]$ variables that are the input of the copula. The function "c" denotes the density of the RS copula. There are a large number of parameters in STMS-RV model, that should be estimated. In addition, the non-linearity in the copula increases the difficulty of the estimation. So we rely on a two-step maximization instead of full one-step maximization. In the first step, we estimate the marginal parameters:

$$\hat{\theta}_m = \arg \max_{\theta_m} L_m(\mathbf{X}; \theta_m).$$

This is straightforward, as it does not depend on the RS and, in addition, it can be simplified further by noting that we can actually estimate each ST-HGARCH model separately:

$$\hat{\theta}_{m,i} = \arg \max_{\theta_{m,i}} \sum_{t=1}^T \log f_i(x_{i,t} | x_i^{t-1}; \theta_{m,i}).$$

Let vector, $\hat{\theta}_m = (\hat{\theta}_{m,i}, \dots, \hat{\theta}_{m,d})$ denotes marginal estimated parameters, in the second step we use $\hat{\theta}_m$ to estimate the parameters of the regime switching copula:

$$\hat{\theta}_c = \arg \max_{\theta_c} L_c(\mathbf{X}; \hat{\theta}_m, \theta_c).$$

3.2 EM algorithm

The parameter of the RS copula are estimated conditional on the estimation of the marginal models [Chollete and et al. \(2009\)](#). Knowing the fact that the Markov chain Z_t is hidden, we can apply the filter of [Hamilton \(1989\)](#). explicitly, the filtered system follows

$$\hat{\xi}_{t|t} = \frac{\hat{\xi}_{t|t-1} \odot \eta_t}{\mathbf{1}'(\hat{\xi}_{t|t-1} \odot \eta_t)} \quad (7)$$

$$\hat{\xi}_{t+1|t} = \mathbf{P}' \hat{\xi}_{t|t}, \quad (8)$$

$$\eta_t = \begin{pmatrix} c^{(1)}(F_1(x_{1,t} | x_1^{t-1}; \theta_{m,1}), \dots, F_d(x_{d,t} | x_d^{t-1}, \theta_{m,d}); \theta_c^{(1)}) \\ c^{(2)}(F_1(x_{1,t} | x_1^{t-1}; \theta_{m,1}), \dots, F_d(x_{d,t} | x_d^{t-1}, \theta_{m,d}); \theta_c^{(2)}) \end{pmatrix} \quad (9)$$

and we obtain all probabilities which are required to evaluate $L_c(\mathbf{X}; \hat{\theta}_m, \theta_c)$ recursively. Where $\hat{\xi}_{t|t} = (P(Z_t = 1 | \mathbf{X}^t, \theta_m, \theta_c), P(Z_t = 2 | \mathbf{X}^t, \theta_m, \theta_c))$ is the (2×1) vector containing the probability of being in each regime at time t , conditional on the observations up to time t ; $\mathbf{1}$ is a (2×1) vector of 1s; and \odot denotes componentwise multiplication of two vectors. The (2×1)

Table 1: Descriptive statistics for the Oil daily log returns.

Series	Mean	Std.dev	Minimum	Maximum	Skewness	Kurtosis
CO	0	0.037	-0.281	0.426	1.334	37.335
CG	-0.011	2.997	-29.973	22.2358	-1.224	22.441
HO	0.014	2.470	-18.444	14.891	-0.374	8.293

vector $\hat{\xi}_{t+1|t}$ gives these probabilities at time $t + 1$ conditional on observations up to time t . The vector η_t contains the copula density at time t , conditional on being in each one of the two regimes. Equation (7) corresponds to a Bayesian updating of the probability of being in either regime given present time observations (η_2). Equation (8) consists in doing one forward iteration of the Markov chain. Iterating over both equations from a given starting value $\hat{\xi}_{1|0}$ and parameter values $\theta_c^{(1)}$ and $\theta_c^{(2)}$ of the copula in each of the two regimes and p_{11} and p_{22} of the Markov chain, one obtains the value of the likelihood:

$$L_c(\mathbf{X}; \theta_m, \theta_c) = \sum_{t=1}^T \log(\mathbf{1}'(\hat{\xi}_{t|t-1} \odot \eta_t)). \quad (10)$$

4 Real Data

We study a financial dataset, consists of a trivariate time series of oil and its products prices of $n = 1763$ observations. Oil and its products prices are central to global economic activity, create significant impacts on current and future economic prosperity. We consider three time series of spot price returns: crude oil (CO), conventional gasoline (CG) and heating oil (HO). We use daily observations from 10 January 2014 to 21 January 2021 from the Energy Information Administration (EIA) of the US Department of Energy; we have 1763 valid observations. We have considered the percentage log returns $r_t = \log(P_t/P_{t-1}) * 100$, where P_t shows the price at t . Some descriptive statistics of the data are listed in Table 1.

All series show very clear signs of non-normality with negative skewness for CG and HO and positive skewness for CO. Further evidence of non-normality is given by the fact that all series have a kurtosis that is well above 3. The estimates of each of the univariate t Student ST-HGARCH(1,d,1) models are presented in Table 2. The degrees-of-freedom parameters ν of all series is around 8, which corresponds to the tails of the conditional distribution that are somewhat fatter than those of the normal distribution. Table 3 shows the Kendall's τ of the data. It is obvious the strongest pair is between CO and HO with $\tau = 0.551$. Fig. 2 shows the pairwise scatter plots of the data. Obviously all pairs show symmetric low and up tail dependencies, so we considered t Student copulas as bivariate building blocks. We compare a four-dimensional pair-copula decomposition with t Student copula for all pairs with the Markov switching pair-copula. According to Kendall's τ results, in the first tree, the first edge is between CO and HO; the second edge is between HO and CG. In the second tree there is one edge between CO and CG conditional on HO. Table 4 and 5 present the results. The results of Table 5 indicate that we have a high and a low dependence regime. The copula correlation coefficient in the more dependent regime, say regime 1 is higher for all pairs of variables, which means that the Oil and its products prices more dependent when the economy is in that regime. This regime is characterized by large correlations.

Table 2: The results of t Student ST-HGARCH(1,r,1) marginal models of the Oil daily log- returns.

Series	γ	β	δ	r	η	ν
CO	0.102	0.102	0.403	0.630	0.129	8.601
CG	0.552	0.711	0.251	0.633	0.298	7.720
HO	0.359	0.734	0.239	0.672	0.299	7.879

Table 3: Kendall's τ coefficients of the Oil daily log- returns.

(CO-CG)	(CO-HO)	(CG-HO)
0.476	0.551	0.533

Table 4: The results of the four-dimensional pair-copula decomposition with t Student copula for all pairs of the Oil daily log- returns.

Tree	Edge	Coefficient	ν	τ
1	CO,CG	0.75	4.013	0.54
	HO,CG	0.74	4.41	0.53
2	CO,CG HO	0.28	6.44	0.18

Discussion and Results

We present additional empirical evidence regarding asymmetric dependence in financial returns through the estimation of a STMS-RV copula model specifically for the prices of oil and its derivatives. The ST-HGARCH model is employed to capture the marginal distributions, while MS R-vine copulas are utilized to represent the dependence structure. The ST-HGARCH model offers a superior characterization of dynamic volatilities, and the MS R-vine copulas provide enhanced flexibility in modelling various shapes and levels of dependencies.

References

- Bedford, T. and Cooke, R.M. (2002), Vines—A New Graphical Model for Dependent Random Variables, *Annals of Statistics*, **30**, 1031–1068.
- Chollete, L., Heinen, A. and Valdesogo, A. (2009), Modeling International Financial Returns with a Multivariate Regime-switching Copula, *Journal of Financial Econometrics*, **7**, (4), 437–480.
- Engle, R. (2002), Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics*, **20**, 339–350
- Hamilton, J.D. (1989), A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, **57**, 357–384.

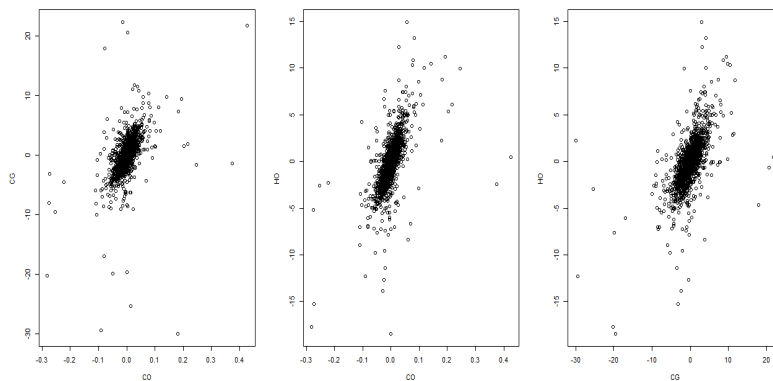


Figure 2: Pairwise scatter plots between crude oil (CO), conventional gasoline (CG) and heating oil (HO) .

Table 5: The results of the MS-copula from t Student D-vine copulas for all pairs of the Oil daily log- returns.

		Regime 1		Regime 2	
Tree	Edge	Coefficient	τ	Coefficient	τ
1	CO,CG	0.92	0.743	0.59	0.40
	HO,CG	0.80	0.59	0.48	0.32
2	CO,CG HO	0.71	0.50	0.40	.27

Mohammadi Basatini, F. and Rezakhah, S. (2019), Smooth transition HYGARCH model: Stability and testing, *Journal of Fluctuation and Noise Letters*, **18** (4), 1950025-1 :1950025-23.

Mohammadi Basatini, F. and Rezakhah, S. (2020), Markov switch smooth transition HYGARCH model : Stability and estimation, *Journal of Communication in Statistics-Theory and Methods*, **49**(10), 2384–2409.

Patton, A. (2004), On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation, *Journal of Financial Econometrics*, **2**, 130–168.

Stöber, J. and Czado C. (2014), Regime switches in the dependence structure of multidimensional financial data, *Journal of Computational Statistics and Data Analysis*, **76**, 672-686.

Valizadeh, T., Rezakhah, S. and Mohammadi Basatini, F. (2021), On Time-Varying Amplitude HGARCH Model, *International journal of finance and economics*, **26**, 2538–2547.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Bayesian Neural Networks using Variational Bayes

Moein Monemi[‡], Morteza Amini², S.M Taheri¹, Ali Fahim¹

¹School of Engineering Science, College of Engineering, University of Tehran, Tehran, Iran

²Department of Statistics, School of Mathematics, Statistics, and Computer Science,
College of Science, University of Tehran, Tehran, Iran

Abstract

In this paper, we aim to investigate a method of approximation the posterior distribution of the parameters of a Bayesian neural network (BNN) using variational bayes (VB), which is a computationally efficient and scalable method. The performance of the BNN is evaluated through a curve fitting task by analyzing the confidence intervals of its predictive outputs.

Keywords: Bayesian neural networks, Regression, posterior distribution, Variational Bayes, Kullback-Leibler divergence

Mathematics Subject Classification (2020): 62F15, 68T08.

1 Introduction

Bayesian Neural Networks (BNNs) are stochastic artificial neural networks trained through Bayesian inference techniques (Bishop (1997), Lampinen and Vehtari (2001), Zhang *et al.* (2018), Jospin *et al.* (2022)). By quantifying model uncertainty, BNNs provide a principled approach to addressing certain challenges such as overfitting and poor generalization in standard neural networks. This is achieved by incorporating prior distributions over model parameters—such as weights and biases—and computing their corresponding posterior distribution.

Although BNNs offer several advantages, they are often confronted with significant computational challenges, primarily due to the intractability of the posterior integrals. To address this issue, sampling-based methods such as Markov Chain Monte Carlo (MCMC) (Hastings (1970)) have been proposed. While these methods offer accurate posterior approximations and strong theoretical foundations, they are typically computationally intensive and scale poorly in high-dimensional settings. To overcome these limitations, more alternatives, such as variational inference methods, have been introduced (Jordan *et al.* (1999), Wainwright and Jordan (2008)). These approaches fall under the category of parametric inference and aim to approximate the posterior distribution with a tractable distribution, such as exponential family, by optimizing a divergence measure, most commonly the Kullback–Leibler (KL) divergence.

*Corresponding author moein.monemi@ut.ac.ir

In this paper, we explore a scalable variational inference framework for training Bayesian neural networks, called *Bayes by Backprop* (Blundell *et al.* (2015)). We discuss its theoretical foundations and practical implications. Subsequently, we evaluate the performance of this method on a curve fitting problem to demonstrate its practical effectiveness.

2 Bayesian Neural Networks

In this section, we briefly describe artificial neural networks (ANNs) and Bayesian neural networks (BNNs) from a statistical perspective. An ANN can be represented as a weighted directed graph $G = (V, W)$, where V denotes the set of nodes, and W represents the set of all learnable parameters in the network, including both weights and biases, which need to be estimated. The goal in an ANN is to approximate a function of the form

$$y = \phi(x, W),$$

where, x and y represent the input and output sets, respectively, which may vary depending on the task.

The objective function in a standard ANN can be formulated using a likelihood function, with the goal of maximizing it, i.e.

$$\hat{W}_{\text{MLE}} = \arg \max_W \{\log L(W \mid x, y)\},$$

where, $L(W \mid x, y)$ is the likelihood function, defined as the joint probability distribution of the observed data given the model parameters. For example, in regression tasks, assuming a Gaussian distribution over the outputs y leads to the sum of squared error loss. In classification tasks, assuming a categorical distribution over outputs yields the cross-entropy error.

In a BNN, instead of estimating a point value for the weights, the goal is to obtain the conditional distribution of W given the observations x and y , which can be derived using Bayes' theorem as follows

$$p(W \mid x, y) = \frac{L(W \mid x, y) p(W)}{\int L(W \mid x, y) p(W) dW}, \quad (1)$$

where, $p(W)$ is the prior distribution over the network parameters, encoding prior beliefs about the model parameters. The denominator of the equation (1), referred to as the *evidence*, is often intractable and represents one of the central challenges in Bayesian inference.

When a point estimate is required, one can instead use the maximum a posteriori (MAP) estimate defined as

$$\hat{W}_{\text{MAP}} = \arg \max_W \{\log p(W \mid x, y)\} = \arg \max_W \{\log L(W \mid x, y) + \log p(W)\},$$

which is equivalent to maximizing a regularized objective function, where the prior acts as a regularizer (Prince (2023)).

A key benefit of the Bayesian framework is the ability to compute a full predictive distribution, known as the *posterior predictive distribution*, which is given by

$$p(y^* \mid x^*, x, y) = \int p(y^* \mid x^*, W) p(W \mid x, y) dW, \quad (2)$$

where, x^* is a new input, and y^* is the corresponding predicted output. Equation (2) is equivalent to averaging the outputs of a collection of neural networks, each sampled from the posterior distribution over the parameters.

In the next section, we briefly discuss variational Bayes (VB), a fast and scalable method for approximating the posterior distribution.

3 Variational Bayes

Variational Bayes (VB) is an optimization-based method that provides an efficient alternative to sampling-based approaches such as MCMC. It belongs to the broader family of *Variational Inference* methods, where the objective is to find an approximate distribution q , referred to as the *Variational posterior distribution*, which belongs to a tractable family of distributions, such as the exponential family. This approximate distribution is then used as an approximation to the posterior distribution.

The KL divergence is used as the objective function for approximating the posterior distribution. It is defined as follows

$$\text{KL}[q(W)||p(W|x, y)] = \int q(W) \log \frac{q(W)}{p(W|x, y)} dW \quad (3)$$

The best approximation in VB is obtained when the equation (3) is minimized, as shown below

$$q^* = \arg \min_q \text{KL}[q(W)||p(W|x, y)]$$

Assuming $p(W|x, y)$ is intractable (which is typically the case), another criterion must be defined to perform the minimization. Since the KL divergence is always non-negative, we can easily observe the following relationship

$$\log p(x, y) \geq \mathbb{E}_{q(W)} \left[\log \frac{p(x, y, W)}{q(W)} \right] :- \mathcal{LB}(q)$$

where, $\mathbb{E}_{q(W)} \left[\log \frac{p(x, y, W)}{q(W)} \right] = \int q(W) \log \frac{p(x, y, W)}{q(W)} dW$. The right-hand side of this inequality is known as the *Evidence Lower Bound* (ELBO), and maximizing this bound is equivalent to minimizing the KL divergence expression defined in equation (3). Variational Bayes can be categorized into two approaches based on the restrictions on q : *Mean Field Variational Bayes* (MFVB) and *Fixed Form Variational Bayes* (FFVB) (Graves (2011)). Since we use the FFVB method in this paper, we briefly review this method.

3.1 Fixed Form Variational Bayes (FFVB)

FFVB is a widely used and scalable method for approximating the posterior distribution. In this method, the variational distribution q is assumed to follow a fixed, tractable form, such as a normal distribution. Unlike MFVB, which factorizes the distribution independently, FFVB aims to identify the optimal values for the *Variational Parameters*, denoted by η , of the assumed variational distribution q . For example, if the variational distribution q is a normal distribution, then $\eta = (\mu_q, \sigma_q^2)$ represents the mean and variance of the variational posterior distribution that needs to be estimated. The objective is to maximize the ELBO with respect to η . Thus, the ELBO can be rewritten as

$$\mathcal{LB}(q|\eta) = \int q(W|\eta) \log \frac{p(x, y, W)}{q(W|\eta)} dW$$

Maximizing the ELBO with respect to the variational parameters η is equivalent to minimizing the negative of the ELBO

$$\eta^* = \arg \min_{\eta} \{-\mathcal{LB}(q|\eta)\} = \arg \min_{\eta} \mathbb{E}_{q(W|\eta)} [\log q(W|\eta) - \log p(x, y|W) - \log p(W)]$$

It can be seen that, by assuming a specific form for the distribution q , the above objective function can be minimized with respect to the variational parameters η without any restrictive assumptions on the distribution of q . The approximated distribution q is then used as the posterior approximation. In the next section, we will explore the *Bayes by Backprop* method, which is developed based on these assumptions.

4 Bayes by Backprop

Bayes by Backprop is a method belonging to the class of FFVB approaches for approximating posterior distribution, particularly in BNNs (Blundell *et al.* (2015)). This method utilizes the reparameterization trick to generate samples from the variational posterior and, to this end, maximizes the ELBO via backpropagation. In this method, a loss function is defined, and the goal is to minimize this loss function. Suppose ϵ is a random variable with probability density $q(\epsilon)$, and f is a function. In this case, if

$$q(\epsilon) d\epsilon = q(W|\eta) dW \quad (4)$$

holds, it can be shown that

$$\frac{\partial}{\partial \eta} \mathbb{E}_{q(W|\eta)}[f(W, \eta)] = \mathbb{E}_{q(\epsilon)} \left[\frac{\partial f(W, \eta)}{\partial \eta} \right]. \quad (5)$$

Let $f(W, \eta) = \log q(W|\eta) - \log p(x, y|W) - \log p(W)$. In this case, the expectation of this expression with respect to $q(W|\eta)$ corresponds to the negative ELBO, which needs to be minimized. The gradient of the function $f(W, \eta)$ with respect to the variational parameters η serves as an unbiased estimator of $\frac{\partial}{\partial \eta} \mathbb{E}_{q(W|\eta)}[f(W, \eta)]$, and thus $f(W, \eta)$ can be used as the objective function in the optimization problem. In this method, a transformation function $W = h(\eta, \epsilon)$ is constructed such that sampling from ϵ , together with the variational parameters η , is equivalent to sampling from W , and the condition in (4) is satisfied.

To satisfy the condition in (4), it is usually assumed that the variational posterior $q(W|\eta)$ is a diagonal Gaussian distribution with variational parameter vector $\eta = (\mu, \sigma^2)$. In this approach, the standard deviation σ is parameterized as $\sigma = \log(1 + \exp\{\rho\})$ to ensure non-negativity. Hence, the variational parameter vector can be written as $\eta = (\mu, \rho)$. Given this assumption, if $q(W|\eta)$ is Gaussian, the transformation function can be defined as $W = h(\eta, \epsilon) = \mu + \log(1 + \exp\{\rho\}) \odot \epsilon$, where \odot denotes element-wise multiplication and $\epsilon \sim \mathcal{N}(0, I_{|W|})$, with $|W|$ representing the total number of weights and biases in the neural network. Under these assumptions, it can be shown that the condition in (4) is satisfied. Algorithm 4.1 illustrates the procedure of applying this method to a BNN.

Algorithm 4.1. Bayes by Backprop

- Generate sample $\epsilon \sim \mathcal{N}(0, I_{|W|})$.
- Initialize μ and ρ .
- Let $\eta = (\mu, \rho)$, $f(W, \eta) = \log q(W|\eta) - \log p(x, y|W) - \log p(W)$, and define $W = \mu + \log(1 + \exp\{\rho\}) \odot \epsilon$.
- Compute the gradients with respect to μ and ρ :

$$\nabla_{\mu} f(W, \eta) = \frac{\partial f(W, \eta)}{\partial W} \cdot 1 + \frac{\partial f(W, \eta)}{\partial \mu}$$

$$\nabla_{\rho} f(W, \eta) = \frac{\partial f(W, \eta)}{\partial W} \odot \frac{\epsilon}{1 + \exp\{-\rho\}} + \frac{\partial f(W, \eta)}{\partial \rho}$$

- Update the variational parameters:

$$\mu = \mu - \lambda \nabla_{\mu} f(W, \eta)$$

$$\rho = \rho - \gamma \nabla_{\rho} f(W, \eta)$$

- Repeat until the stopping criterion is satisfied.

In the following section, we consider the curve fitting problem and briefly describe the implementation procedure of this method.

5 Curve Fitting

In this section, we explain how the posterior distribution of a neural network's parameters can be estimated using the Bayes by Backprop method, as discussed in the previous section. Based on the estimated posterior, we then predict the outputs corresponding to each input data point.

Example 5.1. Assume that (y_1, \dots, y_n) are samples generated from the following curve

$$y_i = 4 \cos(2\pi x_i) + 2 \cos(4\pi x_i) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, 1), \quad i = 1, \dots, n. \quad (6)$$

Note that equation (6) is used solely for simulation purposes and is assumed to be unknown during the modeling process.

Based on Example 5.1, we now describe the model assumptions for fitting the target curve. Let $(x_1, y_1), \dots, (x_n, y_n)$ be the observed or training data such that $x_i \in [-1, 1]$ and W be the unknown parameters of a neural network, including weights and biases. In this experiment, we assume a diagonal gaussian prior distribution over W , i.e.,

$$W \sim \mathcal{N}(0, \sigma_p^2 I_{|W|}),$$

where, σ_p^2 is the prior variance, which can be specified by the user. Considering the random sample $(X_1, Y_1), \dots, (X_n, Y_n)$, the likelihood function is formulated as follows

$$L(W|x_1, y_1, \dots, x_n, y_n) = \prod_{i=1}^n p(y_i|x_i, W) \propto \exp \left\{ -\frac{1}{\sigma_L^2} (y_i - \hat{y}_i)^2 \right\},$$

where, σ_L^2 is the likelihood variance, which can be set, for example, based on the maximum likelihood estimate, and \hat{y}_i is the predicted output of the neural network.

The variational posterior distribution $q(W|\mu, \rho)$ under the Bayes by Backprop method is factorized as follows:

$$q(W|\mu, \rho) = \prod_{j=1}^{|W|} \frac{1}{\sqrt{2\pi} \log(1 + \exp\{\rho_j\})} \exp \left\{ -\frac{1}{2 (\log(1 + \exp\{\rho_j\}))^2} (w_j - \mu_j)^2 \right\}$$

in which, w_j , μ_j , and ρ_j denote the j -th weight parameter, its corresponding mean, and the parameter controlling the variance, respectively.

According to Algorithm 4.1, this method can be implemented for a BNN with an arbitrary number of layers. The experiment was conducted on 800 training samples and 200 test samples. Figure 1 illustrates the estimated curves obtained from a BNN in four different runs on the test dataset. The shaded area represents the 95% confidence interval around the predicted mean. This means that each prediction will fall within the upper and lower bounds of the curve with 95% probability, and the predicted point is the midpoint, which represents the mean of the posterior predictive distribution.

6 Discussion

The application of approximate Bayesian inference in Bayesian neural networks was explored using the Bayes by Backprop method, which is a scalable and efficient approach. We implemented this method on a curve fitting task and demonstrated its effectiveness. The results suggest that FFVB-based methods can be suitable for complex statistical models. Although these approaches offer advantages such as high computational speed, they may yield less accurate approximations of the posterior distribution compared to sampling-based methods. Future work may focus on improving the accuracy of such methods. Nevertheless, FFVB techniques remain highly effective for uncertainty quantification and for constructing credible intervals for parameters and predictions.

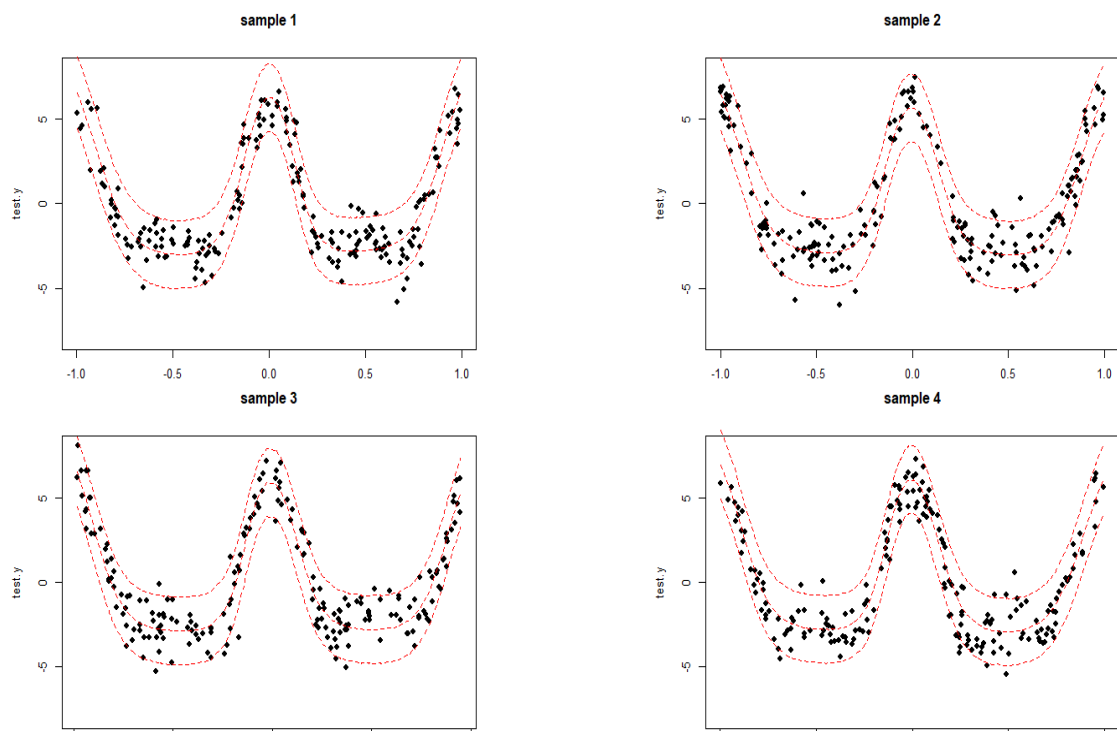


Figure 1: The curves estimated via Bayes by Backprop. The dashed curves represent the predictions of a BNN with 95% confidence intervals, and the points within each curve are the test data.

References

- Bishop, C. M. (1997). Bayesian neural networks. *Journal of the Brazilian Computer Society*, 4, 61-68.
- Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wierstra, D. (2015, June). Weight uncertainty in neural network. In *International Conference on Machine Learning* (pp. 1613-1622). PMLR.
- Graves, A. (2011). Practical variational inference for neural networks. *Advances in neural information processing systems*, 24.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97-109.
- Jospin, L. V., Laga, H., Boussaid, F., Buntine, W., & Bennamoun, M. (2022). Hands-on Bayesian neural networks—A tutorial for deep learning users. *IEEE Computational Intelligence Magazine*, 17(2), 29-48.
- Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., & Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning*, 37, 183-233.
- Lampinen, J., & Vehtari, A. (2001). Bayesian approach for neural networks—review and case studies. *Neural networks*, 14(3), 257-274.
- Prince, S. J. (2023). *Understanding deep learning*. MIT press.
- Wainwright, M. J., & Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2), 1-305.
- Zhang, C., B utepage, J., Kjellstr m, H., & Mandt, S. (2018). Advances in variational inference. *IEEE transactions on pattern analysis and machine intelligence*, 41(8), 2008-2026.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Inhomogeneous Mark Correlation Functions for Point Processes on General State Spaces

Mehdi Moradi^{*1}

¹Department of Mathematics and Mathematical Statistics, Umeå University, Sweden

Abstract

Spatial phenomena in environmental and biological contexts often involve events that are unevenly distributed across space and carry attributes, whose associations/variations are space-dependent. In this paper, we introduce the class of inhomogeneous mark correlation functions, capturing mark associations/variations, while explicitly accounting for the spatial inhomogeneity of events. The proposed functions are designed to quantify how, on average, marks vary or associate with one another as a function of pairwise spatial distances. We develop nonparametric estimators and evaluate their performance through simulation studies covering a range of scenarios with mark association or variation, spanning from nonstationary point patterns without spatial interaction to those characterised by clustering tendencies. Our simulations reveal the shortcomings of traditional methods in the presence of spatial inhomogeneity, underscoring the necessity of our approach. Furthermore, the results show that our estimators accurately identify both the positivity/negativity and effective spatial range for detected mark associations/variations. The proposed inhomogeneous mark correlation functions are then applied to two distinct forest ecosystems: Longleaf pine trees in southern Georgia, USA, marked by their diameter at breast height, and Scots pine trees in Pfywald, Switzerland, marked by their height. Our findings reveal that the inhomogeneous mark correlation functions provide deeper and more detailed insights into tree growth patterns compared to traditional methods.

Keywords: Inhomogeneous mark correlation functions; Mark association; Nonparametric estimation; Spatial inhomogeneity; Spatial point patterns.

Mathematics Subject Classification (2020): 62M30, 62G05, 62P12.

^{*}Corresponding author, mehdi.moradi@umu.se



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Optimizing Generative Adversarial Networks for Financial Time Series: A Stochastic Approach with Novel Loss Metrics

Pegah Moradi^{*1}, Seyed-Mohammad-Mahdi Kazemi²

¹Department of Financial Sciences, Kharazmi University, Tehran, Iran

²Department of Financial Sciences, Kharazmi University, Tehran, Iran

Abstract

Generative models for financial time series often struggle to model practical financial objectives, which limits their applicability. In this study, we propose an improved Generative Adversarial Network framework that combines advanced sequence modeling structures, such as Temporal Convolutional Networks capable of capturing long-range dependencies and volatility patterns, with a novel generator function. This generator function is explicitly constructed to ensure that the resulting stochastic process can transition to its corresponding risk-neutral distribution. It also incorporates a new type of loss function, designed based on incorporating financial performance measures. These components aim to improve the profitability of trading strategies, enhance risk-adjusted returns, and manage the volatility of financial outcomes. Our numerical results demonstrate that the distributional characteristics at both short and long time lags closely match the expected patterns, indicating that the methodology can lead to the generation of more realistic and actionable market scenarios for robust financial decision-making.

Keywords: Generative Adversarial Networks, Temporal Convolutional Networks, Stochastic Process, Time Series

Mathematics Subject Classification (2020): 62M10, 91B84, 65Cxx.

1 Introduction

Time series forecasting has been a fundamental area of research for many years. Most real-world processes naturally follow temporal structures, particularly in finance and economics, where variables such as asset prices, inflation rates, and other macroeconomic indicators evolve over time. In this study, our goal is to move beyond traditional forecasting methods by using Generative Adversarial Networks (GANs) to predict financial time series. GANs have been applied to various financial tasks, including forecasting high-frequency stock market data [Zhou](#)

^{*}Corresponding author, pegah.moradi@khu.ac.ir

and *et al.* (2018), generating synthetic financial time series that replicate key market characteristics such as volatility clustering and leverage effects Jolicoeur-Martineau (2018), and implementing Conditional GANs for trading strategy calibration and aggregation Koshiyama and *et al.* (2021). Here we consider the problem of approximating a realistic asset price simulator by proposing a neural-network-based simulator trained using GAN techniques. This approach positions GANs between purely data-based approaches, such as historical simulations and traditional model-driven methods like Monte Carlo simulations that assume an underlying stock price model like the Black-Scholes model, Heston's stochastic volatility model, or Lévy processes. Unlike these classical methods, GANs offer flexibility without relying on strict assumptions about the underlying price dynamics.

However, applying GANs to sequential data requires models that can effectively capture temporal dependencies. Architectures such as Temporal Convolutional Networks (TCNs) are often integrated within the GAN frameworks to capture these sequential patterns. This specialized architecture allows the generator to produce sequences that respect the inherent order and memory present in time series data, which is essential for financial applications, where past behavior often influences future outcomes.

2 Temporal Convolutional Networks

TCNs are deep learning architectures that have shown strong performance on sequence data and was introduced by Bai and *et al.* (2018). TCNs are built using the concept of dilated causal convolutions, where each output at time t depends only on inputs at time t and earlier. Dilated convolutions allow networks to cover larger contexts without increasing computation. From now on, we assume $N_I, N_O, K, D, T \in \mathbb{N}$, which represent the input and output dimensions, kernel size, dilation factor and the length.

Definition 2.1. Let $X \in \mathbb{R}^{N_I \times T}$ be an N_I -variate sequence of length T and let $W \in \mathbb{R}^{K \times N_I \times N_O}$ be a tensor. For each time step $t \in \{D(K-1)+1, \dots, T\}$ and $m \in \{1, \dots, N_O\}$, we define $*_D$ Operator, which is a dilated causal convolution with dilation D and kernel size K as:

$$(W *_D X)_{m,t} := \sum_{i=1}^K \sum_{j=1}^{N_I} W_{i,j,m} \cdot X_{j,t-D(K-i)}.$$

TCNs are composed by combining dilated causal convolutions with nonlinearities. To improve clarity, a modular design called block module is used. A function $\psi : \mathbb{R}^{N_I \times T} \rightarrow \mathbb{R}^{N_O \times (T-S)}$ that is Lipschitz continuous is a block module with parameters (N_I, N_O, S) .

Definition 2.2. Let $T_0, L, N_0, \dots, N_{L+1} \in \mathbb{N}$, and for each $l \in \{1, \dots, L\}$, let $S_l \in \mathbb{N}$ be such that $\sum_{l=1}^L S_l \leq T_0 - 1$. For $T_l := T_{l-1} - S_l$ it holds:

$$T_L = T_0 - \sum_{l=1}^L S_l \geq 1.$$

Additionally, let $\psi_l : \mathbb{R}^{N_{l-1} \times T_{l-1}} \rightarrow \mathbb{R}^{N_l \times T_l}$ be block modules, and let $w : \mathbb{R}^{N_L \times T_L} \rightarrow \mathbb{R}^{N_{L+1} \times T_L}$ be a 1×1 convolutional layer. The network is defined by the function $f : \mathbb{R}^{N_0 \times T_0} \times \Theta \rightarrow \mathbb{R}^{N_{L+1} \times T_L}$ as follows:

$$f(X, \theta) = w \circ \psi_L \circ \dots \circ \psi_1(X).$$

This is a TCN with L hidden layers. The class of such networks from \mathbb{R}^{d_0} to \mathbb{R}^{d_1} is denoted $TCN_{d_0, d_1, L}$, where $d_0 = N_0$ and $d_1 = N_{L+1}$.

Definition 2.3. Let $f \in TCN_{N_0, N_{L+1}, L}$ be such that each ψ_l is composed of a causal convolution w_l with parameters $(N_{l-1}, N_l, K-l, D_l)$ and an activation function ϕ , i.e. $\psi_l = \phi \circ w_l$. Then $f : \mathbb{R}^{N_0 \times T_0} \times \Theta \rightarrow \mathbb{R}^{N_{L+1} \times T_L}$ is called a Vanilla TCN. If $D_l = D_{l-1}$ for all l , then the dilation factor is said to have a uniform dilation factor D . Similarly, if $K_l = K$ for all l , it is said to have a fixed kernel size K .

TCNs outperform other models in capturing long-term dependencies. We call the number of sequence elements that the TCN can capture the receptive field size [Heaton and Goodfellow \(2016\)](#).

Definition 2.4. Let $f \in \text{TCN}_{d_0, d_1, L}$, and let the sequence of modules S_1, \dots, S_L be defined as in Definition 2.2. The value below is referred to as the receptive field size (RFS).

$$T^{(f)} := 1 + \sum_{l=1}^L S_l.$$

One limitation of TCNs is that the sequence length they can process is constrained by the RFS. Capturing long-term temporal dependencies requires a significantly large RFS, which can lead to computational demands*.

3 Generative Adversarial Networks

GAN models, introduced by [Goodfellow and et al. \(2014\)](#), have become an important concept in deep learning, as they can generate synthetic data that closely follows the characteristics of a given dataset. At their core, GANs employ a competitive framework between two neural networks: a generator and a discriminator. The generator strives to create data samples, that are indistinguishable from real data. Simultaneously, the discriminator tries to accurately differentiate between the real and the synthetic samples produced by the generator. This section begins by outlining the GAN framework for random variables and then expands the discussion to discrete-time stochastic processes, both employed by [Wiese and et al. \(2020\)](#).

3.1 GANs for Random Variables

In this context, assume that $(\mathbb{R}^{N_Z}, \mathcal{B}(\mathbb{R}^{N_Z}))$ and $(\mathbb{R}^{N_X}, \mathcal{B}(\mathbb{R}^{N_X}))$ represent the latent and data measure spaces, respectively. The latent variable Z acts as the input noise (prior), while X represents the target (data) variable. The main purpose of the generator $g : \mathbb{R}^{N_Z} \times \Theta(g) \rightarrow \mathbb{R}^{N_X}$ is to transform the noise into synthetic samples such that the random variable defined by $g_\theta(Z) := g_\theta \odot Z$, for some $\theta \in \Theta(g)$, shares the same distribution as X , denoted by $g_\theta(Z) \stackrel{d}{=} X$.

Definition 3.1. Let $g : \mathbb{R}^{N_Z} \times \Theta(g) \rightarrow \mathbb{R}^{N_X}$ be a neural network parameterized by $\theta \in \Theta(g)$.

$$\tilde{X} : \Omega \times \Theta(g) \rightarrow \mathbb{R}^{N_X}, \quad (\omega, \theta) \mapsto g_\theta(Z(\omega)).$$

where \tilde{X} is the generated random variable, the function g is the generator, and \tilde{X}_θ is the synthetic data generated using parameters θ .

Definition 3.2. Let $\tilde{d} : \mathbb{R}^{N_X} \times \Theta(d) \rightarrow \mathbb{R}$ be a neural network with parameter space $\Theta(d)$ and let $\sigma(x)$ be the activation function. We define the discriminator $d : \mathbb{R}^{N_X} \times \Theta(d) \rightarrow [0, 1]$ by

$$d(x, \eta) := \sigma \odot (\tilde{d}_\eta(x)).$$

The GAN training procedure can be viewed as two players in a zero-sum game. The generator's objective is to produce data points $\{\tilde{x}_{\theta, i}\}_{i=1}^M$ that appear indistinguishable from real samples to the discriminator. Conversely, the discriminator's role is to classify whether a given input comes from the true data distribution or from the generator. The learning procedure is defined in two parts. The discriminator parameters $\eta \in \Theta(d)$ are optimized to maximize the following objective function for a fixed generator parameter $\theta \in \Theta(g)$:

$$L(\theta, \eta) := \mathbb{E}[\log d_\eta(X)] + \mathbb{E}[\log(1 - d_\eta(g_\theta(Z)))] = \mathbb{E}[\log d_\eta(X)] + \mathbb{E}[\log(1 - d_\eta(\tilde{X}_\theta))].$$

This function encourages the discriminator to assign high scores to real data and low scores to generated data. Subsequently, the generator parameters θ are updated to minimize the probability that its outputs are correctly identified as synthetic. The overall training is structured as below, which is known as the standard GAN training objective.

$$\min_{\theta \in \Theta(g)} \max_{\eta \in \Theta(d)} L(\theta, \eta),$$

*Relevant definitions for this section can be found in the electronic appendix, accessible via the following [link](#).

3.2 GANs for Stochastic Processes

To generalize GANs for modeling sequences or time series, we consider a stochastic process $(X_t)_{t \in \mathbb{Z}}$ parameterized by $\theta \in \Theta$. For $s, t \in \mathbb{Z}$ with $s \leq t$, we define:

$$X_{s:t,\theta} := (X_{s,\theta}, \dots, X_{t,\theta}), \quad X_{s:t,\theta}(\omega) := (X_{s,\theta}(\omega), \dots, X_{t,\theta}(\omega)) \in \mathbb{R}^{N_X \times (t-s+1)}.$$

Definition 3.3. Let $(Z_t)_{t \in \mathbb{Z}}$ be a sequence of i.i.d. noise process with values in \mathbb{R}^{N_Z} . Let

$$g : \mathbb{R}^{N_Z \times T^{(g)}} \times \Theta(g) \rightarrow \mathbb{R}^{N_X},$$

be a TCN with receptive field size $T^{(g)}$. Define the neural process \tilde{X} as:

$$\tilde{X} : \Omega \times \mathbb{Z} \times \Theta(g) \rightarrow \mathbb{R}^{N_X}, \quad (\omega, t, \theta) \mapsto g_\theta(Z_{t-(T^{(g)}-1):t}(\omega)).$$

Then the sequence $\tilde{X}_\theta := (\tilde{X}_{t,\theta})_{t \in \mathbb{Z}}$ is called a neural stochastic process, with each $\tilde{X}_{t,\theta}$ being $F_t - \mathbb{B}(\mathbb{R}^{N_X})$ -measurable.

In our framework, the noise input $Z = (Z_t)_{t \in \mathbb{Z}}$ follows a multivariate standard normal distribution, i.e., $Z_t \sim \mathcal{N}(0, I)$. The generator maps this noise sequence into a synthetic process \tilde{X}_θ using a TCN. The discriminator is also defined using a TCN:

$$d : \mathbb{R}^{N_X \times T^{(d)}} \times \Theta(d) \rightarrow [0, 1],$$

where $T^{(g)}$ is the receptive field size and the objective in this temporal setting becomes:

$$\min_{\theta \in \Theta(g)} \max_{\eta \in \Theta(d)} L(\theta, \eta),$$

where

$$L(\theta, \eta) := \mathbb{E}[\log d_\eta(X_{1:T^{(d)}})] + \mathbb{E}[\log(1 - d_\eta(\tilde{X}_{1:T^{(d)},\theta}))].$$

Training follows the same structure as in the random variable case. Samples from a generated neural process $\{\tilde{x}_{1:T^{(d)},\theta}^{(i)}\}_{i=1}^M$ and of the target distribution $\{x_{1:T^{(d)}}^{(i)}\}_{i=1}^M$ are passed through the discriminator, which outputs classification probabilities.

4 Stochastic Volatility Neural Networks

The generator architecture in our framework is constructed by the Stochastic Volatility Neural Network (SVNN). The SVNN consists of volatility, drift, and another network which models the innovations. This structure implies that the volatility and drift are generated based on historical latent variable processed through a shared TCN, whereas the innovation is generated via a separate network. In other words, if we consider $Z = (Z_t)_{t \in \mathbb{Z}}$ as a sequence of i.i.d. Gaussian random variables taking values in \mathbb{R}^{N_Z} . Let a TCN be $g^{(\text{TCN})} : \mathbb{R}^{N_Z \times T^{(g)}} \times \Theta^{(\text{TCN})} \rightarrow \mathbb{R}^{2N_X}$ with RFS $T^{(g)}$, and let another network be $g^{(\varepsilon)} : \mathbb{R}^{N_Z} \times \Theta^{(\varepsilon)} \rightarrow \mathbb{R}^{N_X}$. Suppose that $\alpha \in \Theta^{(\text{TCN})}$ and $\beta \in \Theta^{(\varepsilon)}$ denote parameters of the networks. We define a stochastic process R as:

$$R : \Omega \times Z \times \Theta^{(\text{TCN})} \times \Theta^{(\varepsilon)} \rightarrow \mathbb{R}^{N_X},$$

$$(\omega, t, \alpha, \beta) \mapsto [\sigma_{t,\alpha} \odot \varepsilon_{t,\beta} + \mu_{t,\alpha}](\omega).$$

where the components are defined as:

$$h_t := g_\alpha^{(\text{TCN})}(Z_{t-T^{(g)}:(t-1)}),$$

$$\sigma_{t,\alpha} := |h_{t,1:N_X}|,$$

$$\mu_{t,\alpha} := h_{t,(N_X+1):2N_X},$$

$$\varepsilon_{t,\beta} := g_\beta^{(\varepsilon)}(Z_t).$$

The process R induced by the SVNN is referred to as a log return neural process. Within this framework, $\sigma_\alpha := (\sigma_{t,\alpha})_{t \in \mathbb{Z}}$, $\mu_\alpha := (\mu_{t,\alpha})_{t \in \mathbb{Z}}$, $\varepsilon_\beta := (\varepsilon_{t,\beta})_{t \in \mathbb{Z}}$ are respectively called the volatility, drift, and innovation neural processes. These processes are inspired by the volatility-innovation decomposition methods discussed in [Cont and Tankov \(2003\)](#).

5 Evaluating Metrics and Scoring Criteria

To analyze how well the models perform, we employ a selection of performance metrics and evaluation scores.

Earth Mover Distance (EMD): The EMD serves as a measure of how different two probability distributions are. Let \mathbb{P}^h be the historical distribution and \mathbb{P}^g be the generated distribution of log returns, the EMD is defined as:

$$EMD(\mathbb{P}^h, \mathbb{P}^g) = \inf_{\pi \in \Pi(\mathbb{P}^h, \mathbb{P}^g)} \mathbb{E}_{(X,Y) \sim \pi} [|X - Y|].$$

where $\Pi(\mathbb{P}^h, \mathbb{P}^g)$ is the collection of all possible joint probability distributions and marginal distributions are \mathbb{P}^h and \mathbb{P}^g .

DY Metric: This metric captures the divergence between the logarithmic probabilities of the historical and generated data across discretized intervals. For a given time lag $t \in \mathbb{N}$, the DY metric is defined as:

$$DY(t) = \sum_x |\log P_t^h(A_{t,x}) - \log P_t^g(A_{t,x})|.$$

where P_t^h and P_t^g represent the probability density functions of the historical and generated t -differenced return trajectories for historical and generated data. The term $(A_{t,x})_x$ refers to a partitioning of the real line such that for a fixed t and all x , the approximate condition $\log P_t^h(A_{t,x}) = 5/T$ holds true, with T being the total number of historical log returns.

Autocorrelation Function (ACF) Score: This score evaluates how well the autocorrelation properties of generated time series align with those of real data. Let $r_{1:T}$ be the historical log return series, and let $\{r_{1:\tilde{T},\theta}^{(1)}, \dots, r_{1:\tilde{T},\theta}^{(M)}\}$ be the generated log return series of length $\tilde{T} \in \mathbb{N}$. The autocorrelation is defined as a function of the time lag τ and a series $r_{1:T}$.

$$\mathcal{C}(\tau; r) = \text{Corr}(r_{t+\tau}, r_t).$$

By defining $C : \mathbb{R}^T \rightarrow [-1, 1]^S : r_{1:T} \mapsto (\mathcal{C}(1; r), \dots, \mathcal{C}(S; r))$ as the autocorrelation function up to a lag $S \leq T - 1$, the ACF(f) score is calculated for a function $f : \mathbb{R} \rightarrow \mathbb{R}$ as:

$$ACF(f) := \|C(f(r_{1:T})) - \frac{1}{M} \sum_{i=1}^M C(f(r_{1:T,\theta}^{(i)}))\|_2.$$

The ACF score is computed using constants $S = 250$, $M = 500$, and $\tilde{T} = 4000$ for $f(x) = x$, $f(x) = x^2$, and $f(x) = |x|$, which are serial, squared, and absolute autocorrelation, respectively.

Leverage Effect Score: This metric is quantified by examining the correlation between lagged, squared log returns and the log returns themselves. This is expressed as:

$$\mathcal{L}(\tau; r) = \text{Corr}(r_{t+\tau}^2, r_t).$$

for a given lag τ . If $L : \mathbb{R}^T \rightarrow [-1, 1]^S : r_{1:T} \mapsto (\mathcal{L}(1; r), \dots, \mathcal{L}(S; r))$ denotes the leverage effect function up to lag $S \leq T - 1$, then the leverage effect score is defined as:

$$\|L(r_{1:T}) - \frac{1}{M} \sum_{i=1}^M L(r_{1:T,\theta}^{(i)})\|_2.$$

5.1 Novel Loss Function for the Generator

To better align the GAN’s generative process with financial objectives, a novel economics-driven loss function for the generator (L^G) is introduced by Vuletić and *et al.* (2024). This loss function augments the standard adversarial loss component with terms directly related to financial performance metrics. The general form of this loss function is:

$$L^G(x, \hat{x}) = J^{(G)}(x, \hat{x}) - \alpha \text{PnL}^*(x, \hat{x}) + \beta \text{MSE}(x, \hat{x}) - \gamma \text{SR}^*(x, \hat{x}) + \delta \text{STD}(x, \hat{x}). \quad (1)$$

where x and \hat{x} represent the true and generated values, $J^{(G)}$ is binary cross-entropy, which is the standard generator loss function. α, β, γ , and δ are non-negative hyperparameters that weigh the contribution of each financial term. The additional financial terms in (1) are defined as follows:

Profit and Loss Term (PnL*): This term aims to maximize the profitability of a simulated trading strategy based on the generator’s forecasts.

$$\text{PnL}^*(x, \hat{x}) = \frac{1}{n_{batch}} \sum_{i=1}^{n_{batch}} \text{PnL}_a^*(x_i, \hat{x}_i).$$

where $\text{PnL}_a^*(x_i, \hat{x}_i) = \tanh(k_{tanh} \hat{x}_i) x_i$. The hyperparameter k_{tanh} controls the accuracy of our approximation.

Mean Squared Error Term (MSE): This term encourages the generated values (\hat{x}_i) to be close to the actual realized values (x_i). The generator aims to minimize this term.

$$\text{MSE}(x, \hat{x}) = \frac{1}{n_{batch}} \sum_{i=1}^{n_{batch}} (x_i - \hat{x}_i)^2.$$

Sharpe Ratio Term (SR*): This term encourages the generator to produce forecasts that would lead to a high risk-adjusted return, the generator aims to maximize this term.

$$\text{SR}^*(x, \hat{x}) = \frac{\frac{1}{n_{batch}} \sum_{i=1}^{n_{batch}} \text{PnL}_a^*(x_i, \hat{x}_i)}{\sqrt{\frac{1}{n_{batch}} \sum_{i=1}^{n_{batch}} (\text{PnL}_a^*(x_i, \hat{x}_i) - \frac{1}{n_{batch}} \sum_{i=1}^{n_{batch}} \text{PnL}_a^*(x, \hat{x}))^2}}.$$

Standard Deviation of PnL Term (STD): This term, which the generator aims to minimize, penalizes high volatility in the PnL series and provide more stable performance.

$$\text{STD}(x, \hat{x}) = \sqrt{\frac{1}{n_{batch}} \sum_{i=1}^{n_{batch}} (\text{PnL}_a^*(x_i, \hat{x}_i) - \frac{1}{n_{batch}} \sum_{i=1}^{n_{batch}} \text{PnL}_a^*(x, \hat{x}))^2}.$$

Additionally, novel loss Function hyperparameters are dynamically determined through a gradient norm matching procedure, rather than being fixed parameters for external tuning. By incorporating these financial terms, the generator is not only trained to fool the discriminator but also to produce synthetic financial time series that exhibit desirable characteristics from a trading or investment perspective.

6 Methodology

In this study, we explore four distinct model structures: a pure TCN, a hybrid Fin-TCN, a constrained SVNN, and for baseline comparison, a conventional GARCH model[†]. We employed historical price data for the ticker "Vabesader" (representing Bank Saderat Iran), an actively traded stock on the Tehran Stock Exchange. The dataset consists of log return values from June 8, 2009, to May 6, 2025. We first examine the influence of architectural choices on the

[†]See the [electronic appendix](#) for further details.

performance of the models under consideration, C-SVNN, Fin-TCN, and TCN by conducting a sensitivity analysis focusing on two critical hyperparameters: kernel size (K) and dilation factor (D). K determines the receptive field size (RFS), which affects the model’s ability to capture short-term temporal patterns. Meanwhile, D controls the spacing between kernel elements, enabling the network to capture long-range dependencies. We systematically vary the kernel size $K \in \{2, 3\}$ and dilation factor $D \in \{1, 2, 3\}$ creating six distinct parameter configurations. We exclude $K = 1$ because it limits the RFS to the current timestamp, which is insufficient for modeling temporal dynamics. The statistics of interest are the annualized Sharpe Ratio, mean daily PnL, MAE, and RMSE, as reported in Table 1.

Table 1: Financial performance comparison across different K and D configurations

K	D	Model	MAE	RMSE	PnL	SR
2	1	C-SVNN	0.0340	0.0389	0.0310	0.2541
		Fin-TCN	0.0311	0.0356	0.0364	0.4729
		TCN	0.0311	0.0356	-0.0780	-0.8933
2	2	C-SVNN	0.0322	0.0373	-0.1333	-1.6934
		Fin-TCN	0.0311	0.0356	0.1058	0.8841
		TCN	0.0312	0.0357	-0.0655	-0.7065
2	3	C-SVNN	0.0312	0.0355	0.0011	0.0165
		Fin-TCN	0.0311	0.0356	0.6710	5.8046
		TCN	0.0312	0.0356	-0.2148	-1.0164
3	1	C-SVNN	0.0321	0.0377	-0.5305	-2.8949
		Fin-TCN	0.0311	0.0356	-0.0834	-0.9087
		TCN	0.0312	0.0356	-0.5329	-1.6157
3	2	C-SVNN	0.0286	0.0357	0.1190	0.4244
		Fin-TCN	0.0260	0.0311	-0.0899	-0.6446
		TCN	0.0261	0.0312	-0.1371	-0.8571
3	3	C-SVNN	0.0271	0.0325	0.1014	0.5210
		Fin-TCN	0.0248	0.0302	0.0801	0.8158
		TCN	0.0248	0.0302	-0.1390	-0.5769
		GARCH(1,1)	0.0199	0.0454	-0.3092	-0.7354

Our results show that parameter selection significantly affects model behavior across all architectures, with the Fin-TCN model demonstrating superior financial performance, achieving the highest Sharpe Ratios and PnL values in most configurations. Although the configuration $K = 2$, $D = 3$ yields exceptional performance, it results in a very large RFS that is impractical and may lead to overfitting. Therefore, we employ the more consistent configuration $K = 2$, $D = 2$, which offers robust performance while maintaining a manageable RFS to evaluate the models in the subsequent analysis. Additionally, the GARCH(1,1) model achieves the lowest MAE, while the lowest RMSE is observed in both the Fin-TCN and TCN models.

Furthermore, Table 2 presents the values of the evaluated metrics and scoring criteria for each of the models. Our numerical results highlight that the C-SVNN model can learn a neural process that matches the empirical distribution and dependence properties better than the benchmarked models. Specifically, it outperforms the GARCH(1,1) model in replicating both the empirical distribution and the temporal dependence structure of the observed returns.

Table 2: Evaluated metrics for the four models applied

Metric	C-SVNN	Fin-TCN	TCN	GARCH(1,1)
EMD(1)	0.0079	0.0136	0.0126	0.0134
EMD(5)	0.0208	0.0427	0.0399	0.0423
EMD(20)	0.0600	0.1036	0.0987	0.1031
EMD(100)	0.0879	0.2593	0.2541	0.2587
DY(1)	0.1179	12.1894	6.5783	8.4348
DY(5)	0.0719	16.4804	11.7203	12.4322
DY(20)	0.1522	16.3786	11.5897	14.0046
DY(100)	1.1441	18.3549	16.4555	18.3813
ACF(id)	0.1567	0.4906	0.3717	0.1641
ACF(abs)	0.5061	0.5379	0.5147	0.4415
ACF(sq)	0.038	0.0663	0.1118	0.3146
Leverage__Effect	0.7394	1.2436	0.8513	2.723804

Discussion and Results

This study introduces a novel loss function based on financial metrics like PnL and Sharpe Ratio, integrated with advanced sequence models such as SVNNs and TCNs, to improve the generation of synthetic financial time series. Unlike traditional GANs that often produce data lacking financial relevance despite statistical realism, this method directs the model to generate sequences that align with real-world financial performance criteria. Overall, this approach bridges the gap between statistical accuracy and financial applicability, providing a practical method for generating synthetic market data. It underscores the importance of aligning generative modeling techniques with the unique demands of the financial domain, providing a foundation for more effective applications in quantitative finance.

References

- Bai S., Koltun J. Z. and Koltun V. (2018), An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.
- Cont R. and Tankov P. (2003), Financial modelling with jump processes, Second Edition, New York, Chapman and Hall/CRC.
- Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S. and Bengio Y. (2014), Generative adversarial nets. *Advances in neural information processing systems*, **27**.
- Heaton J., Goodfellow I., Bengio Y. and Courville A. (2016), Deep learning: The mit press, 800 pp, isbn: 0262035618. *Genetic programming and evolvable machines*, **19(1)**, 305-307.
- Jolicœur-Martineau, A. (2018), The relativistic discriminator: a key element missing from standard GAN. *arXiv preprint arXiv: 1807.00734*.
- Koshiyama A., Firoozye N. and Treleaven P. (2021), Generative adversarial networks for financial trading strategies fine-tuning and combination, *Quantitative Finance*, **21(5)**, 797-813.
- Vuletić M., Prenzel F., and Cucuringu M. (2024), Fin-gan: Forecasting and classifying financial time series via generative adversarial networks. *Quantitative Finance*, **24(2)**, 175-199.
- Wiese M., Knobloch R., Korn R., and Kretschmer P. (2020), Quant GANs: deep generation of financial time series. *Quantitative Finance*, **20(9)**, 1419-1440.
- Zhou X., Pan Z., Hu G., Tang S. and Zhao C. (2018), Stock market prediction on high-frequency data using generative adversarial nets, *Mathematical Problems in Engineering*, **2018(1)**, 4907423.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Determination of Minimal Weighted Risk Sample Sizes in Sampling Inspections with Stochastic defect rate

M. Naghizadeh Qomi*

Department of Statistics, University of Mazandaran, Babolsar, Iran

Abstract

Single reliability inspection of manufactured lots is considered to determine optimal sample size using empirical data and prior information. A Beta model is assumed to describe the stochastic behavior of the proportion nonconforming or the defect rate, p . Optimal test plans are derived for the Tsallis q -exponential lifetime distributions by minimizing the classical and expected weighted-average of producer and consumer risks. Integer nonlinear programming is used to obtain optimal sample sizes. The results show that using prior information for defective rates reduces the sample size in the selected sample. A real data example is applied to illustrate the proposed plan.

Keywords: Acceptance sampling inspection plans; Beta prior model; integer nonlinear programming; weighted average risks; Tsallis q -exponential distributions.

Mathematics Subject Classification (2020): 62N05, 62N10.

1 Introduction

One of the challenges for manufacturers and customers in the market is the quality of manufactured products. The manufacturer seeks to produce a quality product that is acceptable to the customer, and the customer seeks to receive quality products. Acceptance sampling plans are one of the important tools in quality control. Therefore, there is a need to inspect production batches. Since full batch inspection is not possible due to time and cost constraints, acceptance sampling plans *ASPs* are used as one of the important quality control tools. *ASPs* have been studied by many authors, the most recent of which are Pérez-González et al. (2023), Naghizadeh Qomi and Fernández (2024), Tripathi et al. (2024) and Naghizadeh Qomi and Tripathi (2025).

Single acceptance sampling inspection plan (*SASIP*) is a conventional approach for determining the acceptability of submitted lots. In this plan, a random sample of units are selected from the submitted lot and the lot will be accepted if the number of defective units are no more than an acceptance number. This type of *ASPs* is developed by several researches. Weibull

*Corresponding author, m.naghizadeh@umz.ac.ir

distribution was analyzed in [Goode and Kao \(1961\)](#) whereas [Aslam et al. \(2010\)](#) and [Gui and Aslam \(2017\)](#) discussed exponential based distributions. Paper by [Gupta and Phyllis \(1961\)](#) was focused in the gamma distribution. For more recently researches, we can cite [Al-Husseini et al. \(2023\)](#), [Saha et al. \(2024\)](#), [Naghizadeh Qomi et al. \(2025\)](#) and [Naghizadeh Qomi and Mahdizadeh \(2025\)](#) which discussed single sampling plans for exponentiated moment exponential, logistic Rayleigh, generalized half-normal and inverted Nadarajah-Haghighi distributions, respectively.

[Al-Nasser and Obeidat \(2020\)](#) have recently conducted a time-censored *SASIP* using single point method (limiting consumer risk) when the lifetime distribution follows the Tsallis q -exponential (TQE) distribution. Our motivation is to construct a time-censored *SASIP* by minimizing the weighted-average risks in order to decide the acceptability of a submitted lot of products using empirical data and prior information.

The remainder of the paper is organized as follows. The TQE distribution is briefly introduced in Section 2. The methodology of *SASIP* is proposed in Section 3. The *SASIPs* with minimal weighted-average risks in both frequentist and Bayesian frameworks are presented in Sections 4 and 5, respectively. A real data example is provided in Section 6. Finally, a brief discussion is presented in Section 7.

2 Tsallis q -exponential distributions

The TQE model was derived by maximizing the Tsallis entropy

$$S_q(f) = \frac{1}{q-1} \left[1 - \int_{-\infty}^{\infty} \{f(x)\}^q dx \right],$$

where $q \neq 1$ is a real parameter, sometimes called entropic-index, and $f(x)$ is a probability density function (pdf). The TQE distribution converges to the ordinary exponential distribution with parameter λ when $q \rightarrow 1$. The TQE distribution is a generalization of some lifetime distributions such as the Lomax model, and it is a particular case of the generalized type II Pareto distribution; see [Al-Nasser and Obeidat \(2020\)](#).

According to Picoli et al. [Picoli et al. \(2009\)](#), the TQE model would be a good representation of both short and long-tailed distributions by choosing a suitable value of q , and therefore could be considered advantageous over many other distributions in the area of acceptance sampling. The pdf of the TQE distribution is given by

$$f(t; q, \theta) = (2 - q)\theta[1 + (q - 1)\theta t]^{1/(1-q)}, \quad t > 0, \quad (1)$$

and the corresponding cumulative distribution function (cdf) is as

$$F(t; q, \theta) = 1 - [1 + (q - 1)\theta t]^{(2-q)/(1-q)}, \quad t > 0, \quad (2)$$

where $q < 2$ ($q \neq 1$) and $\theta > 0$ are the shape and scale parameters, respectively. We denote the random variable T with density (1) by writing $T \sim TQE(q, \theta)$. The mean of the TQE distribution is given by

$$\mu = \frac{1}{\theta(3 - 2q)}, \quad q < \frac{3}{2}. \quad (3)$$

which is inverse proportional to the scale parameter θ when the parameter q is fixed. For simplicity, the mean lifetime is considered the quality characteristic of interest. However, note that in our context it is equivalent to use the mean or median lifetimes, or even a given percentile.

3 Design of time-censoring *SASIPs*

Consider that a large lot of products is submitted to judge its acceptability and the lifespan of products follows the $TQE(q, \theta)$ with pdf given in (1). Moreover, assume that p denotes the

true proportion nonconforming which would approximately be the fraction of defective objects in the lot. Also, let μ be the true mean lifetime of products and μ_0 be the specified mean life. The quality level of a product can be expressed in terms of the ratio $r = \mu/\mu_0$. A lot is said to be good if $r \geq 1$ and is considered bad if $r < 1$. The classical producer and consumer risks are defined respectively as probabilities of good lot rejection and bad lot acceptance.

The decision about lot is based on the number X_n of defective items where X_n assumed to be binomial with parameters $n \in \mathcal{N} = \{1, 2, 3, \dots\}$ and $p \in (0, 1)$. Given the fixed test time $t > 0$, an *SASIP* is characterized by $S \equiv (n, c, f)$, where $f = t/\mu_0$ is test termination. The decision process according to the inspection plan S can be stated as follows:

- *Step 1:* Select n items randomly from the submitted lot.
- *Step 2:* Accept the lot if the number X_n of defective items are no more than a value c known as acceptance number.

The operating characteristic (*OC*) function describes the stochastic behavior of the test plan S and is defined as the lot acceptance probability. Hence, the *OC* function can be expressed by

$$\begin{aligned} \mathcal{L}(p) &\equiv \mathcal{L}(p, n, c) \\ &= P(X_n \leq c|p) = \sum_{i=0}^c \binom{n}{i} p^i (1-p)^{n-i}. \end{aligned} \quad (4)$$

From (2) and (3), the proportion defective before time t is $p \equiv p(f, q, r)$ of the form

$$p = 1 - \left[1 + \frac{(q-1)f}{r(3-2q)} \right]^{(2-q)/(1-q)}.$$

4 *SASIPs* with minimal classical weighted-average risks

Let r_0 and r_1 be the ratio of quality level corresponding to the producer and consumer, respectively. Moreover, $p_0 \equiv p(d, f, r_0)$ and $p_1 \equiv p(d, f, r_1)$ are the acceptable and rejectable defective rates associated with the producer and consumer, respectively. The classical producer and consumer risks (CPR and CCR) are defined respectively as

$$CPR(n, c, p_0) = \sum_{i=c+1}^n \binom{n}{i} p_0^i (1-p_0)^{n-i},$$

and

$$CCR(n, c, p_1) = \sum_{i=0}^c \binom{n}{i} p_1^i (1-p_1)^{n-i}.$$

Following Fernández et al. (2020), consider the classical weighted-average risk (CWR) as

$$CWR(n, c, p_0, p_1) = w_0 CPR(n, c, p_0) + w_1 CCR(n, c, p_1), \quad w_0, w_1 > 0, \quad (5)$$

where w_0 and w_1 are the producer and consumer weights, respectively and $w_0 + w_1 = 1$.

In practice, the acceptance number c is fixed in advance by the decision maker (producer or consumer). Given fixed $c \geq 0$, our aim is to determine the optimal sample size the best *SASIP* that minimizes the CWR (5). Therefore, we have to solve the optimization problem

$$\min\{CWR(n, c, p_0, p_1) : (n, c) \in \Omega_1\}, \quad (6)$$

where $\Omega_1 = \{(n, c) : c \in \mathcal{N}_0 = \mathcal{N} \cup \{0\}, n \in \mathcal{N}, c \leq n\}$ is the feasible region.

Table 1 presents the sample size with minimum CWR, n_c , and their associated risks (CWR, CPR and CCR) for $r_0 = 2(1)5$, $r_1 = 1$, $f = 0.5$, $c = 0(1)3$ and $w_0 = 0.2, 0.5, 0.8$ when $q = 1.2$.

Table 1: Optimal sample size, n_c with minimal CWR, and the corresponding risks (%) for selected values of c , r_0 , and f when $q = 1.2$.

r_0	c	$(w_0, w_1) = (0.2, 0.8)$				$(w_0, w_1) = (0.5, 0.5)$				$(w_0, w_1) = (0.8, 0.2)$			
		n_c	CWR	CPR	CCR	n_c	CWR	CPR	CCR	n_c	CWR	CPR	CCR
2	0	5	27.156	79.828	4.582	2	38.212	47.289	29.136	1	35.372	27.398	53.978
	1	8	24.637	68.974	5.636	5	32.941	41.766	24.116	2	28.900	7.506	78.819
	2	11	22.488	61.643	5.707	8	29.248	38.193	20.304	5	26.356	13.040	57.426
	3	13	20.559	49.572	8.125	10	26.371	28.258	24.485	8	24.068	14.961	45.317
3	0	5	23.033	66.086	4.582	3	31.730	47.733	15.727	1	29.807	19.448	53.978
	1	8	18.356	48.037	5.636	6	24.117	33.108	15.126	4	23.299	17.238	37.440
	2	11	14.923	36.427	5.707	9	19.143	24.732	13.553	7	18.612	13.867	29.684
	3	14	12.291	28.297	5.431	12	15.530	19.097	11.964	10	15.140	11.135	24.485
4	0	5	19.947	55.800	4.582	3	27.228	38.729	15.727	1	26.739	15.065	53.978
	1	8	14.292	34.489	5.636	6	18.816	22.505	15.126	5	18.854	16.600	24.116
	2	12	10.528	26.643	3.622	10	13.509	18.150	8.868	8	13.533	10.631	20.304
	3	15	7.887	17.936	3.580	13	10.036	11.947	8.125	11	10.116	7.038	17.298
5	0	5	17.637	48.097	4.582	3	24.128	32.529	15.727	1	24.798	12.292	53.978
	1	9	11.525	30.544	3.374	7	15.103	20.903	9.303	5	15.443	11.726	24.116
	2	12	7.803	17.560	3.622	11	10.065	14.422	5.707	9	10.244	8.825	13.553
	3	16	5.360	12.427	2.331	14	6.871	8.311	5.431	13	7.024	6.552	8.125

In view of Table 1, the optimal sample size n_c grows when acceptance number c increases while the CWR and CPR decrease. For instance, if $r_0 = 2$ and $w_0 = 0.2$, we obtain $n_c = 5$, $CWR = 27.156\%$, $CPR = 79.828\%$ and $CCR = 4.582\%$ when $c = 0$, whereas $n_c = 8$, $CWR = 24.637\%$, $CPR = 68.974\%$ and $CCR = 5.636\%$ when $c = 1$. Moreover, when w_0 increases, n_c and the CPR decrease, whereas the CCR grows. For instance, if $r_0 = 2$, $c = 2$, we obtain $n_c = 11$, $CWR = 22.488\%$, $CPR = 61.643\%$ and $CCR = 5.707\%$ when $w_0 = 0.2$, whereas $n_c = 8$, $CWR = 29.248\%$, $PR = 38.193\%$ and $CCR = 26.356\%$ when $w_0 = 0.5$.

5 SASIPs with minimal expected weighted-average risks

In designing *SASIPs*, frequentist perspectives consider only empirical data and assume a constant probability p of observing a defective or nonconforming unit in the lot under inspection. however, this assumption is not reasonable and in many situations, exists prior knowledge about the stochastic behavior of the defect rate, p , that can be incorporated to the design of sampling plans. In this section, we present a method to consider a beta distribution of p in developing an optimal failure-censored *SASIPs* for TQE lifetime distribution.

Assume that the defective rate p follows a beta $Beta(a, b)$ prior distribution, where $a, b > 0$ and its pdf is given by

$$h(p) = \frac{p^{a-1}(1-p)^{b-1}}{\mathcal{B}(1, a, b)},$$

where $\mathcal{B}(x, a, b) = \int_0^x p^{a-1}(1-p)^{b-1}dp$. Following Fernández et al. (2020), the expected producer risk (EPR) is the conditional expectation of rejecting an acceptable lot, whereas the expected consumer risk (ECR) is the conditional expectation of accepting a rejectable lot. For a given sample size m and acceptance number d , the expected producer and consumer risks are defined, respectively by

$$EPR(m, d, p_0) = E[1 - \mathcal{L}(p)|p \leq p_0], \quad \text{and} \quad ECR(m, d, p_1) = E[\mathcal{L}(p)|p \geq p_1],$$

which can be expressed as

$$EPR(m, d, p_0) = 1 - \int_0^{p_0} \frac{\mathcal{L}(p)\pi(p)}{P(p \leq p_0)} dp, \quad (7)$$

and

$$ECR(m, d, p_1) = \int_{p_1}^1 \frac{\mathcal{L}(p)\pi(p)}{P(p \geq p_1)} dp, \quad (8)$$

where

$$P(p \leq p_0) = B(p_0, a, b)/B(1, a, b) \quad \text{and} \quad P(p \geq p_1) = \{B(1, a, b) - B(p_1, a, b)\}/B(1, a, b).$$

Using (4) and equations (7)-(8), the EPR and ECR are given respectively by

$$EPR(m, d, p_0) = 1 - \sum_{i=0}^d \frac{\binom{n}{i} B(p_0, a+i, b+m-i)}{B(p_0, a, b)},$$

and

$$ECR(m, d, p_1) = \sum_{i=0}^d \frac{\binom{n}{i} [B(1, a+i, b+m-i) - B(p_1, a+i, b+m-i)]}{[B(1, a, b) - B(p_1, a, b)]}.$$

Now, consider the expected weighted-average risk (EWR) as follows

$$EWR(m, d, p_0, p_1) = w_0 EPR(m, d, p_0) + w_1 ECR(m, d, p_1). \quad (9)$$

Given a fixed constant $d \geq 0$, our interest is to determine the best *SASIP* with acceptance number d , which can be obtained by minimizing (9), or equivalently by solving the optimization problem

$$\min\{EWR(m, d, p_0, p_1) : (m, d) \in \Omega_2\}, \quad (10)$$

where $\Omega_2 = \{(m, d) : d \in \mathcal{N}_0, m \in \mathcal{N}, d \leq m\}$ is the feasible region.

Table 2 shows the optimal values of m_d and the associated expected producer and consumer risks for selected values of $f = 0.5$, $r_0 = 2(1)5$, $r_1 = 1$, $d = 0(1)3$ when $q = 1.2$. Suppose that the prior distribution is $Beta(a, b)$ with mode $5(p_0 + p_1)/6$ and $a + b = 5$. Since the mode of $Beta(a, b)$ is $(a-1)/\{(a+b)-2\}$, then we get $a = 1 + 5(p_0 + p_1)/2$ and $b = 4 - 5(p_0 + p_1)/2$. In light of Table 2 with $w_0 = 0.2$ and $d = 0$, we find $m_d = 3$, $EWR = 17.810\%$, $EPR = 47.669\%$ and $ECR = 5.013\%$. It is observed that m_d tends to increase as d increases, while the EWR, EPR and ECR decrease. Moreover, m_d and EPR decrease when w_0 grows while ECR increase. It is clear that m_d tends to decrease as f increases or r_0 decrease, while EWR, EPR and ECR increase.

Figure 1 displays EWR, EPR and ECR percentages versus w_0 for $r_0 = 2$, $r_1 = 1$, $f = 0.5$, $d = 0$ and $q = 1.2$. When w_0 increases, a decrease in EPR and increase in ECR can be observed. in contrast, the EWR has a slight growth followed by a subsequent decline.

Figure 2, plots minimal CWR and EWR sample sizes versus w_0 when $q = 1.2$, $r_0 = 2$, $r_1 = 1$, $c = 0$, $d = 0$, $f = 0.5$. This figure shows that the minimal EWR sample size, m_0 , is smaller than the minimal CWR sample size, n_0 . The comparison of Tables 1 and 2 shows that, in most cases, optimal sample sizes with minimal EWR are less than the corresponding minimum-CWR sample sizes.

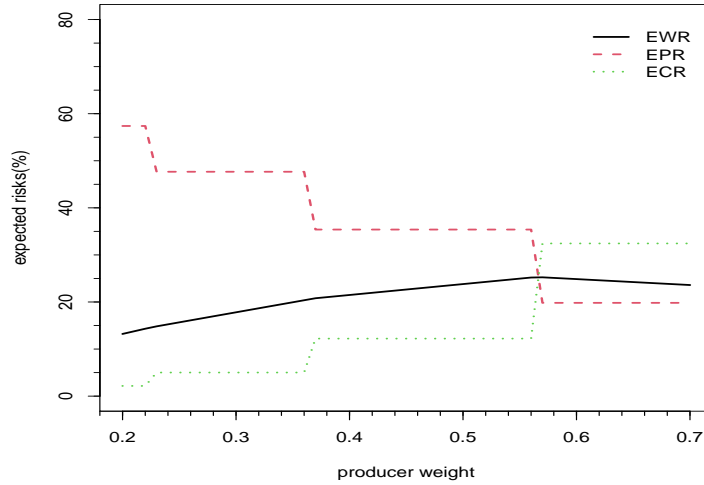
6 Application example

We consider a real data which consist of 30 breaking strengths (in megapascals, MPa) of jute fiber measured using a gauge of length 10 mm. The data are as follows (see Xia et al. (2009) and Al-Nasser and Obeidat (2020))

43.93, 50.16, 101.15, 108.94, 123.06, 141.38, 151.48, 163.40, 177.25, 183.16
 212.13, 257.44, 262.90, 291.27, 303.90, 323.83, 353.24, 376.42, 383.43, 422.11
 506.60, 530.55, 590.48, 637.66, 671.49, 693.73, 700.74, 704.66, 727.23, 778.17.

Table 2: Optimal sample size, m_d with minimal EWR, and the corresponding risks (%) for selected values of d , r_0 , and f when $q = 1.2$.

r_0	d	$(w_0, w_1) = (0.2, 0.8)$				$(w_0, w_1) = (0.5, 0.5)$				$(w_0, w_1) = (0.8, 0.2)$			
		m_d	EWR	EPR	ECR	m_d	EWR	EPR	ECR	m_d	EWR	EPR	ECR
2	0	3	17.810	47.669	5.013	2	23.806	35.388	12.224	1	23.602	19.816	32.435
	1	6	12.802	34.278	3.598	4	16.022	18.499	13.545	3	15.571	10.825	26.645
	2	8	9.604	21.388	4.553	7	11.911	15.918	7.904	5	11.602	6.531	23.436
	3	11	7.628	17.942	3.207	9	9.243	10.052	8.434	8	8.808	6.801	13.489
3	0	3	14.563	35.596	5.549	2	19.434	25.596	13.271	1	19.942	13.841	34.178
	1	6	8.926	20.286	4.057	5	11.387	15.019	7.754	4	11.490	10.031	14.894
	2	9	5.983	12.993	2.979	8	7.470	9.786	5.155	7	7.507	6.923	8.869
	3	12	4.221	8.827	2.246	11	5.196	6.727	3.665	10	5.210	4.901	5.931
4	0	4	12.507	35.679	2.577	2	16.917	20.016	13.819	1	17.956	10.625	35.062
	1	7	6.738	17.154	2.274	6	8.790	13.273	4.307	4	9.061	6.255	15.608
	2	10	4.049	9.214	1.835	9	5.132	7.083	3.181	8	5.279	5.191	5.483
	3	13	2.606	5.263	1.468	12	3.223	4.039	2.407	11	3.266	2.987	3.918
5	0	4	10.837	29.905	2.666	3	14.753	23.494	6.011	2	15.743	16.423	14.155
	1	7	5.322	12.229	2.362	6	6.894	9.326	4.463	5	7.187	6.643	8.454
	2	11	2.920	7.166	1.100	10	3.767	5.623	1.912	8	3.849	3.061	5.689
	3	14	1.716	3.554	0.929	13	2.146	2.760	1.532	12	2.208	2.079	2.509

Figure 1: EWR, EPR and ECR percentages versus w_0 for $r_0 = 2$, $r_1 = 1$, $f = 0.5$, $d = 0$ when $q = 1.2$.

The maximum likelihood (ML) estimates of the scale and shape parameters are calculated as $\hat{\lambda} = 0.0049$ and $\hat{q} = 1.3485$, respectively. A Kolmogorov-Smirnov (K-S) test is employed to check that the TQE distribution fits the aforementioned data well. The K-S statistic is computed as $D = 0.2058$ with the corresponding p -value that equals 0.1362. Thus, we conclude that the TQE distribution has good fit on considered data set.

Table 3 reports the optimal EWR plans, and the associated risks for selected values of $w_0 = 0.5, r_0 = 2(1)5, r_1 = 1$ and $f = 0.5$ when $q = 1.34$ and $p \sim \text{Beta}(a, b)$ with mode $5(p_0 + p_1)/6$ and $a + b = 5$. According to Table 3, the hyperparameter a and sample size m_d decrease and the hyperparameter b becomes larger when the quality level r_0 grows.

Using (3), the mean life can be estimated as $\hat{\mu} = [\hat{\lambda}(3 - 2\hat{q})]^{-1} = 673.54$. Consider now that the termination ratio is $f = 0.5$. Then, the termination time is $t = f\mu_0 = 336.77$. Assume that $r_0 = 2, d = 5$ and $w_0 = 0.5$. From Table 3, we get $p_0 \equiv p_0(0.5, 1.3485, 2) = 0.3670$

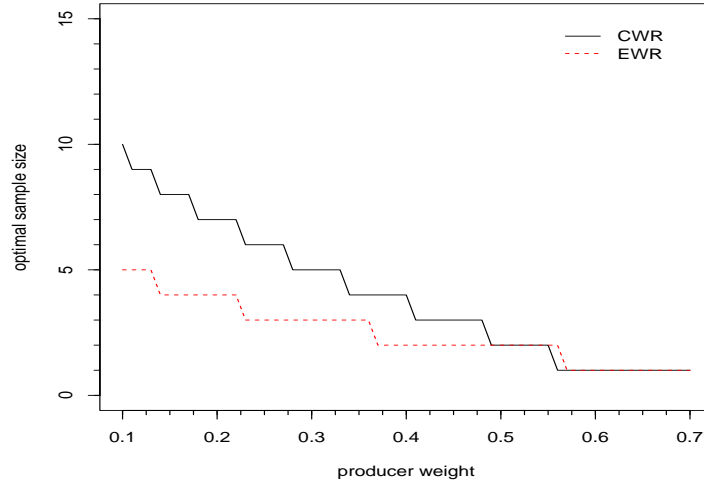


Figure 2: Optimal sample sizes with minimal CWR and EWR versus w_0 when $q = 1.2$, $r_1 = 1$, $c = 0$, $d = 0$, $f = 0.5$ and $r_0 = 2$.

Table 3: Minimum EWR sample sizes for selected values of $w_0 = 0.5$, $d = 5$ and $f = 0.5$ when $q = 1.34$ and $p \sim \text{Beta}(a, b)$ with mode $5(p_0 + p_1)/6$ and $a + b = 5$.

r_0	p_0	p_1	a	b	m_d	EWR	EPR	ECR
2	0.3670	0.5627	3.3242	1.6758	11	7.3850	7.6350	7.1350
3	0.2713	0.5627	3.0850	1.9150	13	3.6980	4.5140	2.8820
4	0.2150	0.5627	2.9442	2.0558	14	2.0480	2.2870	1.8090
5	0.1780	0.5627	2.8517	2.1483	15	1.2180	1.3540	1.0810

and $p_1 \equiv p_1(0.5, 1.3485, 1) = 0.5627$ and then $(a, b) \equiv (a_{p_0, p_1}, b_{p_0, p_1}) = (3.3242, 1.6758)$. Moreover, the optimal EWR sample size is $m_5 = 11$ with corresponding risks $EWR = 7.3850$, $EPR = 7.6350$ and $ECR = 7.1350$. According to these specifications, $m_5 = 11$ fibres selected randomly from the lot must be put on test for $t = 336.77$ MPa. The lot will be accepted if the number of failures is at most $d = 5$.

7 Concluding remarks

The probability p of observing defects in a lot inspection is usually assumed to be constant. There are many practical situations in the industry in which exists prior knowledge about the stochastic behavior of the defect rate that can be incorporated to the design of sampling plans. An approach to determine optimal WR c -defective plan (n_c, c) and optimal EWR d -defective plan, (m_d, d) under time-censoring presented in this paper. Constrained optimization problems have been formulated to determine the optimal plans. The results show that, in most cases, the minimum EWR sample sizes and expected risks are less than the corresponding minimum WR sample sizes and classical risks.

Acknowledgment

We thank the Reviewers for providing valuable comments and suggestions.

References

- Al-Husseini Z., Naghizadeh Qomi M. and MirMostafaei S.M.T.K. (2023), Single acceptance sampling plan based on truncated life tests for the exponentiated moment exponential distribution with application in bladder cancer data. *Iranian Journal of Health Sciences*, **11**, 217-228.
- Al-Nasser A. D. and Obeidat M. A. (2020), Acceptance sampling plans from truncated life test based on Tsallis q -exponential distribution. *Journal of Applied Statistics*, **47**, 685-697.
- Aslam, M., Kundu, D., & Ahmad, M. (2010), Time truncated acceptance sampling plans for generalized exponential distribution. *Journal of Applied Statistics*, **37**, 555-566.
- Fernández A. J., Correa-Álvarez C. D., and Pericchi L. R. (2020), Balancing producer and consumer risks in optimal attribute testing: A unified Bayesian/Frequentist design, *European Journal of Operational Research*, **286**, 576-587.
- Gui, W., & Aslam, M. (2017), Acceptance sampling plans based on truncated life tests for weighted exponential distribution. *Communications in Statistics-Simulation and Computation*, **46**, 2138-2151.
- Gupta, S. S., & Phyllis, A.G. (1961), Gamma distribution in acceptance sampling based on life tests. *Journal of the American Statistical Association*, **56**, 942-970.
- Goode, H. P., & Kao, J. H. K. (1961), *Sampling Procedures and Tables for Life and Reliability Testing Based on the Weibull Distribution (Mean Life Criterion)*. TR-3, DTIC Document: US Department of Defense, Washington, D.C.
- Naghizadeh Qomi M., and Fernández A. J. (2024), Optimal double acceptance sampling plans based on truncated life tests for Tsallis q -exponential distributions. *Journal of Applied Statistics*, **51**, 3456-3467.
- Naghizadeh Qomi M., and Mahdizadeh Z. (2025), Time censored reliability sampling plans for inverted Nadarajah-Haghighi distribution with minimal and limited risks. *Quality and Reliability Engineering International*, 1-11, doi:10.1002/qre.3812.
- Naghizadeh Qomi M., Piadeh M., Pérez-González C.J. and Fernández A.J. (2025), Optimal acceptance sampling plans based on generalized half-normal distribution under time-censoring with conventional and expected limited risks. *Communications in Statistics-Simulation and Computation*, 1-20. doi:10.1080/03610918.2025.2512380
- Naghizadeh Qomi M. and Tripathi H. (2025). Double acceptance sampling inspection plans based on truncated life tests for type I half-logistic Nadarajah-Haghighi distribution. *International Journal of Reliability, Quality and Safety Engineering*, doi: 10.1142/S0218539325500123.
- Pérez-González C. J., Fernández A. J., Giner-Bosch V. and Carrión-García A. (2023), Optimal repetitive reliability inspection of manufactured lots for lifetime models using prior information. *International Journal of Production Research*, **61**, 2214-2230.
- Picoli S., Mendes R. S., Malacarne L. C. and Santos R. P. B. (2009), q -Distributions in complex systems: a brief review. *Brazilian Journal of Physics*, **39**, 468-474.
- Saha M., Tripathi H., Devi A. and Pareek P. (2024), Applications of reliability test plan for logistic Rayleigh distributed quality characteristic. *Annals of Data Science*, **11**, 1687-1703.
- Tripathi H., Kiapour, A. and Naghizadeh Qomi M. (2024), Optimal time truncated double acceptance sampling plan for generalized half normal distribution. *Life Cycle Reliability and Safety Engineering* **13**, 173-180.
- Xia Z. P., Yu J. Y., Cheng L. D., Liu L. F., and Wang W. M. (2009), Study on the breaking strength of jute fibres using modified Weibull distribution. *Composites Part A: Applied Science and Manufacturing*, **40**, 54-59.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



On the stability and convergence of a finite difference scheme for stochastic partial differential equations

Mehran Namjoo^{*1}, Mehran Aminian², Hossein Negarestani³, Hossein Salmei⁴

^{1,2,3,4}Department of Mathematics, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran.

Abstract

In this paper, we introduce a stochastic finite difference scheme for the numerical solution of the Itô stochastic advection–diffusion equation with a one–dimensional white noise process. The main concepts of deterministic difference schemes namely, consistency, stability, and convergence are theoretically established for the stochastic case. Finally, a numerical example is provided to demonstrate the efficiency of the proposed stochastic difference.

Keywords: Stochastic partial differential equations, consistency, stability, convergence.

Mathematics Subject Classification (2020): 60H15, 60H35, 65M12.

1 Introduction

Many physical phenomena and engineering systems traditionally modeled by deterministic partial differential equations (PDEs) exhibit uncertainties due to various stochastic perturbation. Consequently, to achieve a more accurate and comprehensive representation of such phenomena, stochastic partial differential equations (SPDEs) are frequently employed. The investigation of SPDEs has emerged as a prominent research domain within both pure and applied mathematics, as well as in fields such as fluid mechanics, geophysics, and theoretical physics. It is noteworthy that closed-form solutions for SPDEs are typically unattainable, thus necessitating their numerical approximation through powerful computational techniques. A variety of numerical methods are extensively utilized to approximate solutions to SPDEs, including stochastic finite difference (SFD) schemes, finite element methods (FEMs) [Allen and *et al.* \(1998\)](#), and spectral methods. In [Baleanu and *et al.* \(2022\)](#); [Namjoo and Mohebbian \(2016, 2019\)](#); [Karami and *et al.* \(2024\)](#); [Namjoo and *et al.* \(2023\)](#), studied the stability and convergence properties of several SFD schemes applied to SPDEs. In the present study, we propose a stochastic finite difference scheme for the stochastic advection-diffusion equation driven by a one-dimensional Wiener process. The structure of this paper is organized as follows: Section 2 introduces an explicit stochastic finite difference for the numerical solution of SPDEs. Section 3 provides a precise investigation of the scheme's consistency, stability, and

^{*}Corresponding author, namjoo@vru.ac.ir

convergence. Numerical experiments and simulations are presented in Section 4 to empirically validate the theoretical results.

2 An explicit finite difference scheme for SPDEs

In this section, we concentrate on developing an efficient numerical method for approximating the solution of the stochastic advection–diffusion equation, as outlined in [Karami and *et al.* \(2024\)](#).

$$u_t(x, t) + \nu u_x(x, t) = \gamma u_{xx}(x, t) + \sigma u(x, t) \dot{W}(t), \quad x \in [0, 1], \quad t \in [0, 1], \quad (1)$$

with initial and boundary conditions

$$u(x, 0) = u_0(x), \quad u(0, t) = u_1(t), \quad u(1, t) = u_2(t), \quad (x, t) \in [0, 1] \times [0, 1].$$

Here, ν and γ are positive constants, and $W(t)$ denotes a one-dimensional standard Wiener process, where the white noise $\dot{W}(t)$ has a Gaussian distribution with zero mean [Kloeden and Platen \(1995\)](#). Consider a uniform grid in both space and time with stepsizes Δx and Δt , respectively. Let u_k^n denote the approximate solution at the spatial point $x_k = k\Delta x$ and time $t_n = n\Delta t$, where $\Delta x = x_{k+1} - x_k$ and $\Delta t = t_{n+1} - t_n$, for $0 \leq k \leq M-1$ and $0 \leq n \leq N-1$. To develop an SFD scheme for the SPDE (1), the partial derivatives in time and space can be approximated using methods described in [Thomas \(2013\)](#).

$$\begin{aligned} u_t(k\Delta x, n\Delta t) &\approx \frac{u_k^{n+1} - u_k^n}{\Delta t}, \\ u_x(k\Delta x, n\Delta t) &\approx \frac{u_{k+1}^n - u_{k-1}^n}{2\Delta x}, \\ u_{xx}(k\Delta x, n\Delta t) &\approx \frac{u_{k-1}^n - 2u_k^n + u_{k+1}^n}{\Delta x^2}. \end{aligned} \quad (2)$$

In fact, (2) represents a second order approximation of $u_{xx}(x_k, t_n)$, with the truncation error being order $O(\Delta x^2)$. Therefore, the scheme provides an approximation for the stochastic advection–diffusion equation as follows:

$$u_k^{n+1} = (1 - 2\gamma\Theta)u_k^n + \left(\gamma\Theta - \frac{\nu\lambda}{2}\right)u_{k+1}^n + \left(\gamma\Theta + \frac{\nu\lambda}{2}\right)u_{k-1}^n + \sigma u_k^n \Delta W_n, \quad (3)$$

where $\lambda = \frac{\Delta t}{\Delta x}$, $\Theta = \frac{\Delta t}{\Delta x^2}$. The term $\Delta W_n = W((n+1)\Delta t) - W(n\Delta t)$ denotes the increment of the Wiener process, where $\Delta W_n \sim \mathcal{N}(0, \Delta t)$.

3 Qualitative properties of the proposed scheme

In this section, we verify that the stochastic difference scheme (3) is consistent, stable, and convergent. To do this, consider a stochastic partial differential equation of the form $\mathcal{L}\mathcal{U} = \mathcal{F}$. Let $\mathcal{L}_k^n u_k^n = \mathcal{F}_k^n$ denote the discretized scheme of this equation using the proposed finite difference scheme. To analyze consistency, stability, and convergence, we define a norm for the numerical solutions. Let $\{u_k^n\}$ stand for the sequence of approximations generated by the stochastic difference scheme (3). For this sequence, we define $\|u^n\| = (\max_{0 \leq m \leq M} |u_m^n|^2)^{\frac{1}{2}}$, where $u^n = (u_0^n, u_1^n, \dots, u_M^n)$. More details regarding consistency, stability, and convergence can be found in [Roth \(2002\)](#).

Definition 3.1. *The stochastic finite difference scheme $\mathcal{L}_k^n u_k^n = \mathcal{F}_k^n$ is said to be pointwise consistent with the SPDE $\mathcal{L}\mathcal{U} = \mathcal{F}$ at the point (x, t) if for any continuously differentiable function $\Upsilon = \Upsilon(x, t)$ in the mean square sense, the following holds:*

$$\mathbb{E} \left| (\mathcal{L}\Upsilon - \mathcal{F})|_{x_k}^{t_n} - (\mathcal{L}_k^n \Upsilon(x_k, t_n) - \mathcal{F}_k^n) \right|^2 \rightarrow 0,$$

as $\Delta x \rightarrow 0$, $\Delta t \rightarrow 0$, and $(k\Delta x, (n+1)\Delta t) \rightarrow (x, t)$.

Theorem 3.2. *The stochastic difference scheme (3) is mean square consistent, as defined in 3.1.*

Proof. Let $\Upsilon(x, t)$ be a smooth function. Then, one can obtain

$$\begin{aligned} \mathcal{L}(\Upsilon(x, t))|_{x_k}^{t_n} &= \Upsilon(k\Delta x, (n+1)\Delta t) - \Upsilon(k\Delta x, n\Delta t) \\ &+ v \int_{n\Delta t}^{(n+1)\Delta t} \Upsilon_x(k\Delta x, s) ds - \gamma \int_{n\Delta t}^{(n+1)\Delta t} \Upsilon_{xx}(k\Delta x, s) ds \\ &- \sigma \int_{n\Delta t}^{(n+1)\Delta t} \Upsilon(k\Delta x, s) dW(s), \end{aligned}$$

and

$$\begin{aligned} \mathcal{L}_k^n(\Upsilon) &= \Upsilon(k\Delta x, (n+1)\Delta t) - \Upsilon(k\Delta x, n\Delta t) \\ &+ v \frac{\Delta t}{2\Delta x} [\Upsilon((k+1)\Delta x, n\Delta t) - \Upsilon((k-1)\Delta x, n\Delta t)] \\ &- \gamma \frac{\Delta t}{\Delta x^2} [\Upsilon((k+1)\Delta x, n\Delta t) - 2\Upsilon(k\Delta x, n\Delta t) + \Upsilon((k-1)\Delta x, n\Delta t)] \\ &- \sigma \Upsilon(k\Delta x, n\Delta t) \Delta W_n. \end{aligned}$$

Using the property of Itô integrals, the difference's mean square expectation satisfies:

$$\begin{aligned} \mathbb{E}|\mathcal{L}(\Upsilon)|_k^n - \mathcal{L}_k^n(\Upsilon)|^2 &= \mathbb{E} \left| \nu \int_{n\Delta t}^{(n+1)\Delta t} \left(\Upsilon_x(k\Delta x, s) - \frac{\Upsilon((k+1)\Delta x, n\Delta t) - \Upsilon((k-1)\Delta x, n\Delta t)}{2\Delta x} \right) ds \right. \\ &- \gamma \int_{n\Delta t}^{(n+1)\Delta t} \left(\Upsilon_{xx}(k\Delta x, s) - \frac{1}{(\Delta x)^2} [\Upsilon((k-1)\Delta x, n\Delta t) - 2\Upsilon(k\Delta x, n\Delta t) + \Upsilon((k+1)\Delta x, n\Delta t)] \right) ds \\ &- \sigma \int_{n\Delta t}^{(n+1)\Delta t} (\Upsilon(k\Delta x, s) - \Upsilon(k\Delta x, n\Delta t)) dW(s) \Big|^2 \\ &\leq 4\nu^2 \mathbb{E} \left| \int_{n\Delta t}^{(n+1)\Delta t} \left(\Upsilon_x(k\Delta x, s) - \frac{\Upsilon((k+1)\Delta x, n\Delta t) - \Upsilon((k-1)\Delta x, n\Delta t)}{2\Delta x} \right) ds \right|^2 \\ &+ 4\gamma^2 \mathbb{E} \left| \int_{n\Delta t}^{(n+1)\Delta t} \left(\Upsilon_{xx}(k\Delta x, s) - \frac{1}{(\Delta x)^2} [\Upsilon((k-1)\Delta x, n\Delta t) \right. \right. \\ &- 2\Upsilon(k\Delta x, n\Delta t) + \Upsilon((k+1)\Delta x, n\Delta t)] \Big) ds \Big|^2 \\ &+ 4\sigma^2 \int_{n\Delta t}^{(n+1)\Delta t} \mathbb{E} |\Upsilon(k\Delta x, s) - \Upsilon(k\Delta x, n\Delta t)|^2 ds. \end{aligned}$$

Since $\Upsilon(x, t)$ is deterministic, it implies that

$$\mathbb{E}|\mathcal{L}(\Upsilon)|_k^n - \mathcal{L}_k^n(\Upsilon)|^2 \rightarrow 0, \quad \text{as } n, k \rightarrow \infty.$$

Definition 3.3. *A stochastic difference scheme is stable in the mean square sense with respect to a norm if there exist positive constants $\overline{\Delta x}$ and $\overline{\Delta t}$, along with non-negative constants κ and α , such that*

$$\mathbb{E}\|u^{n+1}\|^2 \leq \kappa e^{\alpha t} \mathbb{E}\|u^0\|^2,$$

for all time step $t = (n+1)\Delta t$, where $0 < \Delta x \leq \overline{\Delta x}$, and $0 < \Delta t \leq \overline{\Delta t}$.

Theorem 3.4. *The stochastic difference scheme (3), under the conditions $\nu\lambda^2 \leq \gamma\Theta \leq \frac{1}{2}$, for $t = (n+1)\Delta t$, is stable with respect to the introduced norm.*

Proof. Applying the $\mathbb{E}|\cdot|^2$ for the stochastic difference scheme (3) and utilizing the independence of Wiener process increments, one obtains that

$$\begin{aligned} \mathbb{E}|u_k^{n+1}|^2 &= \mathbb{E} \left| (1 - 2\gamma\Theta)u_k^n + \left(\gamma\Theta - \frac{\nu\lambda}{2} \right) u_{k+1}^n + \left(\gamma\Theta + \frac{\nu\lambda}{2} \right) u_{k-1}^n + \sigma u_k^n \Delta W_n \right|^2 \\ &= \mathbb{E} \left| (1 - 2\gamma\Theta)u_k^n + \left(\gamma\Theta - \frac{\nu\lambda}{2} \right) u_{k+1}^n + \left(\gamma\Theta + \frac{\nu\lambda}{2} \right) u_{k-1}^n \right|^2 \\ &\quad + 2\mathbb{E} \left[\left((1 - 2\gamma\Theta)u_k^n + \left(\gamma\Theta - \frac{\nu\lambda}{2} \right) u_{k+1}^n + \left(\gamma\Theta + \frac{\nu\lambda}{2} \right) u_{k-1}^n \right) (\sigma u_k^n \Delta W_n) \right] \end{aligned}$$

$$\begin{aligned}
& + \sigma^2 \Delta t \mathbb{E} |u_k^n|^2 \\
& = \mathbb{E} \left| \left(1 - 2\gamma\Theta \right) u_k^n + \left(\gamma\Theta - \frac{\nu\lambda}{2} \right) u_{k+1}^n + \left(\gamma\Theta + \frac{\nu\lambda}{2} \right) u_{k-1}^n \right|^2 + \sigma^2 \Delta t \mathbb{E} |u_k^n|^2.
\end{aligned}$$

Using the inequalities $\nu\lambda^2 \leq \gamma\Theta \leq \frac{1}{2}$, one concludes that

$$\begin{aligned}
\mathbb{E} |u_k^{n+1}|^2 & \leq (1 - 2\gamma\Theta)^2 \mathbb{E} |u_k^n|^2 + \left(\gamma\Theta - \frac{\nu\lambda}{2} \right)^2 \mathbb{E} |u_{k+1}^n|^2 \\
& + \left(\gamma\Theta + \frac{\nu\lambda}{2} \right)^2 \mathbb{E} |u_{k-1}^n|^2 \\
& + 2(1 - 2\gamma\Theta) \left(\gamma\Theta - \frac{\nu\lambda}{2} \right) \mathbb{E} (|u_k^n| |u_{k+1}^n|) \\
& + 2(1 - 2\gamma\Theta) \left(\gamma\Theta + \frac{\nu\lambda}{2} \right) \mathbb{E} (|u_k^n| |u_{k-1}^n|) \\
& + 2 \left(\gamma\Theta - \frac{\nu\lambda}{2} \right) \left(\gamma\Theta + \frac{\nu\lambda}{2} \right) \mathbb{E} (|u_{k+1}^n| |u_{k-1}^n|) \\
& + \sigma^2 \Delta t \mathbb{E} |u_k^n|^2 \\
& \leq (1 - 2\gamma\Theta)^2 \mathbb{E} |u_k^n|^2 + \left(\gamma\Theta - \frac{\nu\lambda}{2} \right)^2 \mathbb{E} |u_{k+1}^n|^2 \\
& + \left(\gamma\Theta + \frac{\nu\lambda}{2} \right)^2 \mathbb{E} |u_{k-1}^n|^2 \\
& + (1 - 2\gamma\Theta) \left(\gamma\Theta - \frac{\nu\lambda}{2} \right) \{ \mathbb{E} |u_k^n|^2 + \mathbb{E} |u_{k+1}^n|^2 \} \\
& + (1 - 2\gamma\Theta) \left(\gamma\Theta + \frac{\nu\lambda}{2} \right) \{ \mathbb{E} |u_k^n|^2 + \mathbb{E} |u_{k-1}^n|^2 \} \\
& + \left(\gamma\Theta - \frac{\nu\lambda}{2} \right) \left(\gamma\Theta + \frac{\nu\lambda}{2} \right) \{ \mathbb{E} |u_{k+1}^n|^2 + \mathbb{E} |u_{k-1}^n|^2 \} + \sigma^2 \Delta t \mathbb{E} |u_k^n|^2 \\
& \leq (1 - 2\gamma\Theta)^2 \sup_k \mathbb{E} |u_k^n|^2 + \left(\gamma\Theta - \frac{\nu\lambda}{2} \right)^2 \sup_k \mathbb{E} |u_k^n|^2 \\
& + \left(\gamma\Theta + \frac{\nu\lambda}{2} \right)^2 \sup_k \mathbb{E} |u_k^n|^2 + 2(1 - 2\gamma\Theta) \left(\gamma\Theta - \frac{\nu\lambda}{2} \right) \sup_k \mathbb{E} |u_k^n|^2 \\
& + 2(1 - 2\gamma\Theta) \left(\gamma\Theta + \frac{\nu\lambda}{2} \right) \sup_k \mathbb{E} |u_k^n|^2 + 2 \left(\gamma\Theta - \frac{\nu\lambda}{2} \right) \left(\gamma\Theta + \frac{\nu\lambda}{2} \right) \sup_k \mathbb{E} |u_k^n|^2 \\
& + \sigma^2 \Delta t \sup_k \mathbb{E} |u_k^n|^2 \\
& = \left((1 - 2\gamma\Theta) + \left(\gamma\Theta - \frac{\nu\lambda}{2} \right) + \left(\gamma\Theta + \frac{\nu\lambda}{2} \right) \right)^2 + \sigma^2 \Delta t \sup_k \mathbb{E} |u_k^n|^2 \\
& = (1 + \sigma^2 \Delta t) \sup_k \mathbb{E} |u_k^n|^2.
\end{aligned}$$

Since this holds for all k , it follows that

$$\sup_k \mathbb{E} |u_k^{n+1}|^2 \leq (1 + \sigma^2 \Delta t) \sup_k \mathbb{E} |u_k^n|^2.$$

Note that $\Delta t = \frac{t}{(n+1)}$, which implies that

$$\begin{aligned}
\mathbb{E} \|u^{n+1}\|^2 & \leq (1 + \sigma^2 \Delta t) \mathbb{E} \|u^n\|^2 \\
& \vdots \\
& \leq (1 + \sigma^2 \Delta t)^{n+1} \mathbb{E} \|u^0\|^2 \\
& \leq e^{\sigma^2 (n+1) \Delta t} \mathbb{E} \|u^0\|^2 \\
& = e^{\sigma^2 t} \mathbb{E} \|u^0\|^2.
\end{aligned}$$

By Setting $\kappa = 1$ and $\alpha = \sigma^2$, we deduce that the stochastic difference scheme (3) is stable. \square

Definition 3.5. A stochastic difference scheme $\mathcal{L}_k^n u_k^n = \mathcal{F}_k^n$ that approximates the SPDE $\mathcal{LU} = \mathcal{F}$ is said to be mean square convergent at time $t = (n+1)\Delta t$, if $\mathbb{E} |u^{n+1} - \mathcal{U}^{n+1}|^2 \rightarrow 0$, as $\Delta x \rightarrow 0$, $\Delta t \rightarrow 0$.

Theorem 3.6. *The stochastic difference scheme (3) for $\frac{\nu\lambda}{2} \leq \gamma\Theta \leq \frac{1}{2}$ is convergent in mean square sense with respect to the defined norm.*

Proof. According to the stochastic Lax-Richtmyer theorem, the stochastic difference scheme (3) is convergent in the mean square sense with respect to the specified norm. \square

4 Numerical implementations and results

In this section, we present the numerical results of the stochastic difference scheme (3). All numerical results were obtained using MATLAB 2024.

Example 1. We analyze the performance of the proposed stochastic scheme for the SPDE:

$$u_t(x, t) = 0.01u_{xx}(x, t) + 2u(x, t)\dot{W}(t), \quad (x, t) \in [0, 1] \times [0, 1], \quad (4)$$

with the initial and boundary conditions

$$\begin{aligned} u(x, 0) &= \exp(-100(x - 0.2)^2), & x &\in [0, 1], \\ u(0, t) &= \frac{1}{\sqrt{4t+1}} \exp\left(-\frac{4}{4t+1}\right), \\ u(1, t) &= \frac{1}{\sqrt{4t+1}} \exp\left(-\frac{64}{4t+1}\right), & t &\in [0, 1]. \end{aligned}$$

In the absence of the stochastic term, the exact solution to problem (4) is given by

$$u(x, t) = \frac{1}{\sqrt{4t+1}} \exp\left(\frac{-100(x - 0.2)^2}{4t+1}\right).$$

The stochastic difference scheme is given by

$$u_k^{n+1} = u_k^n + \frac{\Theta}{100}(u_{k-1}^n - 2u_k^n + u_{k+1}^n) + \sigma u_k^n (W((n+1)\Delta t) - W(n\Delta t)), \quad (5)$$

where $\Delta x = \frac{1}{M}$, and $\Delta t = \frac{1}{N}$ for some positive integers M and N . This scheme is conditionally stable under the constraints $\frac{\nu\lambda}{2} \leq \gamma\Theta \leq \frac{1}{2}$. Consequently, if $M = 100$, the stability (or convergence) condition requires $\Delta t \leq 0.005$ or equivalently $N \geq 200$. This is illustrated in Figure 1. It is important to note that the scheme becomes unstable when $N = 100$. According to Table 1, the results of $\mathbb{E}(u(0.2, 1))$ are destroyed when $N < 200$, whereas stable results are achieved for $N \geq 200$.

N	$\mathbb{E}(u(0.2, 1))$
100	$5.4762E + 39$
110	$2.3617E + 38$
120	$3.0925E + 36$
150	$-4.7346E + 25$
200	0.3414
205	0.3447
210	0.3217
300	0.3271

Table 1: Solve the SPDE (4) using the proposed stochastic difference scheme.

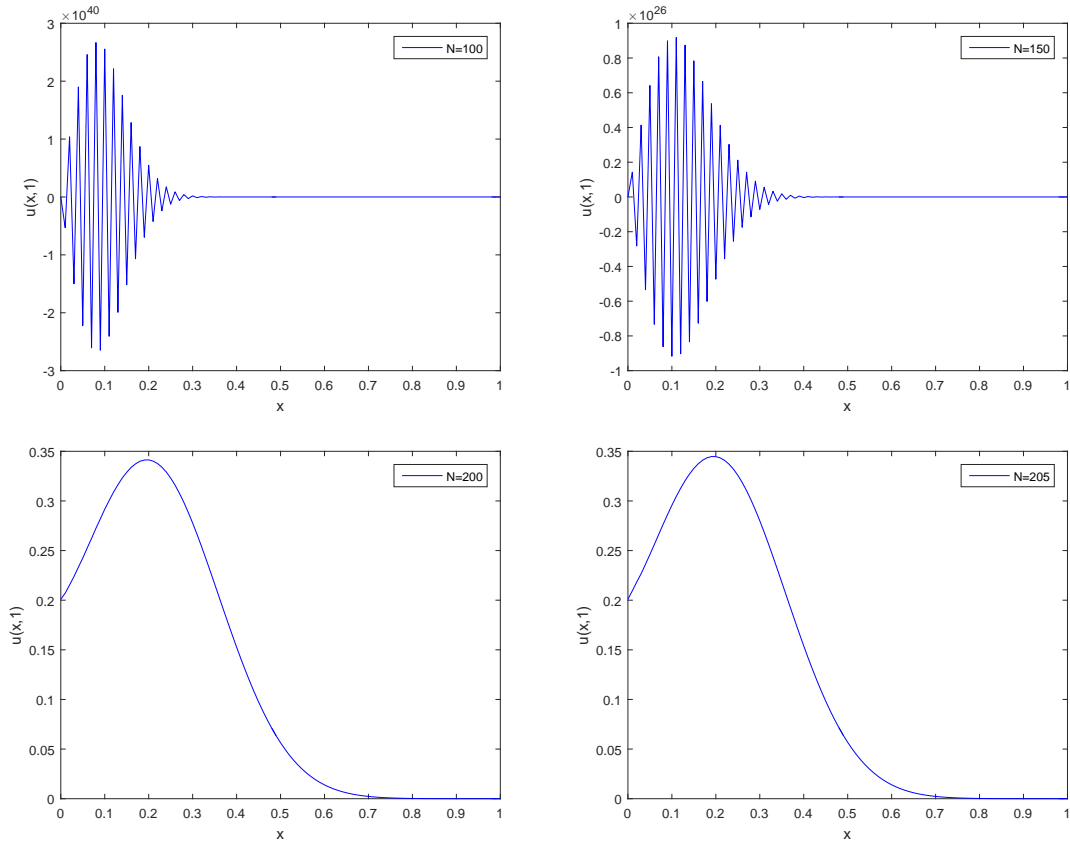


Figure 1: Analysis of the stability of the stochastic difference scheme (5) for different values of N .

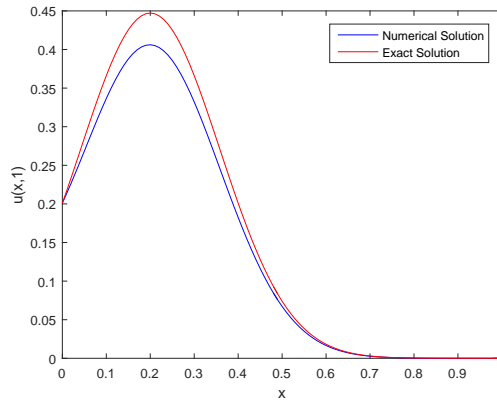


Figure 2: Comparison between the deterministic and stochastic numerical solutions of SPDE (4) using the stochastic difference scheme (5).

5 Conclusion and future work

In this paper, we have presented an explicit stochastic finite difference scheme for a class of SPDEs. We have also analyzed the scheme's consistency, stability, and convergence. The results from our numerical simulations indicate that the proposed scheme is effective for a

variety of SPDEs. Our future research will focus on developing nonstandard stochastic finite difference schemes for these equations.

References

- Baleanu, D., Namjoo, M., Mohebbian, A., and Jajarmi, A. (2022). A Weighted Average Finite Difference Scheme for the Numerical Solution of Stochastic Parabolic Partial Differential Equations. *CMES-Computer Modeling in Engineering and Sciences*, 135(2), 1147-1163.
- Namjoo, M., and Mohebbian, A. (2016). Approximation of stochastic advection diffusion equations with finite difference scheme. *Journal of Mathematical Modeling*, 4(1), 1–18.
- Kloeden, P. F., and Platen, E. (1995). *Numerical solution of stochastic differential equations*. Springer.
- Namjoo, M., and Mohebbian, A. (2019). Analysis of the stability and convergence of a finite difference approximation for stochastic partial differential equations. *Computational Methods in Differential Equations*, 7(3), 334–358.
- Roth, C. (2002). Difference methods for stochastic partial differential equations. *Zeitschrift für Angewandte Mathematik und Mechanik*, 82, 821–830.
- Karami, M., Mohebbian, A., Razaghian, S., Namjoo, M., and Aminian, M. (2024). Numerical solutions for a class of stochastic partial differential equations. *Journal of Mahani Mathematical Research Center*, 13(1).
- Namjoo, M., Aminian, M., Mohebbian, A., Karami, M., and Salmei, H. (2023). Upwind implicit scheme for the numerical solution of stochastic advection-diffusion partial differential equations. *Analytical and Numerical Solutions for Nonlinear Equations*, 8(2), 138–162.
- Thomas, J. W. (2013). *Numerical partial differential equations: Finite difference methods*. Springer. Science & Business Media.
- Allen, E. J., Novosel, S. J., and Zhang, Z. (1998). Finite element and difference approximation of some linear stochastic partial differential equations. *Stochastics: An International Journal of Probability and Stochastic Processes*, 64(1–2), 117–142.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Unsupervised and Supervised Time Series Grouping in Data Science

A. R. Nematollahi*

Department of Statistics, Shiraz University

Abstract

This study provides a structured concise review of grouping (clustering/classification) methods adapted for time series data in data science applications. For clustering (unsupervised learning), we discuss three major paradigms: Feature-based methods, which employ dimensionality reduction to transform raw time series into lower-dimensional representations; Shape-based techniques, such as Dynamic Time Warping, that measure similarity by aligning time series of variable lengths and are widely used in algorithms like k-means, k-medoids and hierarchical clustering; and Model-based approaches, including Gaussian Mixture Models, which assume probabilistic generative processes for the observed data. For classification (supervised learning), we examine not only traditional pipelines that combine feature extraction with machine learning classifiers but also probabilistic model-based methods, such as Gaussian Mixture Models, adapted for time series classification. Current research trends in both fields are also discussed.

Keywords: Time series analysis, clustering, classification, Dynamic Time Warping, Gaussian Mixture Models, Silhouette score.

Mathematics Subject Classification (2020): 62M10, 91B84, 62H30, 91C20.

1 Introduction

Unlike conventional datasets, Time series data present distinct analytical challenges, including temporal dependencies, high dimensionality, and variable sequence lengths, requiring specialized analytical approaches. In this study, we present a brief review of time series grouping methods, covering both unsupervised (clustering) and supervised (classification) learning paradigms.

In practical applications, grouping tasks can be systematically classified along two key axes: temporal characteristics (static/time-invariant versus dynamic/time-varying data) and feature dimensionality (univariate versus multivariate observations).

These fundamental axes define four distinct analytical scenarios in data grouping tasks:

*Corresponding author, ar.nematollahi@shirazu.ac.ir

1. **Static Univariate Data (Single-feature, time-independent observations):** Each observation is a scalar value (e.g., grouping 60 individuals by height sampled from two normal distributions with means 160 cm and 175 cm).
2. **Static Multivariate Data (Multi-feature, time-independent observations):** Each observation is a feature vector, requiring feature scaling (e.g., grouping insurance customers by age and annual income to design tailored insurance plans).
3. **Dynamic Univariate Series (Single-feature time series):** Serve two primary purposes: grouping independent univariate sequences (e.g., clustering 24-hour electricity consumption patterns across households), and segmentation (pattern identification) of individual single-feature series (e.g., detecting warm, cold, or transitional weather periods from 30-day temperature data.)
4. **Dynamic Multivariate Series (Multi-feature time series):** Multivariate sequences with cross-feature dependencies (e.g., stock prices and trading volumes for multiple companies over 50 days).

Univariate time series segmentation, such as that applied to 30-day temperature data, is typically achieved using three key methodologies. These include change-point detection algorithms (e.g., CUSUM) for identifying structural breaks, feature-based clustering (e.g., k-means with DTW) for grouping similar subsequences, and probabilistic models (e.g., HMMs) that define phases as latent states. Although deep learning methods (LSTMs, Temporal CNNs) have shown promising results, they are beyond the scope of this analysis. Interested readers are referred to [Keogh and *et al.* \(2004\)](#) for more details.

This study specifically addresses the grouping of independent dynamic univariate series (single-feature time series), which present unique analytical challenges due to their temporal dependencies and sequential nature. The paper is structured as follows: We begin by introducing time series concepts. Subsequently, our analysis is organized into two components: (1) Unsupervised grouping methods, with a primary focus on clustering techniques; and (2) Supervised classification approaches for labeled time series data. For a comprehensive discussion, we refer readers to [Caiado and *et al.* \(2021\)](#) and related literature.

2 Stochastic Process and Time Series

A stochastic process is defined as a collection of random variables that are ordered in time and defined as a set of points that may be discrete or continuous. A continuous stochastic process is described as $\{X(t), -\infty < t < \infty\}$ while a discrete stochastic process is described as $\{X_t, t = \dots, -2, -1, 0, 1, 2, \dots\}$. Classical statistics typically focuses on estimating population characteristics from independent and identically distributed (i.i.d.) samples, whereas time series analysis deals with sequential observations that are ordered in time and often exhibit temporal dependence. A time series $\{x_t, t = 0, 1, 2, \dots, T\}$ (e.g., daily stock prices or temperatures) represents one realization of a stochastic process, which is formally defined as a collection of random variables indexed by time, denoted by $\{X_t, t = 0, 1, 2, \dots, T\}$ for discrete time. We usually check these models by looking at their first and second moments. It is crucial that the second moments are finite ($E|X_t|^2 < \infty$). The mean function $\mu(t)$ and the autocovariance function $\gamma(t_1, t_2)$ are defined as:

$$\mu(t) = E[X_t],$$

$$\gamma(t_1, t_2) = \text{Cov}(X_{t_1}, X_{t_2}) = E[(X_{t_1} - \mu(t_1))(X_{t_2} - \mu(t_2))].$$

Under stationarity, these simplify to:

$$\mu(t) = \mu, \quad \text{for all } t,$$

$$\gamma(t_1, t_2) = \gamma(\tau), \quad \text{where } \tau = |t_1 - t_2| \quad (\text{covariance depends only on time lag}).$$

In the upcoming sections, we will explore unsupervised (clustering) and supervised (classification) methods. These approaches help data scientists handle time-series data in applications.

3 Unsupervised Time Series Grouping via Clustering

Unsupervised time series clustering groups time series data points into clusters. The goal is to make series in the same cluster more like each other than those in different clusters. This occurs without any pre-assigned labels or categories. We begin with classical methods and then proceed to explore three modern approaches to clustering: feature-based, shape-based, and model-based. These methods improve on traditional techniques by enhancing feature extraction, refining temporal similarity measures, and incorporating probabilistic modeling frameworks, respectively.

3.1 Classical Clustering Algorithms

Traditional time series clustering employs classical algorithms valued for their simplicity and reliability. The main approaches are:

1. **k-means Clustering:** k-means is a well-known unsupervised algorithm. It divides data into k clusters. These clusters have compact shapes and can be split by a straight line using Euclidean distance. It uses a centroid-based method. Each cluster displays its mean, called the centroid.
2. **k-medoids Clustering:** k-medoids is a partitioning algorithm like k-means. But it uses actual data points called *medoids* as cluster centers, not averages. This makes it stronger against outliers and noise. A medoid is less influenced by extreme values than a centroid.
3. **Hierarchical Clustering:** Hierarchical clustering methods construct cluster hierarchies using either agglomerative (bottom-up) or divisive (top-down) approaches. The decision to merge or split clusters relies on a linkage criterion that defines the dissimilarity between two clusters. Common linkage types include: Single linkage, Complete linkage, Average linkage, and Ward's linkage. The result of hierarchical clustering is typically visualized as a dendrogram, a tree-like diagram where the height of each node represents the dissimilarity (linkage value) at which clusters are merged.
4. **DBSCAN Clustering:** DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density-based clustering algorithm that discovers clusters by identifying dense regions in the feature space. Unlike k -means, it does not require the number of clusters to be specified in advance and can detect clusters of arbitrary shape while identifying outliers as noise.

While traditional clustering algorithms (e.g., K-means) retain theoretical value for static data, their i.i.d. assumptions render them inadequate for temporal sequences. They inherently ignore critical temporal relationships—trends, phase shifts, and rate variations—failing to recognize similarity between temporally misaligned or scaled series with identical shapes. Modern research addresses these limitations through three key paradigms: (1) feature-based frameworks (transforming temporal patterns into vector spaces), (2) shape-based approaches (using elastic measures like Dynamic Time Warping), and (3) model-based techniques (probabilistic hierarchical structures). We systematically analyze each paradigm's strengths in subsequent sections.

3.2 Methods based on Feature Extraction or Feature-based Clustering Approach

Time series data's inherent temporal dependencies make standard clustering algorithms inadequate. A common solution is feature-based clustering, which extracts meaningful patterns before grouping. This approach converts time series into fixed-length feature vectors, enabling compatibility with standard clustering algorithms like k-means, k-medoids, hierarchical clustering, and DBSCAN (see Section 3.1). Feature selection is a crucial step that heavily depends on the specific domain and the patterns one aims to identify. These features can be categorized as follows:

- **Statistical Features:** These include basic measures such as mean, variance/standard deviation, minimum/maximum value, skewness, kurtosis, and percentiles.
- **Time-Domain Specific Features:** This category covers characteristics specific to the time-based nature of the data, such as slope/trend, seasonality measures, autocorrelation coefficients, number of peaks/troughs, and energy.
- **Frequency-Domain Features:** These features are extracted by analyzing the time series in the frequency domain. Examples include Fourier coefficients, spectral entropy, and wavelet coefficients.
- **Model-Based Features:** These are parameters derived from a fitted time series model, such as ARIMA coefficients or Hidden Markov Model parameters.

By transforming a time series into a feature vector, the challenge shifts from comparing complex sequences to evaluating simpler vectors within a multidimensional feature space.

3.3 Shape-based (Observation-based) Approach

These methods directly compare the shapes of time series, without first extracting features. The core challenge here is defining an appropriate distance or similarity metric that accounts for time series specific challenges like varying lengths and temporal misalignments.

3.3.1 Dynamic Time Warping (DTW): A Robust Similarity Measure for Time Series

Dynamic Time Warping (DTW) is a powerful similarity measure for time series that addresses the limitations of rigid metrics like Euclidean distance. Formally, DTW measures similarity between two time series $X = (x_1, \dots, x_n)$ and $Y = (y_1, \dots, y_m)$ by finding the optimal alignment that minimizes the cumulative distance, allowing for phase shifts and length variations. Unlike Euclidean distance's strict point-to-point matching, DTW computes a minimum-cost warping path through a cumulative distance matrix, enabling flexible one-to-many and many-to-one alignments (Figure 1). This makes it particularly valuable for clustering time series with temporal variations when used with distance-based algorithms like k-medoids, which selects actual data points as interpretable cluster centers in the warped alignment space.

The method's ability to handle non-linear temporal distortions while preserving meaningful shape similarities explains its widespread use in applications ranging from speech recognition to sensor data analysis. By accommodating timing variations through optimal alignment, DTW provides a more robust foundation for time series comparison and pattern recognition tasks than conventional rigid distance measures.

The DTW algorithm computes similarity between two time series by constructing a cost matrix (or distance matrix) that captures pairwise distances between all points. The algorithm then finds the optimal alignment path through this matrix that minimizes the total accumulated cost. This process involves three key steps:

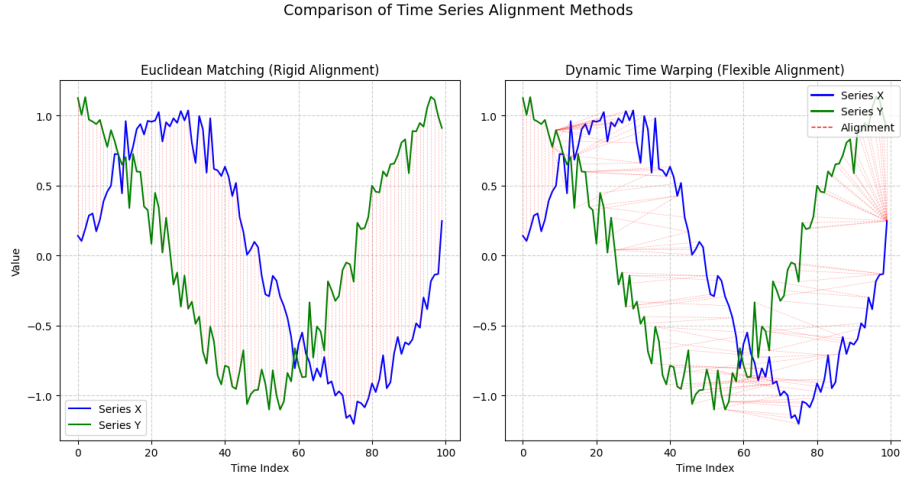


Figure 1: Comparison of alignment methods: (a) Rigid Euclidean matching enforces one-to-one point correspondence, while (b) Dynamic Time Warping (DTW) allows non-linear temporal alignment to accommodate phase shifts and speed variations between time series.

1. **Distance Matrix** ($D(i, j)$): For each cell (i, j) in the $n \times m$ grid, the local distance $D(i, j)$ can be computed using either the Manhattan (*absolute difference*) or Euclidean (*squared difference*) metric.

Both serve as the foundation for cumulative alignment cost in DTW, with the choice depending on data characteristics. For instance, the Manhattan metric demonstrates greater robustness for noisy sensor measurements, whereas the Euclidean norm is preferred for true geometric distances (e.g., map coordinates).

2. **Accumulated Cost Matrix** ($C(i, j)$): The accumulated cost matrix $C(i, j)$ represents the minimum cumulative distance from the starting point $(1, 1)$ to the current point (i, j) following a valid warping path. It is calculated recursively. The term $D(i, j)$ is the cost of aligning x_i and y_j . The minimum of the three preceding cells ensures that the path is optimal, reflecting choices to move vertically (skip a point in Y), horizontally (skip a point in X), or diagonally (align points simultaneously). DTW is a typical dynamic programming problem that can be solved by a more straightforward recursive formula:

$$C(i, j) = D(i, j) + \min\{C(i-1, j), C(i, j-1), C(i-1, j-1)\},$$

which means that the warping distance at cell (i, j) depends not only on the distance between x_i and y_j , but also on the minimum value of the adjacent cells at $(i-1, j-1)$, $(i-1, j)$, and $(i, j-1)$.

3. **Warping Path Constraints:** A warping path $P = (p_1, \dots, p_k)$ where $p_k = (i_k, j_k)$ is a sequence of adjacent cells in the grid that defines the alignment between the two time series. To ensure a meaningful alignment, certain constraints are applied: The path must start at $(1, 1)$ and end at (n, m) . This ensures that all points in both time series are considered in the alignment. $i_{k+1} \geq i_k$ and $j_{k+1} \geq j_k$. This means the path can only move forward in time (or stay at the same time index), preventing it from going backward and creating non-causal alignments. Each step in the path must be to an adjacent cell (vertical, horizontal, or diagonal). This ensures that no points are skipped in either time series, maintaining a continuous mapping.

Example: Figure 2 provides a comparative visualization of Euclidean and Dynamic Time Warping (DTW) alignments for two time series of different lengths. The upper row illustrates

the original signals with their respective alignment connections: rigid point-to-point matching for Euclidean distance and flexible matching for DTW. The middle row displays the corresponding Euclidean distance matrix and DTW accumulated cost matrix, with the optimal warping path (blue arrows) linking geometrically similar points while minimizing the cumulative distance. In this heatmap, lighter colors indicate lower-cost regions, and the blue arrows represent the optimal warping path, which traverses these lighter areas to minimize the total path cost. The lower row shows the time series after DTW warping (aligned in time). This visualization clearly demonstrates DTW’s capacity to accommodate phase shifts and local distortions, in contrast to the strict, linear alignment imposed by the Euclidean approach.

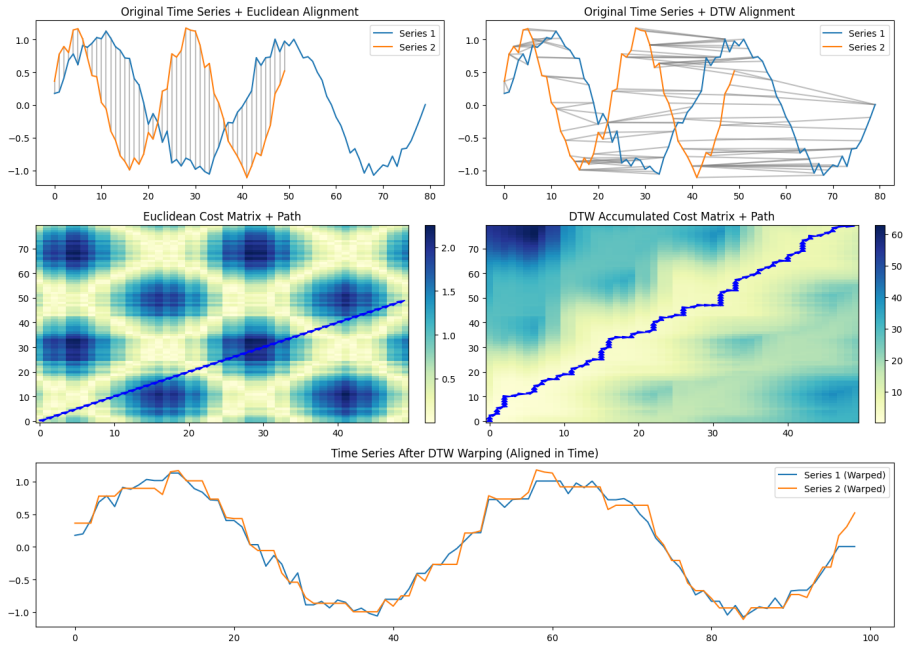


Figure 2: A comparison of Euclidean and Dynamic Time Warping (DTW) alignments for two time series. The top row shows the alignments, the middle row displays the corresponding distance and accumulated cost matrices with their optimal warping paths, and the bottom row illustrates the time-aligned series after DTW warping.

In the Euclidean method, the total cost is the sum of point-to-point distances between two time series. For a fair comparison, this cost is normalized by dividing it by the series length. In contrast, the DTW method constructs a cumulative cost matrix, and its total cost is the minimal accumulated alignment cost, found at cell $C(m, n)$. Since the warping path length is typically longer than the series length, the DTW cost is normalized by the path length. As shown in the following table, the normalized DTW cost is predictably lower than the normalized Euclidean cost.

Table 1: Comparison of Euclidean and DTW alignment costs

Method	Total Cost	Normalized Cost	Path Length
Euclidean	8.002	0.160	50
DTW	7.276	0.087	84

DTW’s ability to handle temporal distortions makes it highly suitable for clustering time series data, where variability in timing is common. Before applying clustering algorithms, a similarity (or dissimilarity) matrix is constructed. This matrix contains the pairwise DTW

distances between every pair of time series in the dataset. This step is computationally intensive, especially for large datasets, as DTW has a quadratic time complexity in the length of the time series ($O(nm)$). With the similarity matrix established, standard clustering algorithms can be employed, including: DTW- k -means, DTW- k -medoids, or hierarchical.

Example: Figure 3 demonstrates the clustering of 30 randomly generated time series (each containing 50 time points) using three methods:

- k -means with DTW barycenter averaging (DBA) to compute three centroids,
- k -medoids with three randomly initialized representative time series (medoids) using DTW distance,
- Hierarchical clustering with average linkage based on a DTW distance matrix.

All methods partitioned the data into three distinct clusters, color-coded as blue, orange, and green. In k -means and k -medoids plots, individual time series are displayed in semi-transparent colors, centroids (k-means) are represented by solid lines, while medoids (k-medoids) are highlighted with dash-dot lines in their respective cluster colors. The hierarchical clustering dendrogram uses the same color scheme for its branches, with vertical heights indicating DTW-based dissimilarity.

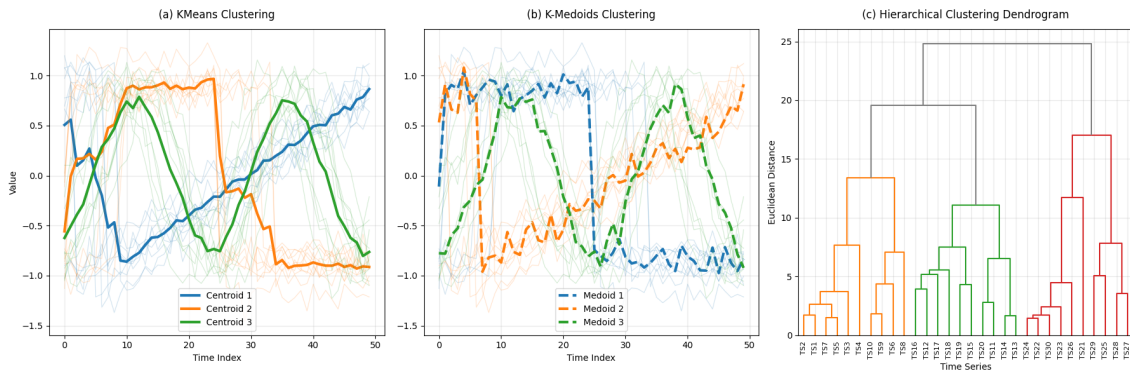


Figure 3: Comparison of clustering methods: (a) k -means with DBA (solid lines = centroids), (b) k -medoids with DTW (solid lines = medoids), and (c) hierarchical clustering dendrogram. Colors (blue/orange/green) denote cluster assignments.

3.3.2 Alternative Shape-Based Clustering Methods for Time Series

Beyond Dynamic Time Warping (DTW), specialized algorithms have been developed to address its limitations such as computational complexity and sensitivity to noise. Alternative approaches to DTW include: (1) edit-based methods (LCSS [Vlachos and *et al.* \(2002\)](#), EDR/ERP [Chen and Ng \(2005\)](#)) that perform sequence alignment via edit operations, and (2) feature-based approaches (Shapelets [Ye and Keogh \(2009\)](#), CID [Batista and *et al.* \(2014\)](#)) that cluster time series using discriminative shape features.

In summary, shape-based time series clustering is particularly effective in applications where patterns and waveforms are more critical than absolute values, such as ECG/EEG analysis for detecting characteristic heart rhythms or brainwave patterns, industrial sensor monitoring for identifying equipment signatures and vibration anomalies, and motion recognition for classifying gesture patterns from accelerometer data. By focusing on the inherent shapes of time series, this approach provides an efficient middle ground between raw data complexity and feature-based methods, though its success heavily depends on domain expertise to select and extract the most discriminative shape features that capture the essential characteristics of the data while maintaining robustness to noise and temporal variations. The

interpretability of shape features makes this method especially valuable in domains where understanding the underlying patterns is as important as the clustering results themselves.

3.4 Model-based Clustering Approach

Model-based clustering assumes that the data points (or the time series themselves) are generated by a mixture of underlying probability distributions. The goal is to estimate the parameters of these distributions and assign each time series to the cluster corresponding to the distribution it most likely belongs to.

3.4.1 GMM approach for Clustering (Unsupervised Learning)

Gaussian Mixture Models (GMMs) offer a probabilistic approach to clustering, assuming that the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. Each Gaussian distribution represents a different cluster. For time series data, directly applying GMMs can be challenging due to the sequential nature and varying lengths. Therefore, a common approach is to first extract meaningful features from each time series, transforming them into fixed-length feature vectors. These feature vectors then become the input for the GMM. Indeed, GMMs are generally applied to the features extracted from time series rather than directly to raw time series of varying lengths. Each component in the GMM represents a cluster, and data points (feature vectors) are assigned to clusters based on their probability of belonging to each Gaussian component.

A Gaussian Mixture Model represents the probability density function of a random variable \mathbf{x} as a weighted sum of K Gaussian component densities:

$$p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k),$$

where K is the number of Gaussian components (clusters), π_k is the prior probability (or mixing coefficient) of component k , satisfying $\sum_{k=1}^K \pi_k = 1$ and $\pi_k \geq 0$, $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the multivariate Gaussian probability density function for component k , defined as:

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}_k|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)\right),$$

where D is the dimensionality of \mathbf{x} , $\boldsymbol{\mu}_k$ is the mean vector of component k , and $\boldsymbol{\Sigma}_k$ is the covariance matrix of component k , and $\boldsymbol{\theta} = \{(\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\}_{k=1}^K$ represents the complete set of model parameters.

The parameters of a GMM are typically learned using the Expectation-Maximization (EM) algorithm. EM is an iterative optimization algorithm for finding maximum likelihood or maximum a posteriori (MAP) estimates of parameters in statistical models, where the model depends on unobserved latent variables. In GMM, the latent variable indicates which Gaussian component generated each data point.

The EM algorithm consists of two steps:

1. **E-step (Expectation):** Calculate the posterior probability (responsibilities) that each data point \mathbf{x}_n belongs to each component k , given the current parameter estimates.

$$\gamma(z_{nk}) = p(z_{nk} = 1|\mathbf{x}_n, \boldsymbol{\theta}^{\text{old}}) = \frac{\pi_k^{\text{old}} \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k^{\text{old}}, \boldsymbol{\Sigma}_k^{\text{old}})}{\sum_{j=1}^K \pi_j^{\text{old}} \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j^{\text{old}}, \boldsymbol{\Sigma}_j^{\text{old}})}.$$

Here, z_{nk} is a binary indicator variable such that $z_{nk} = 1$ if \mathbf{x}_n was generated by component k , and 0 otherwise. $\gamma(z_{nk})$ is the responsibility of component k for data point \mathbf{x}_n .

2. **M-step (Maximization):** Update the model parameters $(\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ to maximize the expected log-likelihood, using the responsibilities computed in the E-step.

$$N_k = \sum_{n=1}^N \gamma(z_{nk}), \quad \boldsymbol{\mu}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) \mathbf{x}_n,$$

$$\boldsymbol{\Sigma}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}})(\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}})^T, \quad \pi_k^{\text{new}} = \frac{N_k}{N},$$

where N is the total number of data points.

The E-step and M-step are iterated until convergence (e.g., when the change in log-likelihood is below a certain threshold).

To apply GMM for clustering time series, let TS_1, TS_2, \dots, TS_N be N time series. After feature extraction, we get N feature vectors, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$, where each $\mathbf{x}_i \in \mathbb{R}^D$. Then, we can apply the EM algorithm to the extracted feature vectors \mathbf{x}_n to learn the parameters $(\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ of the GMM. The number of clusters K needs to be predetermined or selected using criteria like AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion). After the GMM is trained, each time series (represented by its feature vector \mathbf{x}_n) is assigned to the cluster k for which it has the highest posterior probability (responsibility): $\text{cluster}(\mathbf{x}_n) = \arg \max_k (\gamma(z_{nk}))$.

Example: Figure 4 illustrates the application of GMM to time series clustering using extracted statistical features. In this approach, two-dimensional feature vectors (computed from each time series as the mean and standard deviation) serve as inputs to the GMM algorithm, which then identifies natural groupings in this derived feature space. This dimensionality reduction technique enables effective clustering of complex temporal patterns while maintaining computational efficiency.

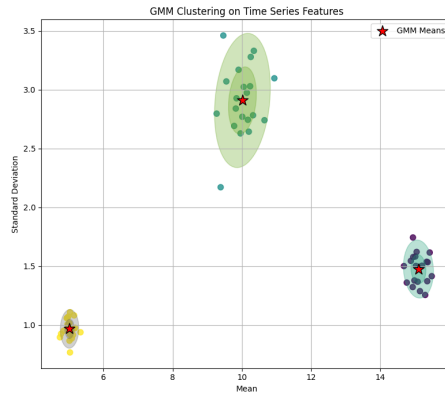


Figure 4: GMM Clustering on Time Series Features. The plot shows the 2D feature space (Mean vs. Std Dev) with GMM-identified clusters and their elliptical boundaries, highlighting the probabilistic nature of GMM assignment.

3.5 Performance Evaluation of Unsupervised Clustering

A widely adopted metric for assessing clustering quality is the Silhouette Coefficient, which quantifies two key properties: how well each data point belongs to its own cluster and how distinct it is from neighboring clusters. Mathematically, it computes a score for each data point ranging from -1 (poor clustering) to +1 (ideal clustering), with values near zero indicating overlapping clusters.

The Silhouette score is the average of the Silhouette values for all data points. A higher score indicates better clustering quality. To calculate the Silhouette value ($s(i)$) for a data point i :

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))},$$

where $a(i)$ is the average Euclidean distance between data point i and all other points in the same cluster that i belongs to, and $b(i)$ is the smallest average Euclidean distance between data point i and all points in any other cluster that i does not belong to.

Example: We determine the optimal number of clusters K for N time series by evaluating $K \in 2, \dots, N - 1$ and selecting the value that maximizes the Silhouette score, which measures both intra-cluster cohesion and inter-cluster separation. Consider 10 univariate time series of length 3; Time Series 1: [1.0, 1.2, 1.1], Time Series 2: [1.1, 1.0, 1.3], Time Series 3: [1.5, 1.7, 1.6], Time Series 4: [1.6, 1.5, 1.8], Time Series 5: [5.0, 5.1, 5.2], Time Series 6: [5.1, 5.0, 5.3], Time Series 7: [5.5, 5.7, 5.6], Time Series 8: [5.6, 5.5, 5.8], Time Series 9: [10.0, 10.1, 10.2], and Time Series 10: [10.1, 10.0, 10.3].

An evaluation using the `silhouette_score` from Python's `sklearn.metrics` library revealed that the optimal number of clusters for this dataset is $K = 3$, with a score of 0.9169. This indicates that clustering the data into 3 clusters provides the best structure in terms of within-cluster cohesion and between-cluster separation for this dataset. This finding is consistent with the known structure of our test dataset:

- **Cluster 1:** Time Series 1 to 4 (values close to 1)
- **Cluster 2:** Time Series 5 to 8 (values close to 5)
- **Cluster 3:** Time Series 9 and 10 (values close to 10)

The data naturally forms three distinct groups, which the Silhouette method successfully identifies.

A graphical representation of these Silhouette scores for different numbers of clusters is very helpful. As shown in Figure 5, the peak of the curve clearly indicates the optimal number of clusters.

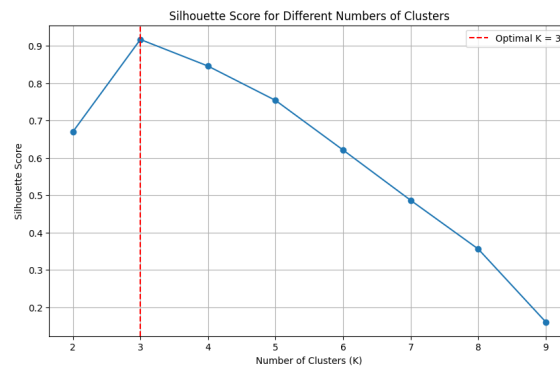


Figure 5: Silhouette Score for Different Numbers of Clusters. The vertical dashed line indicates the optimal number of clusters (K).

By calculating the Silhouette score for clusterings with different numbers of clusters, you can identify the K that best reflects the underlying structure in the data. This method is especially useful when we have no prior knowledge about the number of clusters present in our data.

3.6 Current Research Trends in Clustering (Brief Mention)

GMMs have been extended to more flexible frameworks. For instance, [Setoudehtazangi and *et al.* \(2024\)](#) introduced a mixture of ARCH models by incorporating normal mean–variance mixture distributions. Furthermore, robust mixture of linear experts models have been developed by leveraging normal mean–variance mixtures for unobserved error terms in both censored and uncensored data, as proposed by [Manouchehri and *et al.* \(2025\)](#) and [Setoudehtazangi and *et al.* \(2025\)](#).

Recent advances in deep learning have also revolutionized time series clustering through end-to-end learnable frameworks. Current methods primarily focus on three directions: deep representation learning using autoencoders and adversarial networks, improved deep embedded clustering with attention mechanisms and uncertainty modeling, and neural temporal pattern mining via temporal CNNs and transformers.

4 Supervised Time Series Grouping via Classification

Supervised time series classification refers to the process of building predictive models that can automatically categorize time-dependent data into predefined classes. Unlike conventional classification tasks, this approach deals with the unique challenges of temporal data, where the ordering and timing of observations carry significant information. The trained model learns to identify discriminative patterns in labeled historical time series, enabling it to generalize to new, unseen instances.

4.1 Feature-based extraction with Machine Learning Classifiers (Hybrid Analysis)

Similar to feature-based clustering, this approach starts by extracting relevant features from each time series (e.g., statistical moments, shape descriptors, frequency components). These features then form a fixed-length input for a standard supervised machine learning classifier (e.g., Logistic Regression, Support Vector Machines (SVM), Random Forests, or Neural Networks).

Example: Figure 6 illustrates a feature-based approach for classifying time series into ascending or descending patterns. Panel (a) displays raw synthetic time series with class-labeled coloring, revealing the inherent temporal structures. Panel (b) demonstrates the classification pipeline, where extracted features (slope vs. mean) are projected into a 2D feature space. A linear Support Vector Machine classifier learns an optimal decision boundary in this transformed space, confirming that simple feature representations can effectively capture essential temporal characteristics for classification tasks. This approach demonstrates the advantage of feature engineering in simplifying complex time series analysis.

4.2 Model-Based Approaches in Supervised Learning (Model-Based Classification)

Model-based classification assumes that the observed data points (or time series) are generated from a mixture of underlying probability distributions, where each distribution corresponds to a distinct class.

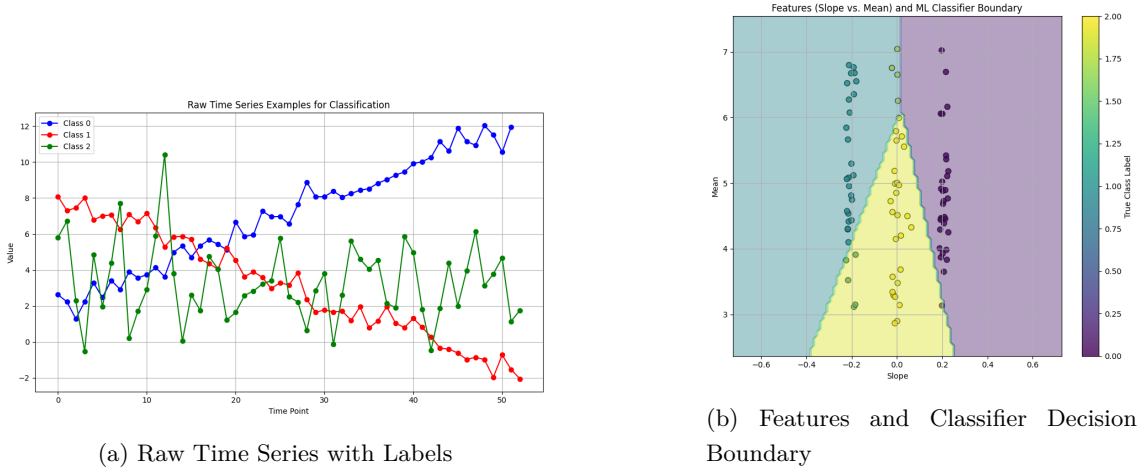


Figure 6: Feature-Based Classification Example. (a) Displays raw synthetic time series data colored by their class labels. (b) Shows extracted features (e.g., slope vs. mean) with the decision boundary learned by a machine learning classifier (Linear SVM)

4.2.1 Gaussian Mixture Models (GMM) for Classification or GMM approach for Classification

Gaussian Mixture Models (GMMs) can be also adapted for classification by training a separate GMM for each class on labeled data. To classify a new time series, features are extracted and evaluated under each class-specific GMM. The instance is assigned to the class with the highest likelihood. This probabilistic approach naturally provides confidence measures. This method, termed GMM classifier or probabilistic GMM-based classification, typically operates on time-series-derived feature vectors, analogous to GMM clustering.

Example: Using slope and mean features, a GMM was trained for each class (Figure 7). The plot shows 2D feature space with Gaussian distributions (ellipses) representing each class. New time series are classified by selecting the class with highest probability.

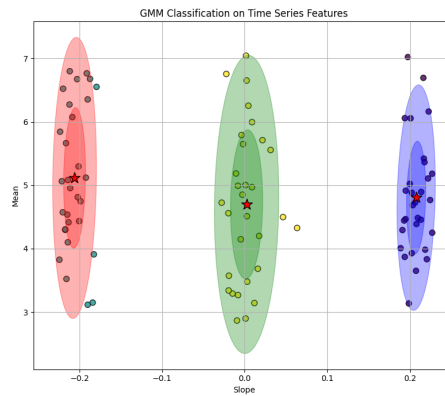


Figure 7: GMM Classification on Time Series Features. The plot shows the 2D feature space with GMM-identified clusters for each class and their elliptical boundaries. A new series would be classified based on which class's GMM generates the highest probability.

Remark: The core distinction between GMM clustering and classification lies in their learning paradigms and objectives. Clustering (unsupervised) employs GMM to discover inherent groupings within unlabeled time series data by analyzing the statistical properties of their

features, without any prior knowledge of categories. In contrast, classification (supervised) trains separate GMMs on labeled data for each predefined class, enabling the model to discriminate between known categories for future predictions. While both methods assume the data follows a mixture of Gaussian distributions, clustering is exploratory by nature, seeking to reveal hidden patterns, whereas classification is discriminative, aiming to categorize new instances into established classes.

4.3 Standard Evaluation Metrics for Classification tasks

Classification performance is typically assessed using several well-established metrics, each providing unique insights into model behavior. The most common measures include *accuracy* (overall correct prediction rate), *precision* and *recall* (for class-specific performance), and the *F1-score* (harmonic mean of precision-recall). For imbalanced datasets, metrics like *Matthews correlation coefficient (MCC)* and *balanced accuracy* are preferred. The *area under ROC curve (AUC-ROC)* evaluates overall discriminative ability across thresholds, while *confusion matrices* provide detailed error breakdowns. Recent trends also incorporate *computational efficiency* metrics (e.g., inference time) and *robustness measures* for real-world deployment scenarios. Choice of metrics depends on the specific application requirements and class distribution characteristics.

4.4 Current Research Trends in Classification (Brief Mention)

Beyond feature-based methods, more advanced approaches operate directly on raw time series data without explicit feature extraction. These include distance-based classifiers like the 1-NN classifier with DTW distance, which classifies new instances by identifying their nearest training sample. Additionally, deep learning models such as CNNs and RNNs (e.g., LSTMs) automatically learn complex temporal patterns from raw data, often achieving state-of-the-art results.

Acknowledgements

Thematic exploration in this work was initially supported by outputs from advanced AI models, including DeepSeek-V2, Gemini, ChatGPT, and Grok. These AI-generated suggestions were not used verbatim; rather, they were systematically reviewed and interpreted within established qualitative research frameworks. The final insights presented herein reflect critical human judgment and have been independently curated to ensure conceptual rigor and scholarly integrity.

References

- Batista, G. E., Keogh, E. J., Tataw, O., & Campbell, J. F. (2014), CID: An Efficient Complexity-Invariant Distance for Time Series, *Proceedings of the 2014 SIAM International Conference on Data Mining (SDM)*.
- Caiado, J., D’Urso, P., & Maharaj, E. A. (2021), *Time series clustering and classification*, Chapman & Hall/CRC.
- Chen, L., & Ng, R. T. (2005), On the design of effective and efficient methods for stock trend prediction, *The VLDB Journal*, **14**(1), 1-14.
- Keogh, E., Chu, S., Hart, D. & Pazzani, M. (2004), Segmenting time series: A survey and novel approach, *Data mining in time series databases*, **57**, 1-22.

- Manouchehri, T., Setoudehtazangi, F., Nematollahi, A.R. & McLachlan, G. (2025), Robust mixture of linear experts modeling based on normal mean–variance mixture distributions, *Submitted*.
- Setoudehtazangi, F., Manouchehri, T., Nematollahi, A.R. & Caporin, M. (2024), Time series clustering based on latent volatility mixture modeling with applications in finance, *Mathematics and Computers in Simulation*, **223**, 543-564.
- Setoudehtazangi, F., Manouchehri, T., Nematollahi, A.R. & Wraith, D. (2025), Robust mixture of linear experts modeling based on normal mean–variance mixture distributions, *Submitted*.
- Vlachos, M., Yu, P. S., & Castelli, V. (2002), Locally causal statistical subsequence: a method for anomaly detection in sensor networks, *Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM)*, 741-744. IEEE.
- Ye, L., & Keogh, E. (2009), Time series shapelets: a new primitive for time series data mining, *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '09)*.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Statistical Inference for Spatio-Temporal Gibbs Point Processes

Morteza Raeisi^{*1}

¹A.R.CA.D. Foundation, Paris, France

Abstract

Gibbs models are flexible point processes that specify point interactions via a probability density relative to the unit-rate Poisson process. They capture local or Markovian dependencies and are useful for modeling pairwise interactions, accounting for clustering, inhibition, or randomness structures, and combining multiple structures across scales through hybridization. We extend two spatial Gibbs point process models to a spatio-temporal framework. However, these single-structure Gibbs models can be too simplistic for capturing the complexity of phenomena in fields such as seismology, epidemiology, and forestry, where multiple structures occur across various spatial and temporal scales. To address this, we introduce multi-scale forms of the two spatio-temporal Gibbs models using a hybridization approach and develop statistical inference method based on composite likelihoods and simulate the models using the birth–death Metropolis–Hastings algorithm. We illustrate the applicability of these new point processes by modeling forest fire occurrences in France and Spain.

Keywords: Composite likelihood; Gibbs point processes; Hybridization; Multi-scale structures; Spatio-temporal modeling.

Mathematics Subject Classification (2020): 60G55, 62M30, 62F12.

^{*}Corresponding author, morteza.raeisi@fondationarcad.org



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Estimating the Effective Reproduction Number of Tuberculosis in a Branching Process

Maryam Rastegar^{*1}, Vahid Fakoor², Mohammad Taghi Shakeri³

¹Department of Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran

²Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran

³Department of Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran

Abstract

Background:

Tuberculosis (TB) remains a major public health challenge in Iran, particularly in border provinces such as Sistan and Baluchestan, where cross-border migration from high TB-burden countries exacerbates transmission dynamics. Accurate estimation of the effective reproduction number (R_t) is crucial for monitoring the epidemic and guiding control strategies, especially in settings with high population mobility.

Methods:

This study employed a branching process model incorporating migration to estimate the Effective Reproduction Number (R_t) of smear-positive pulmonary tuberculosis (SPPTB) in Sistan and Baluchestan Province from March 2011 to March 2022. A Bayesian framework was used to account for uncertainty and integrate prior epidemiological knowledge. Surveillance data from 7,237 registered SPPTB cases were analyzed. The EpiEstim package in R was utilized for dynamic estimation of R_t while descriptive statistics were computed for baseline characteristics and transmission parameters.

Results:

The mean age of patients was 53.24 ± 21.10 years, and 51.6% were male. Afghan nationals accounted for 33.7% of primary cases, contributing to 32.5% of secondary infections. The estimated mean R_t was 1.10 (95% CrI: 0.93 -1.27), indicating sustained transmission. The mean values for immigration and offspring components were 1.22 and 1.07, respectively, highlighting the influence of migration on local transmission. Fluctuations in R_t over time reflected both internal and imported sources of infection.

Conclusion:

The elevated R_t and the significant contribution of migration emphasize the need for targeted TB control strategies in Sistan and Baluchestan. Cross-border coordination, routine TB screening among migrants, and tailored interventions are critical for reducing transmission. The Bayesian branching process model with immigration offers a robust and context-sensitive tool for epidemic monitoring in high-mobility, high-burden regions.

^{*}Corresponding author, mr241361@gmail.com

Keywords: Tuberculosis, Effective Reproduction Number, Branching Process, Bayesian Inference.

Mathematics Subject Classification (2020): xxAxx, xxBxx, xxCxx.

1 Introduction

Tuberculosis (TB) remains a leading cause of morbidity and mortality globally, particularly in regions burdened by poverty, migration, and limited access to healthcare services. Despite progress in TB control, border provinces like Sistan and Baluchestan (S&B) in southeastern Iran continue to face high transmission rates due to socio-economic vulnerabilities and close proximity to high-burden countries such as Afghanistan and Pakistan [Rastegar *et al.* \(2024\)](#). These neighboring countries contribute significantly to cross-border TB transmission, especially in the context of drug-resistant strains and irregular migration patterns [Stop T. Partnership \(2022\)](#).

In such complex epidemiological contexts, the effective reproduction number (R_t) is a critical indicator for understanding disease transmissibility. It reflects the average number of secondary infections caused by an infected individual over time in a partially susceptible population [Huisman *et al.* \(2022\)](#). Accurate and timely estimation of R_t is vital for assessing the impact of interventions and predicting outbreak trajectories. However, traditional compartmental models (e.g., SIR) often fall short in heterogeneous and high-mobility settings like S&B, as they generally assume homogeneous mixing and closed populations (Kucharski *et al.*, 2020).

To address these limitations, this study utilizes a branching process model with immigration, a stochastic modeling framework particularly well-suited for capturing the randomness and heterogeneity of transmission events, especially in regions with frequent population movement. To enhance inference under uncertainty, a Bayesian approach is employed, allowing for integration of prior knowledge and generating credible intervals for parameters such as R_t . Bayesian models are increasingly favored in real-time epidemic analysis due to their flexibility and ability to incorporate uncertainty in reporting, serial interval (SI) distributions, and parameter variability [Rastegar *et al.* \(2024\)](#). SI is defined as the time duration between the onset of symptoms in a primary case and the onset of symptoms in a secondary case infected by the primary one. It is a critical parameter in infectious disease modeling, as it reflects the timing of transmission events and plays a key role in estimating the effective reproduction number (R_t) [Gostic *et al.* \(2020\)](#). Accurate estimation of SI is particularly important for diseases like tuberculosis, where transmission can occur over variable and sometimes extended periods due to long incubation times and delayed diagnosis. In this study, the SI distribution is modeled using a Gamma distribution to capture the right-skewed nature of TB transmission intervals. Incorporating uncertainty in the SI distribution allows for more robust inference of R_t , particularly when using Bayesian frameworks.

This study proposes a Bayesian branching process model incorporating migration to estimate the effective reproduction number of TB in S&B Province using surveillance data from 2011 to 2022. By accounting for cross-border movement and population heterogeneity, this approach generates more accurate and context-sensitive estimates of R_t , providing valuable insights for targeted TB control interventions. The findings will support evidence-based decision-making in border regions and contribute to the methodological advancement of stochastic epidemic modeling in complex settings.

2 Materials and Methods

2.1 Theoretical Framework

Understanding the spread of infectious diseases like TB requires robust mathematical modeling approaches. These models range from deterministic systems that simulate population-wide trends to stochastic models that account for individual-level transmission variability. In the present study, we employed a stochastic branching process model that incorporated migration to simulate the progression and spread of TB cases. This model is particularly appropriate for diseases with variable transmission patterns and in regions affected by significant population movement.

Branching processes (PB) provide a flexible framework for modeling the generation of secondary infections by each infectious individual, accounting for random variation in transmission. Migration was modeled as a factor influencing the importation and exportation of cases between different regions, reflecting real-world population flows in and out of Sistan and Baluchestan (S&B). Bayesian estimation techniques were applied to infer the R_t , integrating prior epidemiological knowledge with observed data to improve the accuracy of inference under uncertainty. These methods have been shown to be particularly powerful in analyzing real-world epidemic data, especially when integrating social, demographic, and ecological factors [Fraser \(2007\)](#).

We assume that each infectious individual infects a random number of susceptible individuals distributed as a random variable ξ . Let us start with s infected individuals. All infected individuals due to contact with them are called the first generation, and let us denote their number by Z_1 . Infected individuals in contact with the first generation form the second generation, with Z_2 individuals, etc. This process can be depicted as a branching process ([Figure 1](#)).

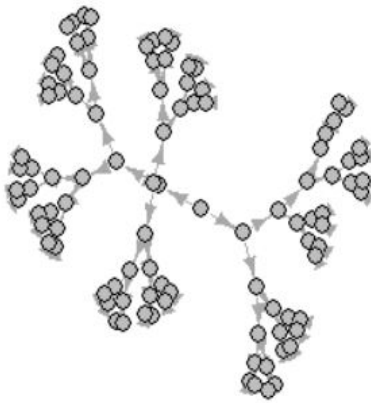


Figure 1: an example of branching process.

Let $\xi_{n,j}$ are independent and identically distributed random variables (i.i.d. r.v.) with the same distribution as ξ . The distribution of ξ is called offspring distribution, and the mean of ξ is denoted by $m = E\xi_{n,j}$. Formally, we define $[Z_n, n = 1, 2, \dots]$ as follows:

$$\begin{aligned}
Z_0 &= s \\
Z_1 &= \xi_{1,0} + \cdots + \xi_{s,0} \\
Z_2 &= \xi_{1,1} + \cdots + \xi_{Z_1,1} \\
&\vdots \\
Z_n &= \xi_{1,n} + \cdots + \xi_{Z_{n-1},n} = \sum_{j=1}^{Z_{n-1}} \xi_{n,j}, \quad n = 1, 2, \dots
\end{aligned}$$

where $\xi_{n,j}$ is the number of infected by the j -th individual of n -th generation.

The sequence of $(Z_n, n = 1, 2, \dots)$ is called the Bienaymé–Galton–Watson process (BG-WBP).

The event $\{Z_n = 0, \text{ for some } n \geq 1 \mid Z_0 = 1\}$ is called extinction. Denote the probability of extinction $q = P(Z_n = 0, \text{ for some } n \geq 1 \mid Z_0 = 1)$. From the theory of the branching process, it is known that for $m \leq 1$, $q = 1$, and for $m > 1$, $q < 1$.

If the process starts with s individuals, the probability of extinction is $\{Z_n = 0 \text{ for some } n \geq 1 \mid Z_0 = s\} = q^s$.

Depending on whether the offspring mean m is less than, equal to, or greater than 1, the process is called subcritical, critical, and supercritical, respectively.

Let us denote by Y the total progeny of BGWBP or the total number of infected individuals by the end of the outbreak. It is defined as follows:

- The generation sizes can then be written as the collection $\{Z_n; n = 1, 2, \dots, n\}$.
- The process $\{Z_n; n \geq 0\}$ can be recursively defined using the representation

$$Y = Z_n = \sum_{j=1}^{Z_{n-1}} \xi_{n,j}, \quad n = 1, 2, \dots \quad (1)$$

The probability distribution of $\xi_{n,j}$ for all n and j is

$$P(\xi_{n,j} = k) = p_k \quad k \geq 0 \quad (2)$$

The sequence $\{p_k : k \geq 0\}$ is referred to as the offspring distribution.

Branching process with immigration

A useful and realistic modification of the BP is the incorporation of an immigration component into the population from an outside source. More precisely, the branching process with immigration (BPI) is defined as

$$X_t = \sum_{j=1}^{X_{t-1}} \xi_{t,j} + I_t, \quad t = 1, 2, \dots \quad (3)$$

where $\xi_{n,j}$ are the same as in (1), and I_n is the number of immigrants entering the n -th generation from an outside source. The random variables I_1, I_2, \dots are assumed to be independent and identically distributed (i.i.d.) and are independent of the process $\{\xi_{n,j}; j \geq 1, n \geq 1\}$.

We denote their distribution by

$$P(I_j = j) = q_j, \quad j \geq 0.$$

Let $E(I_1) = b$, $Var(I_1) = v^2$.

Bayesian estimation of the offspring means in branching processes

The Bayesian estimation method is one of the most commonly used methods developed to estimate R_t in real-time through equation (3), which relies on a branching process model. In a branching process, each infected individual can produce a random number of secondary infections, and the total number of infections at any given time is the sum of all the infections produced by each infected individual.

The branching process $\{X_t; t \geq 0\}$ was defined in (3)

Where $\xi_{t,j}$ can be interpreted as the number of new cases produced by the j^{th} infected case in the t^{th} generation. The probability distribution of $\xi_{t,j}$ for all t and j is

$$P(\xi_{t,j} = k) = p_k \quad k \geq 0$$

The sequence $\{p_k : k \geq 0\}$ is referred to as the offspring distribution. Most often we do not have data on the number of infected ones by each infectious, but on the total number of infected individuals for a given time (X_t). The Poisson process in the branching process by assuming that the number of new cases at any given time is a random variable that follows a Poisson distribution is referred to as the offspring distribution, with the mean of this distribution being the product of R_t and the sum of the past incidence X_{t-s} of cases, weighted by w_s , the probability mass function of the generation time (the time between infection in a case and their infector). In practice, as infection is difficult to observe, the incidence of symptomatic cases can be used instead and w_s can be approximated by SI.

$$p_k = \frac{\exp(-R_t) R_t^k}{k!} \quad k \geq 0$$

The model assumes that the incidence of new cases on day t , X_t , can be represented by a Poisson process:

$$X_t \sim \text{Pois} \left(R_t \sum_{s=1}^t X_{t-s} w_s \right) \quad (4)$$

In practice, as the infection is difficult to observe, the incidence of symptomatic cases can be used instead and SI can approximate w_s . Parametric models for the SI such as the Gamma, Weibull, and log-normal distributions are widely used to model infectious diseases [Cowling et al. \(2009\)](#). Compared to the other distributions, the Gamma distribution has a more flexible shape, which allows it to fit a wider range of SI distributions, including those with shorter or longer tails. The Gamma distribution is also a more natural choice when the SI is measured in discrete time units, such as days or hours, as it models the probability of observing a certain number of days between the onset of symptoms in the primary and secondary cases. The Weibull and log-normal distributions are also similar to the Gamma distribution in terms of empirical form [Firth \(1988\)](#). Here, we assume a Gamma distribution for the SI distribution. The estimation of the shape and scale parameters of the Gamma distribution for the SI can be achieved using the method of uncertainty on the SI distribution as described in [Cori et al. \(2020\)](#).

Following [Reich et al. \(2009\)](#), we perform Bayesian parametric estimation of the SI distribution from such data using data augmentation Markov Chain Monte Carlo (MCMC). If we use a Gamma-distributed SI distribution offset by one day, the MCMC estimation procedure leads to a joint posterior sample for the parameters of the chosen distribution, i.e., a list of possible sets of parameter values, with each parameter set corresponding to a single step in the MCMC chain. Each parameter set, i , in the posterior sample is then converted into a discrete probability mass function, $w_s(i)$. ($s = 0, 1, 2, \dots$) as follows: The probability of the SI lasting 0-time steps, $w_s(0)$, is set to 0 as in [Cori et al. \(2013\)](#), and the probability $w_s(i)$ for any other timestep $s = 1, 2, 3, \dots$ is obtained by integrating the probability density function defined by this parameter set between $s - 0.5$ and $s + 0.5$. For each i , the function $w_s(i)$ is then renormalized to sum to 1. The posterior sample of SI distributions, $w_s(i)$. ($i = 1, 2, \dots, n$), is then used along with disease incidence time series to estimate the R_t as described in the following [Cori et al. \(2013\)](#).

As defined by [Cori et al. \(2013\)](#), R_t is a measure of the transmission potential of an infectious disease at a specific point in time. X_t is calculated as the ratio of the number of new infected cases at time t , denoted by I_t , and the total infection potential across all infected individuals at time t , denoted by Λ_t . If there is a single SI distribution $w_s(i)$. ($s = 1, 2, \dots$),

representing the probability of a secondary case arising a period s after the primary case, each incident case that appeared at a previous time step $t - s$ contributes to the current infectiousness at a relative level given by w_s [Cori et al. \(2013\)](#).

Therefore, conditional on w_s , t can be computed as

$$\Lambda_t = \sum_{s=1}^t X_{t-s} w_s$$

Given a SI distribution w_s , data on the total number of incident cases up to the previous time step $(X_1, X_2, \dots, X_{t-1})$, and R_t , the expected number of incident infected cases at time t is

$$E(X_t | X_1, X_2, \dots, X_{t-1}, w_s) = R_t \Lambda_t$$

Assuming that the number of cases at time step t is drawn from a Poisson distribution. [Cori et al. \(2020\)](#), we assume that R_t is constant over a time period $[t - \tau, t]$, with τ representing the length of the time window over which R_t is estimated. Given R_t and conditional on the previous incidence data X_0, X_1, \dots, X_{t-1} , is then

$$P(X_t | X_0, X_1, \dots, X_{t-1}, w_s, R_t) = \frac{(R_t \Lambda_t)^{X_t} \exp(-R_t \Lambda_t)}{X_t!} = \prod_{k=t-\tau}^t \frac{(R_t \Lambda_k)^{X_k} \exp(-R_t \Lambda_k)}{X_k!}$$

To estimate the R_t of pulmonary tuberculosis, we used the number of total infected individuals Λ_k at a given time k , with a time window of $\tau = 7$ days, and the SI distribution w_s . We employed a Gamma distribution prior and conjugated it with the Poisson likelihood to derive an analytical formulation of the posterior distribution of R_t [Thompson et al. \(2019\)](#).

Using a Bayesian framework with a Gamma-distributed prior for R_t as in [Cori et al. \(2013\)](#), the posterior distribution of R_t given past incidence data and conditional on the SI distribution, w_s , is

$$\begin{aligned} & P(R_t | X_0, X_1, \dots, X_{t-\tau}, X_{t-\tau+1}, \dots, X_t, w_s) \\ & \propto P(X_{t-\tau}, X_{t-\tau+1}, \dots, X_t | X_0, X_1, \dots, X_{t-\tau-1}, w_s, R_t) P(R_t) \\ & = \left(\prod_{k=t-\tau}^t \frac{(R_t \Lambda_k)^{X_k} \exp(-R_t \Lambda_k)}{X_k!} \right) \left(\frac{R_t^{a-1} \exp(-\frac{R_t}{b})}{\Gamma(a) b^a} \right) \\ & = R_t^{a + \sum_{k=t-\tau}^t X_k - 1} \exp \left(-R_t \left(\sum_{k=t-\tau}^t \Lambda_k + \frac{1}{b} \right) \right) \times \prod_{k=t-\tau}^t \frac{(\Lambda_k)^{X_k}}{X_k!} \end{aligned} \quad (5)$$

where a and b are the shape and scale parameters of the Gamma distributed prior to R_t . We used a Gamma-distributed prior, conjugated to the Poisson likelihood, to obtain an analytical formulation of the posterior distribution of R_t . According to the expression above, the posterior distribution for R_t given the incidence data, conditional on the SI distribution w_s , is a Gamma distribution with the shape parameter $a + \sum_{k=t-\tau}^t X_k$, and the scale parameter $\frac{1}{\sum_{k=t-\tau}^t \Lambda_k(w_s) + \frac{1}{b}}$.

To select the parameters a and b for the Gamma prior distribution of R_t , we should consider the type of branching process the data represents. If the data is from a subcritical branching process, we set the mean of the Gamma distribution to be less than one. For a critical branching process, we set the mean to equal one. We set the mean to be greater than one for a supercritical branching process. This approach ensures that the prior distribution reflects the expected behavior of the branching process based on the underlying data. To obtain a sample from the full posterior distribution for R_t given both the incidence and the SI data, we consider each possible SI distribution $w_s(i)$ ($i = 1, \dots, n$). For each i , we draw a sample of size m from the Gamma posterior distribution of R_t , given the incidence data and conditional on the SI distribution $w_s(i)$. We thereby obtain a sample of size $n \times m$ drawn from the posterior distribution of R_t given both the incidence and SI data, from which the posterior mean and 95% credible intervals of R_t can be computed [Cori et al. \(2013\)](#), [Nishiura and Chowell \(2014\)](#), [McBryde et al. \(2009\)](#), [Wallinga and Lipsitch \(2007\)](#).

2.2 Statistical Analysis

Descriptive statistics were computed to summarize the data, with continuous variables expressed as mean \pm standard deviation (SD) and categorical variables reported as frequencies and percentages. The chi-square goodness-of-fit test was applied to evaluate distribution uniformity across qualitative variables. All statistical analyses were performed using SPSS (v22.0), R (v4.2.1), and Microsoft Excel (2014), with a significance threshold of $p < 0.05$ for hypothesis testing. For dynamic estimation of the effective reproduction number (R_t), we implemented a Bayesian inference framework, which accounts for uncertainty by integrating prior epidemiological knowledge with observed transmission data. The EpiEstim package in R was employed to model R_t over time, providing time-varying estimates of disease spread.

2.3 Study Population and Data Source

This study utilized a national historical cohort design, analyzing data from March 2011 to March 2022 obtained from the Health Deputy of Zahedan University of Medical Sciences. The dataset included 7,237 reported cases of SSPTB, diagnosed and registered across S&B province. Given their higher infectivity and public health significance, SSPTB cases were specifically selected to better assess transmission dynamics and estimate reproduction numbers (R_t).

3 Result

A total of 7,237 new smear-positive pulmonary tuberculosis patients were included in this study. The mean age of the patients was 53.24 ± 21.10 years, with a notable age distribution: 2.9% were under 15 years, 22.9% were aged 15–35, 35.3% were 36–63, and 38.9% were over 64 years. Age distribution differences were statistically significant ($P < 0.001$). The cohort comprised 51.6% males and 48.4% females, with a significant gender difference ($P < 0.001$). The majority of patients were Iranian (83.2%), while 16.8% were of other nationalities, a difference that was also statistically significant ($P < 0.001$).

Table 2 presents descriptive statistics of R_t and related parameters for pulmonary tuberculosis (PTB) in Iran from 2018 to 2022, highlighting the dynamic nature of TB transmission in Sistan and Baluchestan. The mean R_t during this period was 1.10, with a 95% credible interval of (0.93, 1.27), indicating that, on average, each infectious individual caused more than one secondary case—suggesting ongoing transmission. This value, which remained above the epidemic threshold, reflects a persistent risk of disease spread. Notably, the immigration mean (b) ranged from 0.72 to 2.27, with an average of 1.22, emphasizing the significant contribution of migration to local TB dynamics. The results indicate that R_t was not solely driven by internal transmission (offspring mean of 1.07) but was also meaningfully influenced by cross-border and internal migration, which introduced new infections into the population.

Figure 2 depicts the monthly trends SSPTB incidence, R_t , and SI distribution in Sistan and Baluchestan Province from March 2011 to March 2022. The epidemic curve shows fluctuating case counts with peaks reaching 75 monthly cases and troughs near zero, indicating intermittent outbreaks superimposed on endemic transmission. R_t estimates varied between 0.0–2.0, with sustained periods above 1.0, suggesting ongoing transmission, while sharp declines likely reflected intervention impacts. The SI distribution peaked at low probabilities (0.00–0.02), consistent with rapid transmission in close-contact settings. Notably, incidence peaks correlated with R_t surges (> 1.5), revealing periodic epidemic growth phases amidst generally stable endemic transmission patterns.

Table 1: Baseline Characteristics and Risk Factors of New Smear-Positive Pulmonary Tuberculosis Patients in S&B province (n = 7237).

Characteristics	New smear-positive PTB patients		P Value
Age (year)	Mean±SD	53.24±21.10	—
Age (year)	<15 (n %)	211 (2.9)	<0.001*
	15–35 (n %)	1655 (22.9)	
	36–63 (n %)	2556 (35.3)	
	>64 (n %)	2815 (38.9)	
Gender	Female (n %)	3501 (48.4)	<0.001*
	Male (n %)	3736 (51.6)	
Nationality	Iranian (n %)	6022 (83.2)	<0.001*
	Others (n %)	1215 (16.8)	

* Significant at a level of 0.05

Table 2: Descriptive statistics of R_t and SI of PTB disease in Iran during 2011–2022

descriptive statistics	Mean (b) (Immigration Mean Estima- tion)	SD (b) (Immigration Standard deviation Estimators)	Mean (m) (Offspring Mean Estima- tion)	SD (m) (Offspring Standard deviation Estimation)	Mean (R_t) Total	SD (R_t) Total
Minimum	0.72	0.10	0.63	0.04	0.64	0.039
1st Quartile	0.99	0.13	0.91	0.05	0.92	0.051
Median	1.17	0.14	0.98	0.06	1.02	0.054
Mean	1.22	0.15	1.07	0.06	1.10	0.057
3rd Quartile	1.37	0.16	1.21	0.07	1.25	0.062
Maximum	2.27	0.28	2.01	0.10	2.05	0.098
95%CrI	(0.77, 1.67)		(0.89, 1.25)		(0.93, 1.27)	

* Standard deviation (SD), 95% credible interval (95% CrI).

4 Discussion:

A total of 7,237 new smear-positive pulmonary tuberculosis (SPPTB) cases were analyzed from March 2011 to March 2022 in S&B Province. Most patients were Iranian nationals (83.2%), while 16.8% were of other nationalities, predominantly Afghan. Descriptive statistics of R_t during 2018–2022 showed a mean R_t of 1.10 (SD = 0.057), with a 95% credible interval (CrI) of (0.93, 1.27). The immigration mean (b) was 1.22 and offspring mean (m) was 1.07, indicating a notable contribution of immigration to transmission dynamics. The maximum estimated R_t reached 2.05, emphasizing periods of intense transmission risk. These results highlight the dual role of internal and external (migration-related) factors in maintaining TB transmission in the region.

Tuberculosis remains a significant public health concern in regions with high population mobility, such as S&B. This province, situated on Iran's southeastern border, is uniquely vulnerable due to its proximity to Afghanistan and Pakistan—countries with some of the highest TB and drug-resistant TB incidence rates globally [World Health Organization \(2021\)](#). In the current study, 16% of SPPTB cases were Afghan nationals. This high transmission rate from migrants strongly implicates cross-border movement in sustaining TB transmission.

The elevated R_t values estimated in this study support these concerns. The mean R_t of 1.10, exceeding the threshold of 1, indicates that the epidemic is self-sustaining. The immigration mean (1.22) being higher than the offspring mean (1.07) suggests that imported cases via migration are a major driver of ongoing transmission. These findings are consistent

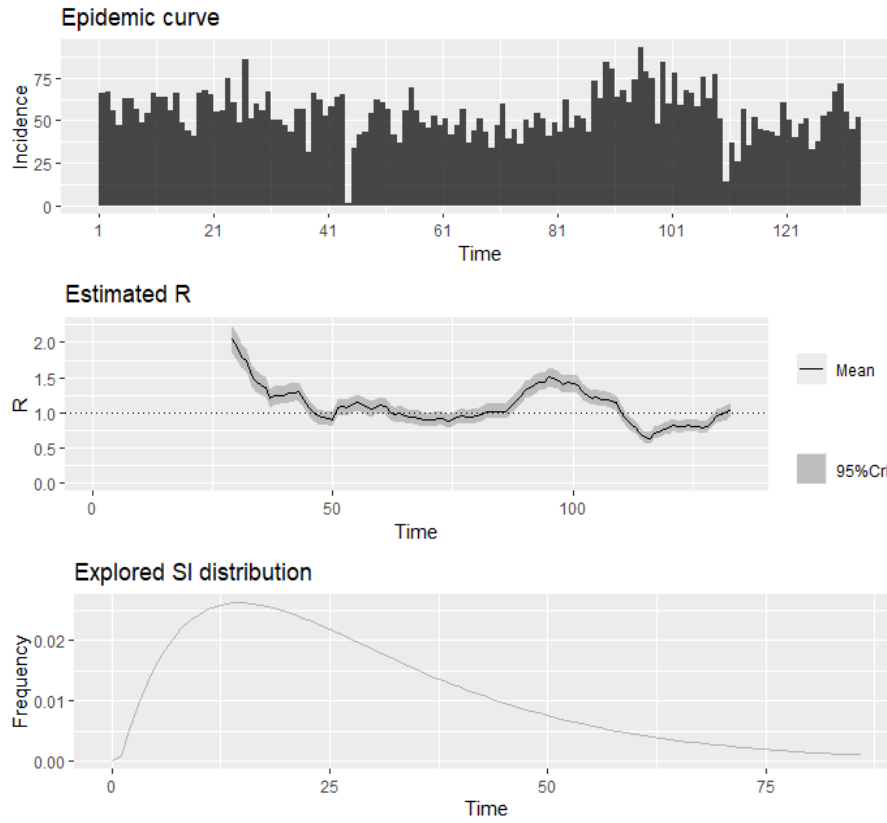


Figure 2: New cases, R_t and SI of SPPTB in SB province from March 2011 to March 2022 monthly.

with previous literature that highlights how population displacement and migration amplify the spread of infectious diseases, especially TB [Kimbrough *et al.* \(2012\)](#).

Bayesian estimation through branching process modeling provided robust insight into the dynamic contribution of both internal transmission and migration. The use of such probabilistic models is particularly valuable in high-mobility regions where surveillance data may be incomplete or uncertain [Cori *et al.* \(2017\)](#). These findings further underline the importance of integrating border health monitoring systems, routine TB screening among migrants, and culturally adapted treatment services in TB control programs.

5 Conclusion

The findings of this study underscore that the R_t of TB in Sistan and Baluchestan is influenced not only by internal community transmission but significantly by immigration from neighboring high-burden countries. The observed R_t levels above 1 highlight the need for urgent and targeted interventions. Given S&B's geographic location and socio-political context-including porous borders and high population mobility-the elevated R_t underscores the need for targeted TB control strategies that address both transmission and migration-related risks. To reduce TB transmission, especially smear-positive cases that are more infectious, Iran's national TB control strategy must include comprehensive policies addressing migrant health. These include routine screening, cross-border coordination, and ensuring access to care for undocumented populations. Addressing these factors is essential for achieving TB elimination goals in this high-risk region.

List of abbreviations

BP	Branching Process
TB	Tuberculosis
PTB	Pulmonary tuberculosis
SPPTB	Smear-positive pulmonary tuberculosis
EXTB	Extra-pulmonary tuberculosis
R_t	Effective reproduction number
SI	Serial Interval
S&B	Sistan and Baluchestan

Ethics approval and consent to participate

Informed consent was waived by the Research Ethics Committee of Mashhad University of Medical Sciences, Code: IR.MUMS.FHMPM.REC.1401.054.

Consent for publication

Not Applicable

Competing interests

The authors declare that they have no conflicts of interest.

References

- Rastegar M., Fakoor V., Nazar E., Nasehi M., Sharafi S., Shakeri M. T. (2024), Effective Reproduction Number of Smear-Positive Pulmonary Tuberculosis in Iran: A Registry-Based Study (2011-2021), *Journal of Research in Health Sciences*, **24**(4), e00633.
- Stop T. Partnership (2022), *Global Plan to End TB 2023–2030*, United Nations Office for Project Services.
- Huisman J. S., Scire J., Angst D. C., Li J., Neher R. A., Maathuis M. H., et al. (2022), Estimation and worldwide monitoring of the effective reproductive number of SARS-CoV-2, *Elife*, **11**, e71345.
- Rastegar M., Nazar E., Nasehi M., Sharafi S., Fakoor V., Shakeri M. T. (2024), Bayesian estimation of the time-varying reproduction number for pulmonary tuberculosis in Iran: a registry-based study from 2018 to 2022 using new smear-positive cases, *Infectious Disease Modelling*, **9**(3), 963-974.
- Gostic K. M., McGough L., Baskerville E. B., Abbott S., Joshi K., Tedijanto C., et al. (2020), Practical considerations for measuring the effective reproductive number, R_t , *PLoS Computational Biology*, **16**(12), e1008409.
- Fraser C. (2007), Estimating individual and household reproduction numbers in an emerging epidemic, *PloS One*, **2**(8), e758.
- Cowling B. J., Fang V. J., Riley S., Peiris J. M., Leung G. M. (2009), Estimation of the serial interval of influenza, *Epidemiology (Cambridge, Mass)*, **20**(3), 344.
- Firth D. (1988), Multiplicative errors: log-normal or gamma?, *Journal of the Royal Statistical Society: Series B (Methodological)*, **50**(2), 266-268.
- Cori A., Cauchemez S., Ferguson N. M., Fraser C., Dahlgvist E., Demarsh P. A., et al. (2020), Package ‘EpiEstim’, CRAN: Vienna Austria.

- Reich N. G., Lessler J., Cummings D. A., Brookmeyer R. (2009), Estimating incubation period distributions with coarse data, *Statistics in Medicine*, **28**(22), 2769-2784.
- Cori A., Ferguson N. M., Fraser C., Cauchemez S. (2013), A new framework and software to estimate time-varying reproduction numbers during epidemics, *American Journal of Epidemiology*, **178**(9), 1505-1512.
- Thompson R. N., Stockwin J. E., van Gaalen R. D., Polonsky J. A., Kamvar Z. N., Demarsh P. A., et al. (2019), Improved inference of time-varying reproduction numbers during infectious disease outbreaks, *Epidemics*, **29**, 100356.
- Nishiura H., Chowell G. (2014), Early transmission dynamics of Ebola virus disease (EVD), West Africa, March to August 2014, *Eurosurveillance*, **19**(36), 20894.
- McBryde E., Bergeri I., van Gemert C., Rotty J., Headley E., Simpson K., et al. (2009), Early transmission characteristics of influenza A (H1N1) v in Australia: Victorian state, 16 May–3 June 2009, *Eurosurveillance*, **14**(42).
- Wallinga J., Lipsitch M. (2007), How generation intervals shape the relationship between growth rates and reproductive numbers, *Proceedings of the Royal Society B: Biological Sciences*, **274**(1609), 599-604.
- Organization WH (2021), *WHO global lists of high burden countries for tuberculosis (TB), TB/HIV and multidrug/rifampicin-resistant TB (MDR/RR-TB), 2021–2025: background document*.
- Kimbrough W., Saliba V., Dahab M., Haskew C., Checchi F. (2012), The burden of tuberculosis in crisis-affected populations: a systematic review, *The Lancet Infectious Diseases*, **12**(12), 950-965.
- Cori A., Donnelly C. A., Dorigatti I., Ferguson N. M., Fraser C., Garske T., et al. (2017), Key data for outbreak evaluation: building on the Ebola experience, *Philosophical Transactions of the Royal Society B: Biological Sciences*, **372**(1721), 20160371.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



European Option Pricing under Market Uncertainty and Liquidity Risk: A Heston-Based Stochastic Volatility Model

Mahdi Rezaei Bahrmand*

Department of Systems and Control, K.N. Toosi University of Technology, Tehran, 16317-14191, Iran

Abstract

This paper addresses the study of European options in markets where the underlying asset is subject to liquidity-related risks. In the absence of liquidity effects, the asset price dynamic is governed by the Heston stochastic volatility model. Liquidity risk is introduced through market liquidity as a stochastic process, which adjusts the asset price via a liquidity discount mechanism. By applying a change of measure, we derive the dynamics of the model under the risk-neutral framework. An explicit form of the characteristic function for the logarithm of the asset price is then obtained, enabling the derivation of a closed-form solution for European option prices. The numerical example and analysis of the sensitivity of key parameters confirm the theoretical results.

Keywords: Option Pricing, Heston Model, Liquidity.

Mathematics Subject Classification (2020): 65CXX, 60HXX.

1 Introduction

Following the foundational work of [Black and Scholes \(1973\)](#), which modeled asset prices using normally distributed log-returns, subsequent models have sought to address more realistic market dynamics. These advancements introduced multiple stochastic factors, such as classical stochastic volatility frameworks [Heston \(1993\)](#) and [Stein and Stein \(1991\)](#), regime-switching dynamics [He and Lin \(2023\)](#), multifactor volatility specifications [Cortazar and *et al.* \(2009\)](#) and [El-Khatib and *et al.* \(2024\)](#), and hybrid models incorporating both stochastic volatility and interest rates [Grzelak and Oosterlee \(2011\)](#). Among these, the Heston model [Heston \(1993\)](#) has remained widely adopted due to its analytical tractability and its capacity to capture essential volatility features like mean reversion and non-negativity. Although many existing models offer significant strengths, they often overlook the essential role of liquidity risk—an aspect widely recognized as critical in real-world financial risk management. Incorporating liquidity considerations into derivative pricing frameworks remains a necessary yet underdeveloped area. Despite its practical importance to both market practitioners and risk

*Corresponding author, Rezaeimahdi2018@gmail.com

professionals, research on this topic is still emerging, with no broadly accepted framework established to date. Liquidity-aware option pricing models vary [Li and *et al.* \(2018\)](#) and [Loeper \(2018\)](#), with some focusing on trader impact [Leippold and Schärer \(2017\)](#) and [Ludkovski and Shen \(2013\)](#), others using bid-ask spreads [Zhang and *et al.* \(2019\)](#), and some applying price discounts based on trading speed during illiquid periods [Subramanian and Jarrow \(2001\)](#) and [Longstaff and *et al.* \(2005\)](#). Growing empirical evidence highlights the importance of market liquidity in shaping asset prices, motivating its integration into stock price dynamics [Feng and *et al.* \(2016\)](#). [Madan and Cherny \(2010\)](#) introduced a framework treating the market as a counterparty and modeling liquidity through a stress parameter, later refined to better capture its effect on returns [Albrecher and Schoutens \(2013\)](#) and [Corcuera and *et al.* \(2010\)](#). This led to derivative pricing models that discount stock prices based on liquidity conditions [Caldarera and *et al.* \(2005\)](#). [Feng and *et al.* \(2014\)](#) advanced this by introducing stochastic market liquidity, which has since been applied to pricing various options, including European [Puneet and *et al.* \(2022\)](#), American [Zhang and *et al.* \(2019\)](#), and Asian options [Li and *et al.* \(2018\)](#). To address the limitations of prior models that rely on the Black-Scholes framework, we propose an option pricing model combining Heston's stochastic volatility with a liquidity discount factor. Despite involving four stochastic drivers, we derive a closed-form solution for European options using a benchmark transformation. Numerical results highlight the effects of both liquidity and volatility risks, supported by sensitivity analysis.

The structure of the paper is as follows. Section 2 introduces the proposed option pricing framework under the real-world measure. In Section 3, we apply a transforming the physical measure into an equivalent martingale framework and present the pricing closed-form formula. Section 4 offers numerical results, and final remarks are provided in the concluding section.

2 Option pricing based on the Heston volatility model

We consider a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and adopt the Heston stochastic volatility model given by:

$$\begin{aligned} \frac{dP_t^H}{P_t^H} &= \mu dt + \sqrt{v_t} dB_{1,t}, \\ dv_t &= \eta(\theta - v_t) dt + \sigma \sqrt{v_t} dB_{2,t} \end{aligned} \quad (1)$$

where P_t^H denotes the stock price, μ is the expected return, v_t represents the instantaneous variance, θ is the long-run mean of the variance, η is the rate of mean reversion, σ is the volatility of volatility, and $\{B_{i,t}\}_{t \geq 0}$ for $i = 1, 2$ are standard Brownian motions with $dB_{1,t}dB_{2,t} = \rho_1 dt$. It is important to emphasize that the dynamics described in equation (1) are not directly suitable for derivative pricing, as real-world stock markets often experience imbalances in supply and demand, resulting in liquidity issues. Therefore, accurately modeling liquidity risk within stock dynamics demands careful and deliberate treatment.

Following [Black and Scholes \(1973\)](#) and [Feng and *et al.* \(2014\)](#), we model liquidity risk through supply-demand interaction, assuming a constant stock supply \bar{P} , which adjusts prices in response to market demand fluctuations. The demand is modeled by the function $d(P_t, \delta_t, \epsilon_t)$, defined as

$$d(P_t, \delta_t, \epsilon_t) = \left(\frac{P_t \epsilon_t}{\omega \delta_t} \right) \Upsilon, \quad (2)$$

where $Z(\cdot)$ is a smooth, strictly increasing function that captures how demand responds to price, liquidity sensitivity δ_t , and an external factor ϵ_t . In this framework, the observed stock price P_t incorporates liquidity risk by discounting the Heston price P_t^H via a stochastic discount factor ϵ_t :

$$P_t = \frac{1}{\epsilon_t} P_t^H. \quad (3)$$

The liquidity-adjusted price is derived from market equilibrium:

$$\left(\frac{\delta_t \omega}{\epsilon_t P_t} \right) \Upsilon = \bar{P}, \quad \Rightarrow \quad P_t = \frac{1}{\epsilon_t} \cdot \frac{\delta_t \omega}{\Upsilon^{-1}(\bar{P})}. \quad (4)$$

When liquidity is perfect ($\epsilon_t = 1$), the price reduces to:

$$P_t^H = \frac{\delta_t \omega}{Z^{-1}(\bar{P})}. \quad (5)$$

The liquidity discount factor ϵ_t evolves as a geometric Brownian motion:

$$\frac{d\epsilon_t}{\epsilon_t} = \left(\frac{1}{2} \lambda^2 M_t^2 - \lambda M_t \right) dt - \lambda M_t dB_{3,t}, \quad (6)$$

with its explicit solution:

$$\epsilon_t = \exp \left(-\lambda \int_0^t m_s ds + \lambda \int_0^t m_s dB_{3,s} \right). \quad (7)$$

Applying Itô's product rule gives the dynamics of the liquidity-adjusted stock price:

$$\frac{dP_t}{P_t} = \left(\mu + \lambda M_t + \frac{1}{2} \lambda^2 M_t^2 \right) dt + \sqrt{v_t} dB_{1,t} + \lambda M_t dB_{3,t}. \quad (8)$$

The market liquidity M_t follows an Ornstein–Uhlenbeck process:

$$dM_t = \alpha(\beta - M_t) dt + \delta dB_{4,t}, \quad (9)$$

where $B_{4,t}$ is a Brownian motion with $dB_{1,t}dB_{4,t} = 0$, $dB_{2,t}dB_{4,t} = 0$, and $dB_{3,t}dB_{4,t} = \rho_2$. Decoupling liquidity and price dynamics simplifies analysis and remains financially valid, despite reduced model flexibility from not using full correlation.

This model captures the dual impact of stochastic volatility and liquidity risk, forming a two-factor pricing framework.

3 European option pricing

To enable European option pricing within the liquidity-adjusted Heston model, an equivalent martingale measure must first be established, as the current model is defined under the physical measure.

3.1 A risk-adjusted probability measure

In a market where both stochastic volatility and liquidity risk are involved, the system is inherently incomplete due to the presence of multiple sources of uncertainty. Consequently, there is no unique risk-neutral measure, and one must be selected under suitable conditions.

To handle correlations among the driving Brownian motions, the stock price process is reformulated as:

$$\frac{dP_t}{P_t} = \left(\mu + \lambda M_t + \frac{1}{2} \lambda^2 M_t^2 \right) dt + \sqrt{v_t} (\rho_1 dB_{2,t} + \sqrt{1 - \rho_1^2} dB_{5,t}) + \lambda M_t (\rho_2 dB_{4,t} + \sqrt{1 - \rho_2^2} dB_{6,t}), \quad (10)$$

where $B_{2,t}, B_{4,t}, B_{5,t}, B_{6,t}$ are independent Brownian motions. We define the Radon-Nikodym derivative for transitioning from the physical measure \mathbb{P} to a risk-neutral measure \mathbb{Q} :

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp \left(- \sum_{i=1}^4 \int_0^t \eta_{i,s} dB_{i,s} - \frac{1}{2} \int_0^t \eta_{i,s}^2 ds \right), \quad (11)$$

which yields the following transformed Brownian motions under \mathbb{Q} :

$$\begin{aligned} dB_{2,t}^{\mathbb{Q}} &= dB_{2,t} + \eta_{1,t} dt, \\ dB_{4,t}^{\mathbb{Q}} &= dB_{4,t} + \eta_{2,t} dt, \\ dB_{5,t}^{\mathbb{Q}} &= dB_{5,t} + \eta_{3,t} dt, \\ dB_{6,t}^{\mathbb{Q}} &= dB_{6,t} + \eta_{4,t} dt. \end{aligned} \quad (12)$$

Assuming that the market prices of volatility and liquidity risks are proportional to their respective levels, we let:

$$\sigma\sqrt{v_t}\eta_{1,t} = \eta_1 v_t, \quad \delta\eta_{2,t} = \eta_2 M_t. \quad (13)$$

This leads to the transformed dynamics:

$$\begin{aligned} dv_t &= \bar{\eta}(\bar{\theta} - v_t)dt + \sigma\sqrt{v_t}dB_{2,t}^{\mathbb{Q}}, \\ dM_t &= \bar{\alpha}(\bar{\beta} - M_t)dt + \delta dB_{4,t}^{\mathbb{Q}}, \end{aligned} \quad (14)$$

where the adjusted parameters are:

$$\bar{\eta} = \eta + \eta_1, \quad \bar{\theta} = \frac{\eta\theta}{\eta + \eta_1}, \quad \bar{\alpha} = \alpha + \eta_2, \quad \bar{\beta} = \frac{\alpha\beta}{\alpha + \eta_2}. \quad (15)$$

Substituting the transformed Brownian motions into the stock price process gives:

$$\frac{dP_t}{P_t} = \bar{\mu}dt + \sqrt{v_t}(\rho_1 dB_{2,t}^{\mathbb{Q}} + \sqrt{1 - \rho_1^2} dB_{5,t}^{\mathbb{Q}}) + \lambda M_t(\rho_2 dB_{4,t}^{\mathbb{Q}} + \sqrt{1 - \rho_2^2} dB_{6,t}^{\mathbb{Q}}), \quad (16)$$

where

$$\bar{\mu} = \mu + \lambda M_t + \frac{1}{2}\lambda^2 M_t^2 - \sqrt{v_t}(\rho_1 \eta_{1,t} + \sqrt{1 - \rho_1^2} \eta_{3,t}) - \lambda M_t(\rho_2 \eta_{2,t} + \sqrt{1 - \rho_2^2} \eta_{4,t}). \quad (17)$$

By enforcing the no-arbitrage condition under \mathbb{Q} (i.e., requiring that the discounted stock price is a martingale), we must have $\bar{\mu} = r$, the risk-free rate. Solving for this yields:

$$\mu + \lambda M_t + \frac{1}{2}\lambda^2 M_t^2 - \sqrt{v_t}(\rho_1 \eta_{1,t} + \sqrt{1 - \rho_1^2} \eta_{3,t}) - \lambda M_t(\rho_2 \eta_{2,t} + \sqrt{1 - \rho_2^2} \eta_{4,t}) = r. \quad (18)$$

Under the risk-neutral measure, the final form of the stock price SDE becomes:

$$\frac{dP_t}{P_t} = rdt + \sqrt{v_t}dB_{1,t}^{\mathbb{Q}} + \lambda M_t dB_{3,t}^{\mathbb{Q}}, \quad (19)$$

where the volatility and liquidity processes evolve according to the dynamics in Equation (4).

3.2 Closed-form formula

With the risk-neutral dynamics established, we now address the pricing of European options where both stochastic volatility and liquidity effects are present. Let $C(P_t, M_t, v_t, t)$ denote the option value at time t , with $\mathbb{F}_t^{\mathbb{Q}}$ representing the filtration generated by Brownian motions under the risk-neutral measure \mathbb{Q} . The pricing formula becomes:

$$C(P_t, M_t, v_t, t) = \mathbb{E}^{\mathbb{Q}} \left[e^{-r(T-t)} (P_t - K)^+ \mid \mathcal{F}^{B_2^{\mathbb{Q}}} \vee \mathcal{F}^{B_4^{\mathbb{Q}}} \vee \mathcal{F}^{B_5^{\mathbb{Q}}} \vee \mathcal{F}^{B_6^{\mathbb{Q}}} \right] = e^{-r(T-t)} (J_1 - K J_2) \quad (20)$$

Define $x_t = \ln P_t$, then:

$$J_1 = \int_{\ln K}^{\infty} e^{x_T} l(x_T) dx_T, \quad J_2 = \int_{\ln K}^{\infty} l(x_T) dx_T, \quad (21)$$

where $l(x_T)$ is the conditional density function of x_T . We utilize the characteristic function

$$f(\phi) \equiv f(\phi; x_t, M_t, v_t, t, T) = \mathbb{E}^{\mathbb{Q}} [e^{j\phi x_T} \mid \mathcal{F}^{B_2^{\mathbb{Q}}} \vee \mathcal{F}^{B_4^{\mathbb{Q}}} \vee \mathcal{F}^{B_5^{\mathbb{Q}}} \vee \mathcal{F}^{B_6^{\mathbb{Q}}}] = \int_{-\infty}^{+\infty} e^{j\phi x_T l(x_T)} dx_T \quad (22)$$

to express:

$$J_1 = f(-j) \left(\frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \Re \left(\frac{e^{-j\phi \ln K}}{j\phi} \frac{f(\phi - j)}{f(-j)} \right) d\phi \right), \quad (23)$$

$$J_2 = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \Re \left(\frac{e^{-j\phi \ln K}}{j\phi} f(\phi) \right) d\phi. \quad (24)$$

The characteristic function satisfies the PDE:

$$\frac{\partial f}{\partial t} + \frac{1}{2}(\lambda^2 m^2 + v) \frac{\partial^2 f}{\partial x^2} + \frac{1}{2}\sigma^2 v \frac{\partial^2 f}{\partial v^2} + \frac{1}{2}\delta^2 \frac{\partial^2 f}{\partial m^2} + \rho_1 \sigma v \frac{\partial^2 f}{\partial x \partial v} + \rho_2 \lambda \delta m \frac{\partial^2 f}{\partial x \partial m}$$

$$+ \left(r - \frac{1}{2}(\lambda^2 m^2 + v) \right) \frac{\partial f}{\partial x} + \bar{\eta}(\bar{\theta} - v) \frac{\partial f}{\partial v} + \bar{\alpha}(\bar{\beta} - m) \frac{\partial f}{\partial m} = 0, \quad (25)$$

with terminal condition: $f(\phi; x_T, M_t, v_T, T) = e^{j\phi x_T}$. Assuming an affine form:

$$f(\phi) = \exp(I_1(\phi) + I_2(\phi)m + I_3(\phi)m^2 + I_4(\phi)v + j\phi x), \quad (26)$$

leads to ODEs for I_1, I_2, I_3, I_4 , with closed-form Riccati solutions for I_3 and I_4 :

$$I_3 = \frac{A_1 - (2j\delta\lambda\rho_2\phi - 2\bar{\alpha})}{4\delta^2} \cdot \frac{1 - e^{A_1\tau}}{1 - B_1e^{A_1\tau}}, \quad (27)$$

$$I_4 = \frac{A_2 - (j\sigma\rho_1\phi - \bar{\eta})}{\sigma^2} \cdot \frac{1 - e^{A_2\tau}}{1 - A_2e^{A_2\tau}}. \quad (28)$$

Given I_3 , we solve:

$$I_2 = -\frac{\bar{\alpha}\bar{\beta}}{\delta^2} (A_1 - 2(j\delta\lambda\rho_2\phi - \bar{\alpha})) \frac{(1 - e^{\frac{1}{2}A_1\tau})^2}{1 - B_1e^{A_1\tau}}. \quad (29)$$

Finally, I_1 is integrated explicitly:

$$I_1 = \frac{1}{4} \{ (A_1 - 2(j\delta\lambda\rho_2\phi - \bar{\alpha}))\tau - 2\ln\left(\frac{1 - B_1e^{A_1\tau}}{1 - B_1}\right) \} + \frac{\bar{\eta}\bar{\theta}}{\sigma^2} \{ [A_2 - (j\sigma\rho_1\phi - \bar{\eta})]\tau - 2\ln\left(\frac{1 - B_2e^{A_2\tau}}{1 - B_2}\right) \} + \int_0^\tau \left(\frac{1}{2}\delta^2 I_2^2(\varphi; s) + \bar{\alpha}\bar{\beta} I_2(\varphi; s) \right) ds + jr\tau\varphi. \quad (30)$$

4 Numerical Validation and Sensitivity Analysis

To validate the analytical pricing formula, we first benchmark it against Monte Carlo simulations. This comparison confirms the accuracy of the formula and enables further investigation

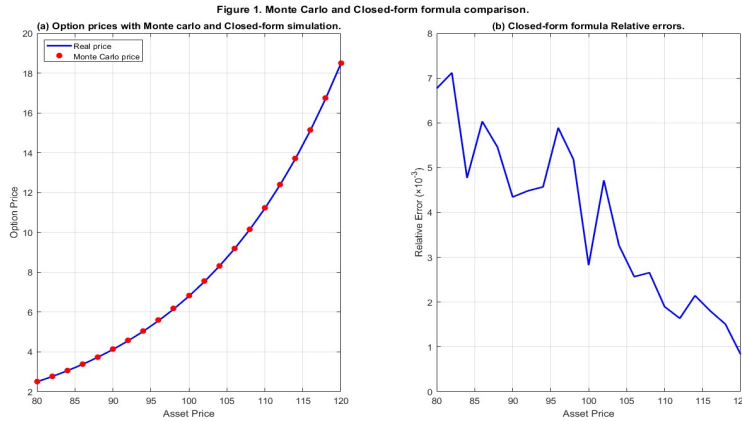


Figure 1: Write the caption here.

into the effects of volatility and liquidity risks. Unless otherwise specified, the following parameters are used for numerical evaluation:

$$\bar{\eta} = 7, \bar{\theta} = 0.06, \sigma = 0.17, \bar{\alpha} = 3, \bar{\beta} = 0.3, \delta = 0.1, \lambda = 0.6, \rho_1 = -0.6,$$

$$\rho_2 = -0.3, r = 0.02, \tau = 1, P_0 = 100, K = 100, M_0 = 0.1, v_0 = 0.03.$$

Figure 1 demonstrates strong agreement between the option prices computed via the closed-form formula and those obtained from simulations. In particular, Figure 1 (a) shows pointwise consistency, and Figure 1 (b) confirms a maximum relative error below 0.6%, establishing the formula's precision. Next, we examine how the liquidity sensitivity parameter λ influences option prices. Three model variations are compared:

- **CV Model:** Ignores stochastic volatility ($\bar{\eta} = \sigma = 0$); based on [Feng and et al. \(2014\)](#).
- **CL Model:** Omits stochastic liquidity ($\bar{\alpha} = \delta = 0$).
- **Full Model:** Incorporates both stochastic volatility and liquidity.

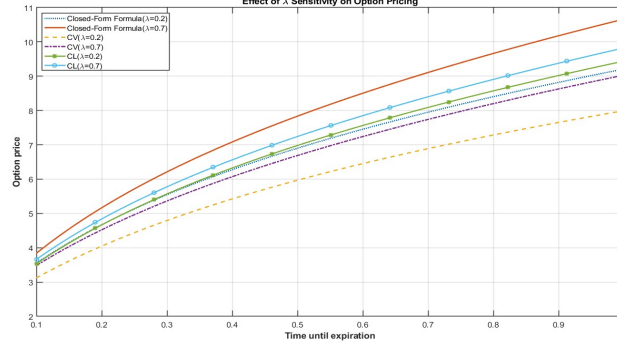


Figure 2: Write the caption here.

Results in Figure 2 show that increasing λ leads to higher option prices across all models. This is, expected as greater sensitivity to market liquidity implies a stronger discounting effect, raising option value. Moreover, both stochastic volatility and stochastic liquidity are shown to have a significant impact on pricing, motivating a deeper analysis of their combined effects. This gives a fully analytical form for the characteristic function, enabling efficient computation of the European option price under both volatility and liquidity risks.

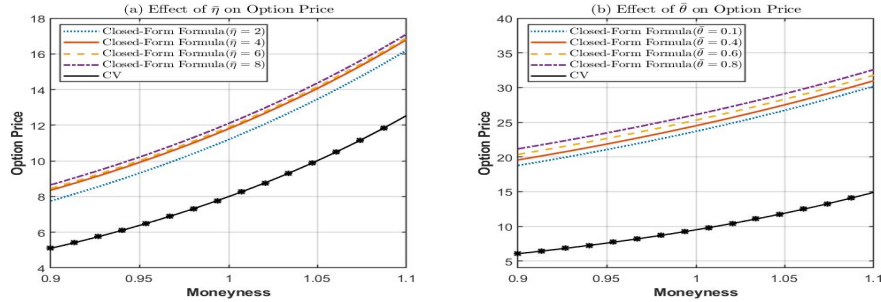


Figure 3: Write the caption here.

To evaluate the role of stochastic volatility, we benchmark against the CV model ([Feng and et al. , 2014](#)). Figure 3 (a) shows that option prices increase with the mean reversion speed $\bar{\eta}$, due to faster convergence toward a higher long-term volatility level $\bar{\theta}$. However, this effect tapers off as $\bar{\eta}$ becomes large. Conversely, if $\bar{\theta}$ were below the current variance, the trend would reverse. Figure 3 (b) confirms that raising $\bar{\theta}$ also elevates option values by amplifying expected volatility. Across both cases, the CV model yields the lowest prices, given its constant and relatively low volatility assumption, though this could change with different parameter settings. Figure 4 (a) illustrates how option prices respond to changes in the volatility of volatility parameter σ . A noteworthy pattern emerges: when the option is deep out-of-the-money (low moneyness), increasing σ tends to reduce its value. In contrast, for deep in-the-money options (high moneyness), greater σ leads to higher option prices. This asymmetry can be explained by exercise probability: high volatility decreases the chance of realization for out-of-the-money options while potentially amplifying gains for in-the-money options. To assess the role of stochastic liquidity, we compare our full model with the CL model under varying liquidity parameters. Figure 4 (b) shows that option prices rise with the

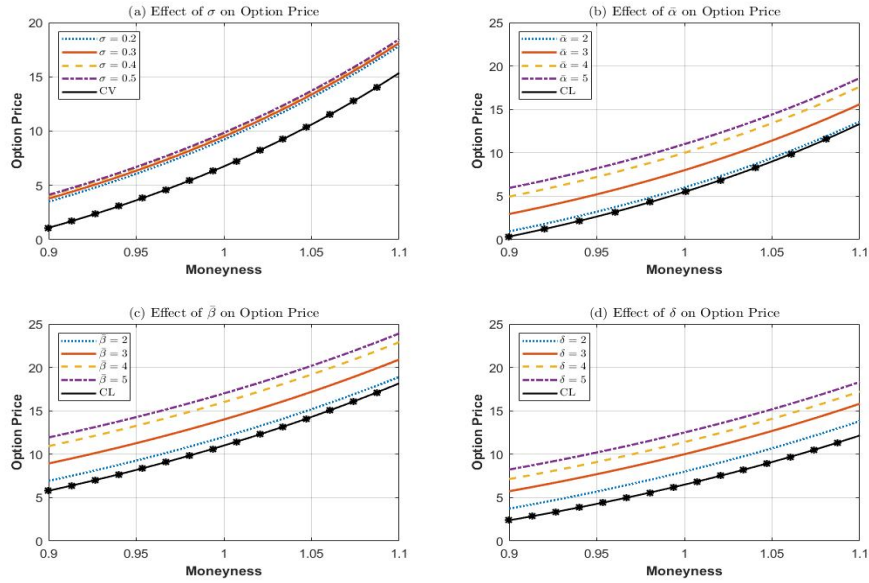


Figure 4: Write the caption here.

mean-reversion speed $\bar{\alpha}$, since faster convergence to a higher long-term illiquidity level increases perceived liquidity risk, thereby raising the option premium. Figure 4 (c) examines the impact of the equilibrium liquidity level $\bar{\beta}$, where a clear upward trend in option prices is observed. Higher $\bar{\beta}$ reflects more persistent illiquidity, leading to greater pricing impact. The CL model, which assumes a lower fixed liquidity level, consistently yields lower option prices. This pattern could reverse if the CL model were set with a higher illiquidity baseline. Finally, Figure 4 (d) evaluates how volatility in market liquidity affects pricing. As liquidity volatility rises, option prices increase. This is intuitive: higher uncertainty in liquidity conditions discourages option sellers from offering contracts at lower prices due to elevated exposure to liquidity shocks.

Discussion and Results

In response to the demand for more realistic financial market modeling, this paper introduces a novel option pricing framework that incorporates both stochastic volatility and liquidity risks. The model extends the classical Heston stochastic volatility framework by introducing a liquidity discounting mechanism that captures stochastic market liquidity conditions. A risk-neutral measure is constructed via benchmark transformation to facilitate derivative pricing. Within this framework, we derive a closed-form solution for European option prices. Numerical experiments validate the model's effectiveness and demonstrate its capability to reflect key empirical characteristics of financial markets, particularly those related to volatility and liquidity dynamics.

References

- Albrecher, H. and Schoutens, W. (2013). Implied liquidity: Model sensitivity. *Journal of Empirical Finance*, **23**, 48–67.
- Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**, 637–654.
- Caldarera, A. and Brunetti, C. (2005). Asset prices and asset correlations in illiquid markets. Available at SSRN 625184.

- Cortazar, G., Lopez, M., and Naranjo, L. (2017). A multifactor stochastic volatility model of commodity prices. *Energy Economics*, **67**, 182–201.
- Corcuera, J. M., Guillaume, F., Madan, D. B., and Schoutens, W. (2010). Implied liquidity: Towards stochastic liquidity modeling and liquidity trading. *International Journal of Portfolio Analysis and Management*, 80–91.
- Feng, S.-P., Hung, M.-W., and Wang, Y.-H. (2014). Option pricing with stochastic liquidity risk: Theory and evidence. *Journal of Financial Markets*, **18**, 77–95.
- Grzelak, L. A. and Oosterlee, C. W. (2011). On the Heston model with stochastic interest rates. *SIAM Journal on Financial Mathematics*, **2**, 255–286.
- He, X.-J. and Lin, S. (2023). Analytically pricing European options under a hybrid stochastic volatility and interest rate model with a general correlation structure. *Journal of Futures Markets*, **43**, 951–967.
- He, X.-J. and Lin, S. (2024). Analytically pricing foreign exchange options under a three-factor stochastic volatility and interest rate model: A full correlation structure. *Expert Systems with Applications*, **246**, 123203.
- Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, **6**, 327–343.
- Hu, Z., Yang, B.-Z., He, X.-J., and Yue, J. (2024). Equilibrium pricing of European crude oil options with stochastic behavior and jump risks. *Mathematics and Computers in Simulation*, **219**, 212–230.
- Feng, S. P., Hung, M. W., and Wang, Y. H. (2016). The importance of stock liquidity on option pricing. *International Review of Economics and Finance*, **43**, 457–467.
- Leippold, M. and Schärer, S. (2017). Discrete-time option pricing with stochastic liquidity. *Journal of Banking and Finance*, **75**, 1–16.
- Li, Z., Zhang, W.-G., and Liu, Y.-J. (2018). Analytical valuation for geometric Asian options in illiquid markets. *Physica A*, **507**, 175–191.
- Loeper, G. (2018). Option pricing with linear market impact and nonlinear Black-Scholes equations. *Annals of Applied Probability*, **28**, 2664–2726.
- Longstaff, F. A., Mithal, S., and Neis, E. (2005). Corporate yield spreads: Default risk or liquidity? New evidence from the credit default swap market. *Journal of Finance*, **60**, 2213–2253.
- Ludkovski, M. and Shen, Q. (2013). European option pricing with liquidity shocks. *International Journal of Theoretical and Applied Finance*, **16**, 1350043.
- Madan, D. B. and Cherny, A. (2010). Markets as a counterparty: An introduction to conic finance. *International Journal of Theoretical and Applied Finance*, **13**, 1149–1177.
- Puneet, P., Song-Ping, Z., and Xin-Jiang, H. (2022). A closed-form pricing formula for European options in an illiquid asset market. *Financial Innovation*, **8**, 1–18.
- El-Khatib, Y., Makumbe, Z. S., and Vives, J. (2024). Approximate option pricing under a two-factor Heston–Kou stochastic volatility model. *Computational Management Science*, **21(1)**, 3.
- Stein, E. M. and Stein, J. C. (1991). Stock price distributions with stochastic volatility: An analytic approach. *Review of Financial Studies*, **4**, 727–752.
- Subramanian, A. and Jarrow, R. A. (2001). The liquidity discount. *Mathematical Finance*, **11**, 447–474.
- Zhang, Y., Ding, S., and Duygun, M. (2019). Derivatives pricing with liquidity risk. *Journal of Futures Markets*, **39**, 1471–1485.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Markovian Queueing Systems and Application in Real Data

Mehdi Shams^{*1}, Mohammad Ali Mirzaie²

¹Department of mathematics, University of Kashan, Kashan, Iran

²Department of mathematics, University of Kashan, Kashan, Iran

Abstract

A probabilistic model is an approach to modeling a situation where we want to know what will happen in the future or to gather information from our systems. One branch of the probabilistic model is queueing theory. In this paper, we first examine and analyze stochastic processes, especially the Poisson process and its simulation, and the simple Markovian queueing systems with the M/M/c model. Finally, we perform a queue modeling with **R** software for a series of real data.

Keywords: Poisson process, stochastic processes, queueing theory, simulation.

Mathematics Subject Classification (2020): 60G05, 60K25, 60G55.

1 Introduction

Stochastic processes have wide applications in most branches of science. Queueing theory is a state modeling approach in which customers enter the system with the main process and exit the system after receiving service from the servers. Markovian queues are useful for interpreting some queues in which customers enter the system with the Poisson process and receive service from the servers when it follows an exponential distribution.

In Szavits-Nossan and Grima (2024), solving stochastic gene expression models using queueing theory is described. M/M/c queueing systems with requeueing and service failure and repair are reviewed in Subramanian (2011). In studies such as Dai *et al.* (2024); Hassan *et al.* (2024), the applications of queueing models in call centers are examined.

In this paper, Markovian queue and the basic theory of the stochastic process are reviewed and finally data modeling of a call center with the software **R** is investigated. Finally, statistical inferences about the arrival and answer parameters of the processes are also examined. The section 2 reviews the necessary prerequisites and the section 3 is the authors' research, which analyzes and models queue data in a case report.

^{*}Corresponding author, mehdishams@kashanu.ac.ir

2 stochastic process and queueing theory

The set $\{X_t : t \in T\}$ with index set T and support E has been called stochastic process. Also, the set $X_t - X_s$, $t > s$ is an increment. For the stochastic process $\{X_t : t \geq 0\}$, we say the increments are independent, if $X_0, X_{t_1} - X_0, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ are independent random variables. Also, we say the increments are stationary, if for all s, t where $s < t$, the distribution of increment only depend on distance between s and t .

Definition 2.1. (Ross, 1996) *The stochastic process $\{N_t : t \geq 0\}$ is called the Poisson process, if we have these conditions*

1. $N_0 = 0$;
2. $\{N_t : t \geq 0\}$ with independent increment;
3. $\{N_t : t \geq 0\}$ with stationary increment;
4. The increments $N_t - N_s$, $0 \leq s < t$ has the Poisson distribution, i.e. $N_t - N_s \sim P(\lambda(t-s))$ with following probability mass function:

$$P(N_t - N_s = k) = \frac{e^{-\lambda(t-s)}(\lambda(t-s))^k}{k!}, \quad k = 0, 1, \dots$$

In the Poisson process $\{N_t : t \geq 0\}$, we have:

$$E(N_t) = \lambda t, \quad C(t, s) = \text{Cov}(N_t, N_s) = \begin{cases} \lambda \min\{t, s\}, & t \neq s, \\ \lambda t, & t = s. \end{cases}$$

Therefore, it can be said that the Poisson process is not a stationary process.

Another definition of the Poisson process is given in the following lemma, and the two definitions are equivalent.

Lemma 2.2. (Ross, 1996) *The stochastic process $\{N_t : t \geq 0\}$ is called the Poisson process, if the following conditions hold:*

1. $N_0 = 0$;
2. $\{N_t : t \geq 0\}$ with independent increment;
3. $\{N_t : t \geq 0\}$ with stationary increment;
4. For all $t > 0$ and $h > 0$:

$$P(N_{t+h} - N_t = 0) = 1 - \lambda h + o(h);$$

$$P(N_{t+h} - N_t = 1) = \lambda h + o(h);$$

$$P(N_{t+h} - N_t \geq 2) = o(h).$$

We accept the following theorem without proof.

Theorem 2.3. (Ross, 1996) *Let us consider $\{N_t : t \geq 0\}$ occurring by a Poisson process with rate λ . The times between two events follow independent identically distributed exponential random variables with mean $\frac{1}{\lambda}$ and probability density function $f(t) = \lambda e^{-\lambda t}$, $t > 0$.*

Let T_n , $n = 1, 2, \dots$ be a sequence of interarrival times. We define the arrival time of the n^{th} event or the waiting time until the n^{th} event as $S_n = \sum_{i=1}^n T_i$. According to Theorem 2.3, we can conclude that S_n has a gamma distribution, i.e., $S_n \sim \Gamma(n, \lambda)$ (Ross, 1996). So, we can simulate the Poisson process with following algorithm (Dobrow, 2016):

1. Let $S_0 = 0$;
2. Generate S_1, \dots, S_n from an exponential distribution with mean $\frac{1}{\lambda}$;
3. Calculate cumulative summation.

This algorithm can be shown in Figure 1.

In Figure 1, the simulation of the Poisson process with the mentioned algorithm is shown. Another algorithm to simulate the Poisson process is working with the Poisson distribution. From Definition 2.1, it follows that, the number of events in $(0, t)$ follows the Poisson distribution $N(t) \sim P(\lambda t)$. On the other hand, we have the following theorem.

Theorem 2.4. (Ross, 1996) Let $N(t) = n$. Then, the joint distribution of S_1, \dots, S_n have the same distribution as the order statistics corresponding to n independent random variables $U(0, t)$ with probability density function $f(x) = \frac{1}{t} I_{(0,t)}(x)$.

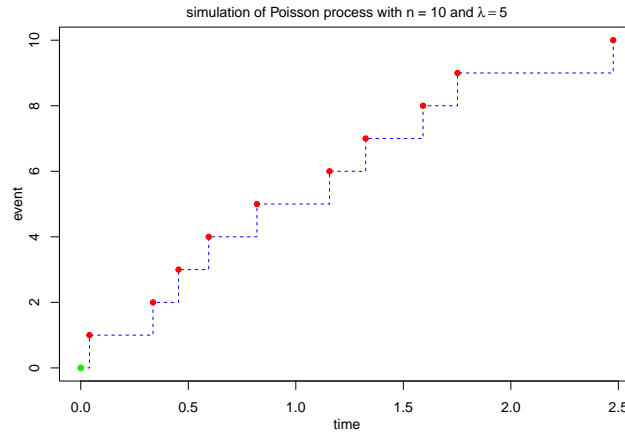


Figure 1: Simulate the Poisson process with $n = 10$ and $\lambda = 5$

The following algorithm is proposed to simulate the Poisson process based on Theorem 2.4 (Dobrow, 2016):

1. Generate a number from the Poisson distribution $P(\lambda t)$, namely n ;
2. Generate X_1, \dots, X_n from $U(0, t)$;
3. Sort the previous step.

In Figure 2, we simulate the Poisson process with second algorithm.

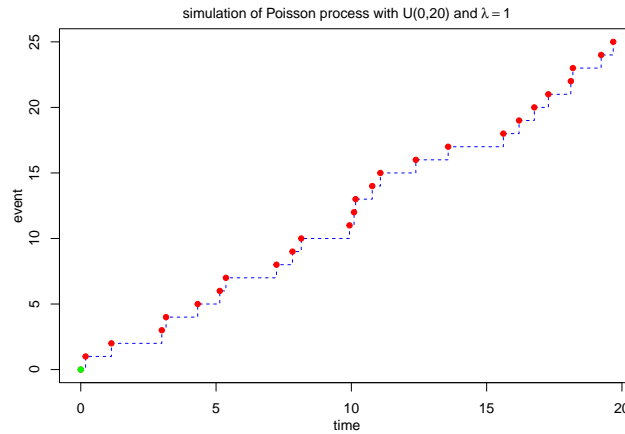


Figure 2: Simulate the Poisson process with $U(0, 20)$ and $\lambda = 1$

Consider a birth and death queueing system with c servers where customers enter the system according to a Poisson process and receive service according to an exponential distribution. We denote this system by $M/M/c$ and the queueing discipline is FIFO (First in First out). This means that any customer who logs in earlier will be served earlier. The $M/M/c$ is a birth and death queueing model with arrival rates $\lambda_n = \lambda$, $n \geq 0$ and departure rates

$$\mu_n = \begin{cases} n\mu, & \text{if } n \leq c \\ c\mu, & \text{if } n \geq c \end{cases}$$

Some performance measures are of particular importance in the queue model, which are mentioned below (Gross *et al.*, 2008):

The mean number of clients in the server (the expected number of busy servers) is equal to $E(B) = r = \frac{\lambda}{\mu}$, say the traffic intensity. The expected number of idle servers will be $E(I) = c - E(B) = c(1 - \rho)$.

$$p_n = \frac{(c\rho)^n}{n!} p_0, \quad n = 0, \dots, c$$

$$p_{c+n} = \rho^n p_c = \rho^n \frac{(c\rho)^c}{c!} p_0, \quad n = 0, 1, 2, \dots$$

where $p_0 = \left(\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(c\rho)^c}{c!} \cdot \frac{1}{1-\rho} \right)^{-1}$. The average number of customers in the system (queue) is $L = \frac{\lambda}{\mu} + p_0 \frac{\rho(c\rho)^c}{(1-\rho)^2 c!}$ ($L_q = p_0 \frac{\rho(c\rho)^c}{(1-\rho)^2 c!}$) and the average time in the system (queue) for a customer is $W = \frac{1}{\mu} + p_0 \frac{\rho(c\rho)^c}{\lambda(1-\rho)^2 c!}$ ($W_q = p_0 \frac{\rho(c\rho)^c}{\lambda(1-\rho)^2 c!}$). The probability of a queue forming in the system is equal to:

$$P[N \geq c] = \sum_{n=c}^{\infty} p_n = \frac{\frac{r^c}{c!}}{\frac{r^c}{c!} + (1-\rho) \sum_{j=0}^{c-1} \frac{r^j}{j!}}.$$

3 Application in real data modeling

In this section, we work with real data and show how to work with and model queue data. We use the data in Table 1.

Table 1: Data for call center with 4 servers in one year.

call_id	date	call_started	call_answered	call_ended	service_length
1	2021-01-01	8:00:00 AM	8:00:00 AM	8:14:22 AM	863
2	2021-01-01	8:02:42 AM	8:02:42 AM	8:07:31 AM	289
3	2021-01-01	8:08:24 AM	8:08:24 AM	8:10:13 AM	108
4	2021-01-01	8:09:37 AM	8:09:37 AM	8:13:45 AM	247
5	2021-01-01	8:11:10 AM	8:11:10 AM	8:15:28 AM	258
6	2021-01-01	8:22:33 AM	8:22:33 AM	8:28:13 AM	340
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
51708	2021-12-31	5:58:15 PM	5:58:15 PM	5:58:19 PM	4

This data were collected from a call centre with 4 servers in one year from 8 o'clock to 18 o'clock (a 10-hour shift work schedule). The variables considered include call started, call answered, call ended, and service length in second. At first, the service time (service-length) was examined, and it can be said that the service time follows an exponential distribution with a mean of approximately 300 seconds (5 minutes).

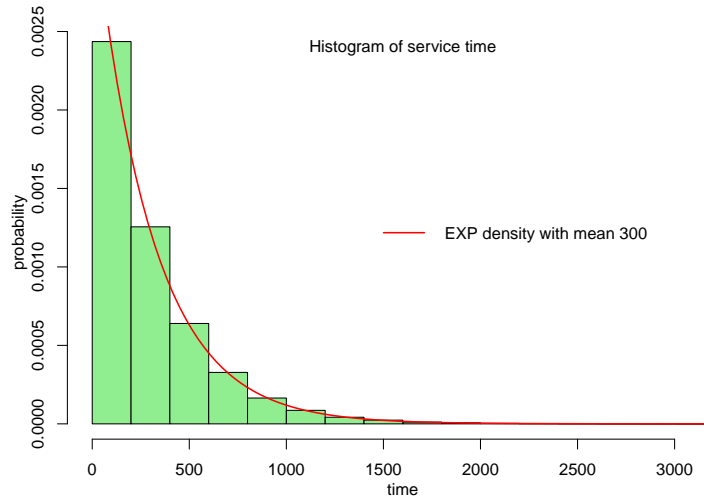


Figure 3: The histogram of the service time

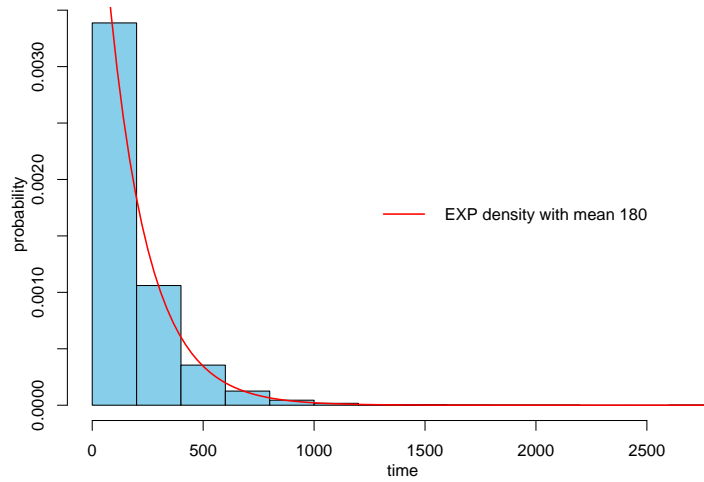


Figure 4: The histogram of times between each arrival

In Figure 3, the histogram of observations has an exponential density function. Also, the P-value of KS (Kolmogorov-Smirnov goodness of fit) test for service time is 0.7745, which means the service time follows an exponential distribution. At second step, the distribution of the entries into the system should be examined. The times between each calls was calculated and the histogram is plotted in Figure 4. The P-value of KS test is 0.1178, which means the arrival process is the Poisson process. From Figures 3 and 4, we can say that customers who calling this system follow a Poisson process, and therefore the distance between calls has an exponential distribution with mean $\frac{1}{\lambda} = 180$ seconds (3 minutes or 0.05 hours), i.e, $\lambda = 20$ customers per hour, thus we have $20 \times 10 = 200$ calls in a 10-hour shift work schedule, and take service from an exponential distribution with mean $\frac{1}{\mu} = 300$ seconds (5 minutes or 0.083 hours), i.e, $\mu = 12$ customers per hour, thus we have $12 \times 10 = 120$ persons in a 10-hour shift

work schedule, so we can say we have $M/M/4$ queue. In **R** software with package *queueing* we can calculate the performance measures of this queue.

In the following command, we observe performance measures of queueing systems with the $M/M/4$ model:

```
> library(queueing)
> mm4_input=NewInput.MMC(lambda=200,mu=120,c=4,n=4)
> mm4queue=QueueingModel.i_MMC(mm4_input)
> summary(mm4queue)
  lambda mu c k m      R0      P0      Lq
1    200 120 4 NA NA 0.4166667 0.1859321 0.07319701
      Wq  X      L      W      Wqq
1 0.0003659851 200 1.739864 0.008699318 0.003571429
      Lqq
1 1.714286
> Report(mm4queue)
The inputs of the model M/M/c are:
lambda: 200, mu: 120, c: 4, n: 4, method: Exact

The outputs of the model M/M/c are:

The probability (p0, p1, ..., pn) of the n = 4 clients
in the system are:
0.1859321 0.3098869 0.2582391 0.1434661 0.05977756
The traffic intensity is: 1.666666666666667
The server use is: 0.416666666666667
The mean number of clients in the system is: 1.7398636778863
The mean number of clients in the queue is: 0.0731970112196379
The mean number of clients in the server is: 1.666666666666667
The mean time spend in the system is: 0.00869931838943152
The mean time spend in the queue is: 0.00036598505609819
The mean time spend in the server is: 0.008333333333333333
The mean time spend in the queue when there is queue
is: 0.00357142857142857
The throughput is: 200
```

According to the information given in Table 2, relationship $p_0 = 0.18593$ shows that 18.593% of the time the system is empty and the servers have free time. Also, since $\rho = 0.41667 < 1$, so the system is in steady state. The mean number of clients in the system (queue) is $L = E(N) = 1.73986$ ($L_q = E(N_q) = 0.0732$), where random variables N and N_q represent the number of customers in the system and queue. The mean number of clients in the server (the expected number of busy servers or the traffic intensity) is equal to $E(B) = r = 1.66667$, the expected number of idle servers will be $E(I) = 2.33333$. The mean number of clients in the queue, when there is queue $L_{qq} = 1.71429$, the mean time spend in the system (queue) in a 10-hour shift work schedule is $W = 0.0087$ ($W_q = 0.00037$), the mean time spend in the queue in a 10-hour shift work schedule, when there is queue is $W_{qq} = 0.00357$. The probability of a

queue in the system is calculated as follows:

$$\begin{aligned}
P[N \geq 4] &= \sum_{n=4}^{\infty} p_n \\
&= 1 - p_0 - p_1 - p_2 - p_3 \\
&= 1 - 0.1859321 - 0.3098869 - 0.2582391 - 0.1434661 = 0.1024758.
\end{aligned}$$

Also, the variance of the number of customers in the system (queue) is equal to $Var(N) = 2.009866$ ($Var(N_q) = 0.1724064$), the variance of time spent in the system (queue) is equal to 7.192468×10^{-5} (2.480234×10^{-6}).

Table 2: Performance measures of queueing systems with the $M/M/4$ model

r	ρ	p_0	$P[N \geq 4]$	L
1.66667	0.41667	0.18593	0.10247	1.73986
L_q	L_{qq}	W	W_q	W_{qq}
0.0732	1.71429	0.0087	0.00037	0.00357

To test the hypothesis for the parameter λ , we propose two approaches. The first is a one-sample t-test based on the CLT (Central Limit Theorem) and the second is the GLRT (Generalized Likelihood Ratio Test). It is important to note that the power of the GLRT is greater than the power of the t-test, because in GLRT we also use the assumption that the data follow an exponential distribution.

Suppose $T_1, \dots, T_n \stackrel{iid}{\sim} f(t) = \lambda e^{-\lambda t}$. Note that, the GLRT based on asymptotic distribution $\chi^2_{(1)}$ at level $\alpha = 0.05$ for testing $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda \neq \lambda_0$ is given by

$$\phi(\mathbf{T}) = \begin{cases} 1, & -2n \ln \lambda_0 + 2\lambda_0 \sum_{i=1}^n T_i - 2n \ln(\bar{T}) - 2n > \chi^2_{(1)}(0.95) \\ 0, & \text{o.w.} \end{cases}$$

Note that the likelihood ratio statistic is equal to $\lambda(\mathbf{t}) = \frac{\sup_{\lambda_0} L(\lambda)}{\sup_{\lambda > 0} L(\lambda)} = \frac{\lambda_0^n e^{-n\lambda_0 \bar{t}}}{L(\hat{\lambda})}$ where $L(\lambda) = \lambda^n e^{-n\lambda \bar{t}}$ is a likelihood function, $\bar{t} = \frac{1}{n} \sum_{i=1}^n t_i$ and $\hat{\lambda} = \frac{1}{\bar{t}}$ is a MLE for λ . So, $\lambda(\mathbf{t}) = \frac{\lambda_0^n e^{-n\lambda_0 \bar{t}}}{\frac{1}{\bar{t}}^n} e^{-n}$. We reject H_0 whenever $-2 \ln \lambda(\mathbf{t}) > \chi^2_{(1)}(0.95)$ holds, in which case the test function is equal to $\phi(\mathbf{t}) = 1$.

To find the CI (confidence Interval) for $\frac{1}{\lambda}$ and $\frac{1}{\mu}$, we propose two approaches. The first is a CLT-based confidence interval with equal tails and the second is a confidence interval based on the exponential distribution. In Table 3, the hypothesis test is analyzed based on CLT and GLRT, also we generate confidence interval with CLT and direct approach. The first column includes the time between two input events. We consider $H_0 : \lambda = \frac{1}{180}$ against $H_1 : \lambda \neq \frac{1}{180}$ with two approaches. Also, we generate confidence interval based on CLT and direct method. In direct method, we generate $(1 - \alpha)\%$ confidence interval estimate with equal tails for $\frac{1}{\lambda}$ and $\frac{1}{\mu}$, i.e. $(\frac{2 \sum_{i=1}^n T_i}{b}, \frac{2 \sum_{i=1}^n T_i}{a})$ where $a = \chi^2_{(2n)}(\frac{\alpha}{2})$ and $b = \chi^2_{(2n)}(1 - \frac{\alpha}{2})$. The second column includes the time of service and $H_0 : \mu = \frac{1}{300}$ against $H_1 : \mu \neq \frac{1}{300}$ were analyzed. The confidence interval computed as same as arrival process.

Table 3: Statistical inference about parameters of arrival and answer processes

	arrival process	answer process
t-test	1.2147	-0.68052
P-value	0.2245	0.4962
GLRT	1.604769	0.4636167
P-value	0.2052287	0.4959386
Lower bound of the 95% CI based on CLT	179.3813	296.5179
Upper bound of the 95% CI based on CLT	182.6354	301.6873
Lower bound of the 95% CI based on direct method	179.4526	298.7445
Upper bound of the 95% CI based on direct method	182.5846	303.9586

According to the information given in Table 3, since all P-values are greater than 0.05, the null hypotheses based on CLT and GLRT approaches are accepted, meaning that the arrival process and answer process rates are $\lambda = 20$ and $\mu = 12$ customers per hour, respectively. Similarly (179.4526, 182.5846) and (298.7445, 303.9586) are 95% confidence interval with equal tails for $\frac{1}{\lambda}$ and $\frac{1}{\mu}$, respectively. Also, (179.3813, 182.6354) and (296.5179, 301.6873) are 95% asymptotic confidence interval for $\frac{1}{\lambda}$ and $\frac{1}{\mu}$, respectively.

References

- Dai T., Lu Y., Zivlak N., Lalic D.C. and Lalic B., (2024), How does queueing information in pre-sales call centers affect customer repurchase behavior, *Asia-Pacific Journal of Operational Research*, World Scientific Publishing Company, **41(04)**, 2440006.
- Dobrow R.P., (2016), *Introduction to stochastic processes with R*, John Wiley and Sons.
- Gross D., Shortle J.F., Thompson J.M. and Harris C.M. (2008), *Fundamentals of Queueing Theory. 4th Edition*, John Wiley and Sons Inc., New Jersey.
- Hassan N.A., Abdallah N.M.S. and Attwa R.A., (2024), Optimizing multi-skill call center staffing using queueing models: A study of service level, *Journal of Applied Research and Technology*, **22(2)**, 230-242.
- Ross S.M. *Stochastic Processes*, 2nd Edition, John Wiley and Sons.
- Subramanian M.G., Ayyappan G. and Sekar G. (2011), M/M/c Retrial queueing system with breakdown and repair of services, *Asian Journal of Mathematics and Statistics*, **4(4)**, 214-223.
- Szavits-Nossan J. and Grima R. (2024), Solving stochastic gene-expression models using queueing theory: a tutorial review, *Biophysical Journal*, **123(9)**, 1034-1057.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



SDE-Based Models for Trajectories on Linear Networks

Ali Sharifi^{*1}

¹Department of Mathematics and Mathematical Statistics, Umeå University, Sweden

Abstract

The growing ability to monitor movements along directed linear networks, such as road networks, poses new modelling challenges, as existing methods are often designed for unconstrained movements, such as animal movement, and do not account for the routing rules imposed by an underlying network structure. To capture travel dynamics within such environments, we propose Cox-Ingersoll-Ross models modulated by a latent Markov chain. Our framework allows for selective parameter variation across latent states, enabling flexible modelling of heterogeneous movement behaviour. We develop novel EM-based estimators, establish their consistency, and provide algorithms for simulating trajectories on networks. Through a simulation study, we evaluate the performance of our estimators under a range of scenarios and employ our proposal to model ambulance movements, classified by priority level, in northern Sweden.

Keywords: Ambulance movements; Cox-Ingersoll-Ross models; EM algorithm; Latent; Network-constrained movement; Trajectory simulation.

Mathematics Subject Classification (2020): 62M05, 60J25, 62P30.

^{*}Corresponding author, ali.sharifi@umu.se



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Application of Functional Linear Regression with Impact Points to Real Data: A Case Study

Alireza Shirvani^{*1}

¹Department of Statistics, Velayat University, Iranshahr, Iran

Abstract

This study demonstrates the practical application of the methodology introduced in Shirvani *et al.* (2024) for detecting and testing the significance of impact points within a functional linear regression framework. Using data from a Google AdWords campaign, the analysis confirms that the functional predictor at the identified impact points significantly influences the response variable, highlighting the utility of the approach in real-world settings. The results underscore the importance of impact point detection in functional data modeling.

Keywords: Functional linear regression, Impact point, Hypothesis test.

Mathematics Subject Classification (2020): 62G10.

1 Introduction

Functional Data Analysis (FDA) is a powerful statistical tool for modeling data observed over a continuum, such as time or space (for more details, see Ramsay and Silverman (2005)). A key topic within FDA is the identification of impact points, discrete observation times at which the corresponding functional values have a significant influence on the response variable, considered by Kneip *et al.* (2016).

In Shirvani *et al.* (2024), a rigorous hypothesis testing approach was proposed to evaluate whether the functional predictor evaluated at specific impact points in the domain has a statistically significant effect on the response. The framework involves constructing an appropriate test statistic whose distribution under the null hypothesis is well characterized.

The investigation of impact points in functional linear models has received considerable attention. Cardot *et al.* (2003) pioneered early work on hypothesis testing in this context, providing foundational tools. Crambes *et al.* (2009) and Kneip *et al.* (2016) further refined estimation procedures and impact detection techniques. Recent developments, such as those by Liebl *et al.* (2020), have enhanced impact point identification through improved estimation methods and bootstrap approaches. The testing procedure outlined in Shirvani *et al.* (2024) builds upon these advancements, offering a statistically rigorous and practically applicable framework.

^{*}Corresponding author, a.shirvani@velayat.ac.ir

This paper applies the methodology from Shirvani *et al.* (2024) to a real dataset obtained from a Google AdWords campaign, aiming to demonstrate its effectiveness in practical data analysis and its capacity to identify meaningful impact points.

2 Model

We analyze a functional linear model where the response variable Y is scalar, and the predictor variable X belongs to $L^2(\mathcal{I})$, with $\mathcal{I} = [a, b] \subset \mathbb{R}$ being a bounded interval, and $L^2(\mathcal{I})$ representing the space of all square-integrable functions on \mathcal{I} . The data consist of independent observations (X_i, Y_i) , for $i = 1, \dots, n$, of the pair (X, Y) . The functional predictor X over \mathcal{I} satisfies

$$E \left[\int_{\mathcal{I}} X^2(t) dt \right] < \infty.$$

For simplicity, we assume the variables are centered, meaning

$$E(Y) = 0 \quad \text{and} \quad E[X(t)] = 0 \quad \text{for a.e. } t \in [a, b].$$

Then, the Y_i 's are generated according to the following functional linear regression model with points of impact:

$$Y_i = \int_{\mathcal{I}} \beta(t) X_i(t) dt + \sum_{r=1}^S \beta_r X_i(\tau_r) + \varepsilon_i, \quad i = 1, \dots, n, \quad (1)$$

where the errors ε_i are i.i.d. with zero mean and finite variance σ^2 , the slope function β is an unknown element of $L^2(\mathcal{I})$, and the errors ε_i are independent of the X_i .

Note that the overall effect of the entire trajectory $X_i(\cdot)$ on Y_i is captured by

$$\int_{\mathcal{I}} \beta(t) X_i(t) dt$$

in Model (1). The model also incorporates an unknown number $S \in \mathbb{N}$ of specific time points τ_1, \dots, τ_S , called “points of impact,” such that the response Y_i is significantly influenced by the values $X_i(\tau_1), \dots, X_i(\tau_S)$.

Model (1) involves the unknown parameters: the function $\beta(t)$, the number of impact points $S \geq 0$, and the impact points τ_r , as well as the coefficients β_r , for $r = 1, \dots, S$, all of which must be estimated from the data.

3 Data Description and Application

The dataset under consideration includes daily impressions and clicks. This real-world dataset, obtained from a Google AdWords campaign—an essential platform of Alphabet (Google’s parent company)—provides a practical example for applying the proposed impact point detection method. Google AdWords operates on a Pay-Per-Click (PPC) mechanism, where advertisers bid for sponsored impressions that appear alongside search results when users enter relevant keywords. The auction system allocates a limited number of sponsored impressions based on an ad-rank procedure that accounts for bid amounts and quality scores. Crucially, advertisers only pay when a user clicks on their ad.

To model the relationship between the total annual clicks and the trajectories of daily impressions, we used a functional linear regression framework with points of impact. Specifically, the response variable is the logarithm of the total annual clicks:

$$Y_i = \log \left(\sum_{t=1}^{365} \text{clicks}_{it} \right),$$

and the functional predictor is given by the logarithm of daily impressions:

$$X_i(t) = \log(\text{impressions}_{it}),$$

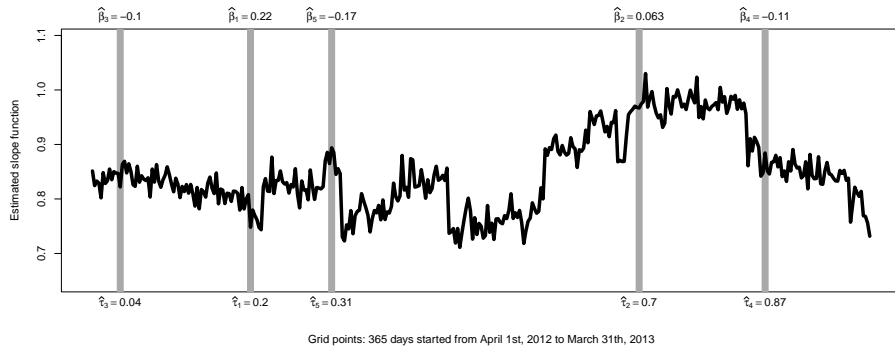


Figure 1: Estimated slope function and points of impact using Kneip *et al.* (2016) method. The horizontal axis shows 365 days of the year, and the vertical axis displays the estimated slope function. The impact points are indicated by vertical lines.

for $i = 1, \dots, 903$ and $t = 1, \dots, 365$. The data originate from a Google AdWords campaign conducted between April 1, 2012, and March 31, 2013.

Applying the method proposed by Kneip *et al.* (2016), the estimated impact points are summarized in Figure 1. Five impact points were identified at the following dates: June 14 ($\hat{\tau}_1, \hat{\beta}_1 = 0.22$), December 13 ($\hat{\tau}_2, \hat{\beta}_2 = 0.06$), April 14 ($\hat{\tau}_3, \hat{\beta}_3 = -0.10$), February 10 ($\hat{\tau}_4, \hat{\beta}_4 = -0.11$), and July 22 ($\hat{\tau}_5, \hat{\beta}_5 = -0.17$). These impactful days could correspond to specific seasonal effects, holiday periods, or promotional spikes, providing valuable insights in marketing and campaign planning. To assess whether including these impact points significantly enhances the model, the hypothesis testing procedure suggested by Shirvani *et al.* (2024) was employed. The resulting p-value was zero, indicating that the impact points have a statistically significant effect on the response variable.

4 Conclusion

This application illustrates that the impact points approach from Kneip *et al.* (2016) effectively identifies significant impactful days in a functional predictor. The results align with domain expectations and provide valuable interpretability for practical applications. Importantly, the hypothesis test introduced by Shirvani *et al.* (2024) assesses the overall significance of the entire set of impact points collectively, rather than testing each coefficient individually. The fact that the p-value is zero indicates that not all impact point coefficients are zero—some have a meaningful effect—confirming the presence of significant impact points and emphasizing the importance of this test in validating the collective impact of those points.

In the future, it will be important to develop a hypothesis test that can assess the significance of individual impact points separately, rather than evaluating the collective set. Such a test would allow for a more detailed understanding of which specific points have a meaningful influence on the response variable, providing greater specificity and interpretability in impact point detection.

References

- Cardot, H., Ferraty, F., Mas, A., and Sarda, P. (2003), Testing hypotheses in the functional linear model, *Scandinavian Journal of Statistics*, **30**, 241–255.
- Crambes, C., Kneip, A., and Sarda, P. (2009), Smoothing splines estimators for functional linear regression, *Annals of Statistics*, **37**, 35–72.

- Kneip, A., Dette, H., and Gertheiss, J. (2016), Variable selection in functional linear models, *Annals of Statistics*, **44**(3), 1060–1088.
- Liebl, D., Rameseder, S. and Rust, C. (2020), Improving estimation in functional linear regression with points of impact: insights into Google AdWords, *Journal of Computational and Graphical Statistics*, **29**, 814–826.
- Ramsay, J., and Silverman, B. (2005), *Functional Data Analysis*, 2nd Ed., Springer.
- Shirvani, A., Khademnoe, O. and Hosseini-Nasab, M. (2024), Hypothesis testing for points of impact in functional linear regression, *Computational and Applied Mathematics*, **43**, 201.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Application of statistical models in designing a system to save animal lives

Morteza Taheri Saif Abad^{*1}

¹Department of Statistics, Faculty of Science, Shiraz University, Shiraz, Iran

Abstract

Animal-related road accidents have emerged as a major cause of wildlife mortality in recent years. To mitigate this pressing problem, it is essential to implement advanced animal detection and identification systems in vehicles. Using statistical models, it is possible to identify animal behavior patterns and also assess risks, which will lead to improved resource management and relief efforts. In this article, an attempt has been made to explain how road accidents can be reduced by using an early warning system to prevent accidents, using statistical models such as Markov chains and reliability parameters. This system can help in quick and effective decision-making by analyzing data and predicting critical situations.

Keywords: Stress-Strength Parameter, Markov Process, Reliability, Warning System.

Mathematics Subject Classification (2020): 62N05, 60Jxx.

1 Introduction

Environmental protection plays an important role in the lives of future generations. Natural and unnatural factors have caused a decrease in the animal population. Fragmentation of human habitats has increased the rate of wildlife collisions with vehicles, creating an urgent need for intelligent mitigation systems. In some countries, such as Canada, systems have been designed to reduce the rate of wildlife collisions with vehicles. These systems suffer from three fundamental limitations:

- 1) Temporal proximity: Threshold-based warnings cannot anticipate all emerging hazards.
- 2) Behavioral oversimplification: Deterministic models fail to capture random movement patterns.
- 3) System fragility: There is no quantitative measure of reliability for field deployment.

To eliminate the limitations mentioned, statistical models can be used to improve the performance of systems. Statistical models play an important role in modeling animal behavior and designing systems to save animals from danger. For example, you can refer to [Lebreton and *et al.* \(1993\)](#), [Festing \(2004\)](#), [Festing \(2006\)](#) and [Williams and *et al.* \(2020\)](#).

In this paper, a system consisting of 4 parts is designed that can save the lives of some animals. Different methods of obtaining system reliability and their combination were used to calculate

^{*}Corresponding author, taherisaifmorteza@gmail.com

the overall system reliability in order to better predict the system lifetime based on different system states and calculate the field deployment reliability. Different operating modes of the system were considered and a Markov model was proposed to obtain the overall system failure probability and system availability. Also, by considering the different states of animals in terms of being in danger, a Markov model was proposed to predict upcoming dangers by considering the random behavior of animals.

By combining reliability calculation methods, Markov models, Hazard rate function, and Checking the presence of danger at a specific time and using them to improve system performance, nature can be protected to a great extent.

2 System Design

We want to design a system that warns the driver before an accident. Therefore, this system should consist of four parts:

1. Motion sensors: Detecting animal body heat in the dark or in adverse weather conditions
2. Smart cameras: Image processing to detect the type of animal and its direction of movement,
3. Warning system: Send audio/visual alerts to the driver,
4. Emergency braking: Automatic braking activation if the driver does not react.

Systems that are designed from parts like the system above usually face various errors and risks, some of which are mentioned below:

- 1- Sensor error in detecting animals,
- 2- Delay in sending the warning,
- 3- Camera failure.
- 4- System hardware failure

3 System reliability

One of the most important points in designing various systems is that the system must have high reliability. Suppose that the random variable X represents the lifetime of a system. The reliability of a system is represented by $R(t)$ and is defined as follows

$$R(t) = P(X > t) \quad t > 0 \quad (1)$$

Using the formula (1), the reliability of many systems can be easily calculated. But in sensitive cases, its reliability must be increased by considering all the possible risks that a system may face. Therefore, we will use several methods to increase the reliability of the system.

3.1 Reliability for hardware components

For the hardware components of the system, a stress-strength reliability model can be used, considering the stresses imposed on the system and the system capacity. Suppose that the stresses imposed on the system, for example, the number of processing requests per second, are represented by a random variable X and the strength of the system, for example, the server processing power, is represented by a random variable Y . In this case, stress-strength reliability is equal to

$$R = P(Y > X) \quad (2)$$

Calculating the formula (2) is simple in most cases, and in cases where the calculation of the formula is complex or several different stresses enter the system, it can be easily calculated using nonparametric methods.

3.2 Probability of complete failure

Calculating the probability of the system reaching a complete failure state is another issue that must be examined to increase the reliability of the system. To calculate this probability, we need to examine the system's operating process and its states. For most systems, the system operating states can be considered as follows:

- a) Operational: The system is healthy.
- b) Partial failure: Some sensors have failed or the alarm system has been muted.
- c) Complete failure: The system is inoperable.

The best way to examine the transition between states and determine the probability of reaching a complete system failure is to model the system performance using processes such as Markov, semi-Markov, or a renewal process.

Example 3.1. *Suppose that the transition between system states can be represented by a Markov process with a transition probability matrix P as follows. In this case, by obtaining the stationary distribution of the process, the probability of the system reaching a complete failure can be calculated.*

$$P = \begin{bmatrix} 0.94 & 0.04 & 0.02 \\ 0.33 & 0.60 & 0.07 \\ 0 & 0 & 1 \end{bmatrix}$$

By solving the Markov equations $\pi = \pi P$ where $\pi = (\pi_1, \pi_2, \pi_3)$, the long-term probability of each state is given by:

$$\pi_1 = 0.80, \quad \pi_2 = 0.15, \quad \pi_3 = 0.05$$

That is, the system is operational 80 percent of the time, has a partial failure 15 percent of the time, and a complete failure 5 percent of the time.

3.3 System Availability

In this process, system availability is the percentage of time the system is in *a* or *b*. Assuming the conditions of Example 1 are met, then

$$Availability = \pi_1 + \pi_2 = 0.80 + 0.15 = 0.95$$

With 95 percent communication capability, this system can reduce animal collisions by up to 80 percent. By improving the Markov model and adding layers of redundancy, higher accuracy can be achieved.

3.4 Equipment Lifetime Prediction

The system designed in section 2 consists of different parts that have different functions and different structures. Therefore, it is better to calculate a separate reliability function for each. One of the best reliability models for predicting the lifetime of equipment is the Weibull reliability function model, which is as follows:

$$R(t) = e^{-(t/\alpha)^\beta} \quad (3)$$

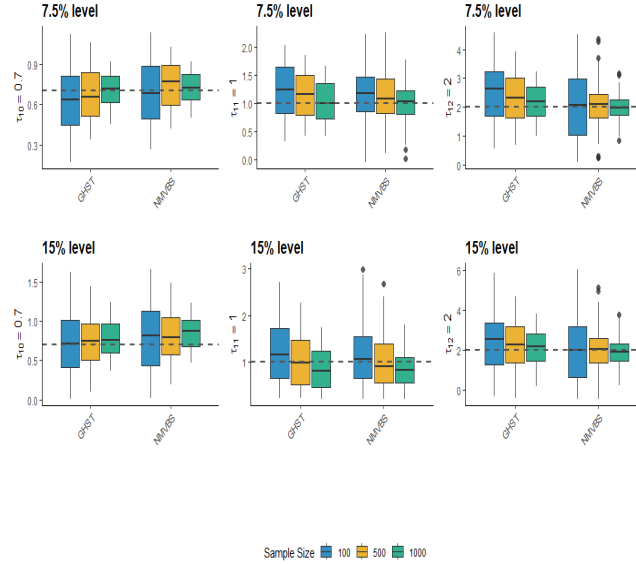
Where α is the scale parameter (characteristic life) and β is the shape parameter (failure rate). The Weibull model is a suitable model for predicting the time to replace sensors before failure.

Example 3.2. *Suppose that for the system discussed in Section 1, we have the following conditions:*

- a) *The sensors have an exponential lifetime with a failure rate of 0.002.*
- b) *The stress applied to the camera system has a Weibull(2, 1.5) distribution and its strength has a Weibull(2.5, 2) distribution.*
- c) *The lifetime of the warning system has a Weibull(50, 1.8) distribution.*

d) The stress applied to the brake system has a $N(50, 10)$ distribution and its strengt has a $N(65, 8)$ distribution.

Figure ?? shows the reliability of such a system. Since the order of the steps in this system is serial, the reliability of the sum is less than the reliability of the parts.



3.5 Using spares

In some parts of system, such as the animal detection part where sensors are used, it is better to use several sensors in parallel due to their high sensitivity, thereby increasing the reliability of the system.

Example 3.3. Suppose we have three sensors with a reliability of 0.9 placed in parallel. In this case, the total reliability is

$$R_{total} = 1 - (1 - 0.9)^3 = 0.999$$

3.6 Reduce recognition error

Use machine learning to reduce recognition error. For example, use the YOLO algorithm to recognize animals in images. Collect thousands of images of animals in different conditions. Test the system in different conditions.

4 Improving sensor performance

Here, we present a Markov model for modeling animal behavior, movement patterns, and environmental hazards to improve sensor performance. For this model to function properly, in addition to vehicle sensors, sensors need to be placed where animal populations are high and the likelihood of them crossing the road is high. For example, electronic signs should be placed on the side of the road that indicate the possibility of animals crossing or flashing lights should be placed on the side of the road to scare the animal away.

The Markov chain states for improving sensor operation are defined as follows:

- 1- Safe state: the animal is away from the road.
- 2- Possible danger state: the animal is near the road.
- 3- High danger state - the animal is on the road.
4. Accident state

Example 4.1. Suppose that the transition matrix in a Markov chain is as follows:

$$\begin{pmatrix} 0.7 & 0.3 & 0 & 0 \\ 0.2 & 0.5 & 0.3 & 0 \\ 0.1 & 0.2 & 0.6 & 0.1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

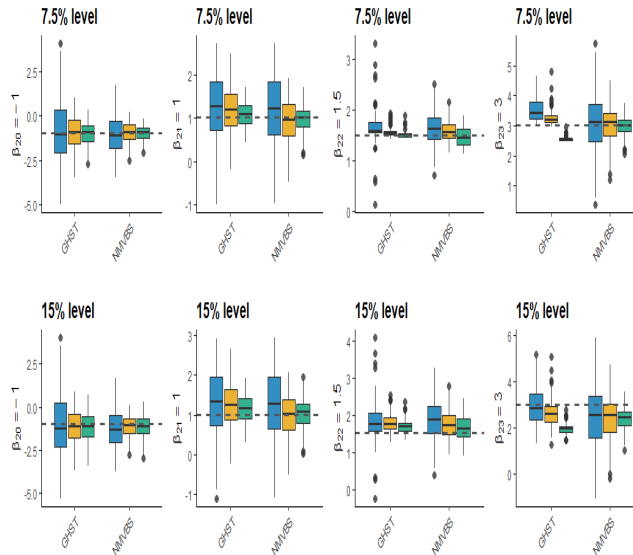
Then the probability that an animal will be in the forest or away from the road after five transition steps when it is in step zero is 0.04.

By monitoring the current state of the animal, the system calculates the probability of being in a danger state, and if the probability of danger is higher than a certain level, an alarm is activated. For example, if the Markov model predicts that deer will move towards the road with a probability of over 75 percent, warning lights for drivers will be activated.

4.1 Hazard Rate Function

Another way to improve sensor performance is to use hazard functions to model the probability of an accident at different time intervals. If the hazard probability exceeds a certain threshold, the warning system is activated. Suppose we have data that includes the time until the animals reach the road (Time-to-Event) and whether an accident occurred or not (Censored Data). Models such as the Cox Proportional Hazards model or the Weibull distribution can be used to predict the hazard.

Example 4.2. suppose that the vector (5, 10, 3, 8, 12, 7, 15, 20, 4, 6) represents the time until an accident occurs or does not occur, and the vector (1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0) represents the occurrence or non-occurrence of an accident. Figure ?? shows a simulation in R for the cumulative hazard function graph for animal collisions. Also, the probability of an accident occurring in less than 7 minutes is 0.4.

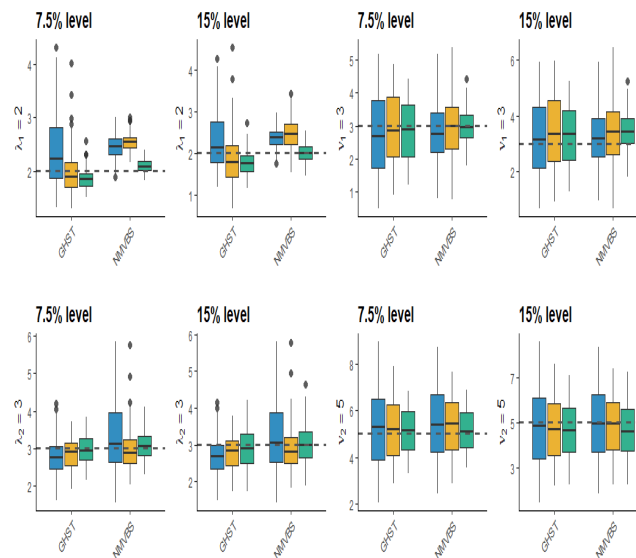


4.2 Checking the presence of danger at a specific time

The number of accidents varies due to the variable number of cars and animals at different times. For this reason, the probability of the presence of danger at a specific hour or day should also be checked.

Example 4.3. Suppose that the number of animals near the road at a specific time and place can be represented by $Pois(5)$. Suppose that the number of cars moving on the road near

these animals is between 10 and 100. Suppose that the probability of an accident is equal to $\frac{\text{Numberofcars} * \text{Numberofanimals}}{1000}$ and if this probability is above 0.7, the warning system is turned on. Figure ?? shows the risk graph over time for thirty consecutive days.



If the number of animals and cars is updated, using this method, dangerous days are identified and the driver is warned on that day to be more careful.

5 Conclusion

Statistical models are applicable and efficient in almost any field. In this paper, we have shown that statistical models can be used to design highly reliable animal rescue systems. The application of different reliability methods such as 1 and 2 was shown in different sections. We also showed that statistical models such as Markov and Weibull can be used to predict failures. Of course, the mentioned method is expensive and one of the main problems that makes vehicle manufacturers reluctant to design such systems is its high cost.

References

- Lebreton, J. D., Pradel, R. and Clobert, J. (1993), The statistical analysis of survival in animal populations, *Trends in Ecology and Evolution*, **8**, 91-95.
- Festing, M. F. (2004), Good experimental design and statistics can save animals, but how can it be promoted?. *Alternatives to Laboratory Animals*, **32**, 133-135
- Festing, M. F. (2006), Design and statistical methods in studies using animal models of development. *Ilar Journal*, **47**, 5-14.
- Williams, H. J., Taylor, L. A., Benhamou, S., Bijleveld, A. I., Clay, T. A., de Grissac, S. and Börger, L. (2020), Optimizing the use of biologgers for movement ecology research. *Journal of Animal Ecology*, **89**, 186-206.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Lookback options in stochastic intensity models

Mahdieh Tahmasebi*

¹ Department of Applied Mathematics, Tarbiat Modares University, P.O. Box 14115-134, Tehran, Iran.

Abstract

In this manuscript, we first consider the martingale representation theorem when the underlying asset is driven by a process with stochastic intensity, using a powerful tool called Malliavin calculus. We will obtain an expression for the Malliavin derivative concerning a Wiener-Poisson space. Then, using this Clark-Ocone formula, we price a barrier lookback option depending on the maximum of the asset.

Keywords: Processes with stochastic intensity, Clark-Ocone formula, Malliavin calculus.

Mathematics Subject Classification (2020): 60H07, 60G55, 60J76.

*tahmasebi@modares.ac.ir



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Bayesian Estimation for the Parameters of the Lindley-Burr XII Distribution

Bahman Tarvirdizade^{*1}, Salman Babayi¹

¹Department of Mathematics, Faculty of Sciences, Urmia University, Urmia, Iran

Abstract

In this paper, the Bayesian inference for the parameters of a new flexible three-parameters distribution called the Lindley-Burr XII distribution is considered. The estimation of parameters are obtained by using the Markov Chain Monte Carlo (MCMC) methods under squared error, linear-exponential, and Stein's loss functions. An application to real data set is provided for illustrative purposes. Finally, a Monte Carlo simulation study is conducted to investigate and compare the performance of different types of Bayes estimators presented in this paper.

Keywords: Lindley-Burr XII distribution, Bayesian estimation, Monte Carlo simulation.

Mathematics Subject Classification (2020): 60E05, 62F15, 65C05.

1 Introduction

The Lindley and Burr XII distributions are two well-known distributions that have been extensively used over the past decades for modeling data in many fields such as business, economics, actuarial modeling, queuing problems, quality control, medicine, reliability and life testing problems. Although these distributions are very useful for modeling lifetime data, they have a restriction to accommodate some nonmonotone hazard rate functions. The hazard rate function (hrf) of Lindley distribution is increasing and that of Burr XII distribution can be decreasing or unimodal. Thus, these distributions may not provide a reasonable parametric fit for modeling phenomena with non-monotone hazard rates such as the bathtub-shaped and modified bathtub-shaped hazard rates which are often encountered in practice. As a solution, several authors have considered modified or generalized forms and extensions of these distributions to give them more flexibility in describing various types of data. For instance, new extended generalized lindley distribution by [Maya and Irshad \(2017\)](#), odd log-logistic Marshal-Olkin Lindley distribution by [Alizadeh et al. \(2017\)](#), logistic Burr XII distribution by [Guerra et al. \(2023\)](#) and Marshall-Olkin Weibull-Burr XII distribution by [Alsadat et al. \(2023\)](#).

Recently, [Makubate et al. \(2021\)](#) has introduced a new flexible three-parameter distribution using the combination of the Lindley and the Burr XII distributions in a serial system which is called the Lindley-Burr XII (L-BXII) distribution. The cumulative distribution function

^{*}Corresponding author, b.tarvirdizade@urmia.ac.ir

(cdf) and the probability density function (pdf) of the L-BXII distribution with parameters α , β and θ are given by

$$F(x) = 1 - (1 + x^\alpha)^{-\beta} \left(1 + \frac{\theta x}{\theta + 1}\right) e^{-\theta x}, \quad x > 0, \alpha, \beta, \theta > 0 \quad (1)$$

and

$$f(x) = \left(\frac{\alpha \beta x^{\alpha-1}}{1 + x^\alpha} \left(1 + \frac{\theta x}{\theta + 1}\right) + \frac{\theta^2 (1 + x)}{1 + \theta} \right) (1 + x^\alpha)^{-\beta} e^{-\theta x}, \quad x > 0, \alpha, \beta, \theta > 0 \quad (2)$$

respectively. The L-BXII distribution could be applied effectively for analyzing different types of lifetime data since it has a simple form of hrf which can accommodate decreasing, increasing, unimodal, bathtub-shaped and modified bathtub-shaped hazard rates. Some statistical properties and estimation of the parameters of the L-BXII distribution using maximum likelihood method were studied by [Makubate *et al.* \(2021\)](#). Also, it was shown that the L-BXII distribution provides a better fit than other three-parameter distributions for two real lifetime data, while its hrf is very simple in comparison with those of the competitor distributions. In this paper, we apply a Bayesian approach to estimate the parameters of the L-BXII distribution based on a complete sample. First the likelihood function of the parameters are presented and then, the Bayes estimates of the parameters based on symmetric and asymmetric loss functions are obtained.

The rest of the paper is organized as follows. In Section 2, the Bayes estimates of unknown parameters under the squared error, linear-exponential (LINEX) and Stein loss functions via the Metropolis-Hastings method are obtained. In Section 3, a real data set is analyzed to illustrate the estimation procedures discussed in the previous section. In Section 4, to investigate and compare the performance of the different methods of estimation presented in this paper, outcomes of a Monte Carlo simulation study are presented. Finally, some conclusions are provided in Section 5.

2 Bayesian Estimation

In this section, we discuss Bayesian estimation of the parameters of the L-BXII distribution under different loss functions. One of the most commonly used loss function is squared error loss (SEL) function which is given by $L(\Theta, \delta(X)) = (\delta(X) - \Theta)^2$, where δ is a decision rule based on the data and Θ is the unknown parameter. The symmetric nature of SEL function gives equal weight to overestimation and underestimation of the parameters under consideration. However, in life testing, overestimation may be more serious than underestimation or vice versa. In these cases, the use of an asymmetric loss function which assigns greater importance to overestimation or underestimation may be more appropriate. For this purpose, [Varian \(1975\)](#) proposed a convex but asymmetric loss function which is known as LINEX loss function and is defined as

$$L(\Theta, \delta) = e^{c(\delta - \Theta)} - c(\delta - \Theta) - 1, \quad c \neq 0 \quad (3)$$

The shape parameter c is known and gives the degree of asymmetry. If $c > 0$, the overestimation is more serious than underestimation and if $c < 0$, underestimation is more serious than overestimation. If c close to zero, the LINEX loss is approximately SEL and therefore almost symmetric. Under the LINEX loss function (3), the Bayes estimator of Θ that minimizes the posterior risk $E[L(\Theta, \delta(X))|X]$ is given by

$$\delta_{BL}(X) = -\frac{1}{c} \log E \left(e^{-c\Theta} | X \right), \quad (4)$$

provided that the expectation exists and is finite. Another useful asymmetric loss function is the Stein loss function which is also known as entropy loss function. This loss function has the form

$$L(\Theta, \delta) = \frac{\delta}{\Theta} - \log \frac{\delta}{\Theta} - 1, \quad (5)$$

This loss is a convex function of δ and more penalized underestimation than overestimation. Under this loss function the Bayes estimator $\delta_{BST}(X)$ that minimizes the posterior risk $E[L(\Theta, \delta(X))|X]$ is given by

$$\delta_{BST}(X) = \left\{ E \left(\frac{1}{\Theta} | X \right) \right\}^{-1}, \quad (6)$$

provided that the expectation exists.

In Bayesian inference, we need to determine the likelihood function and prior distributions for unknown parameters. Let $\mathbf{x} = (x_1, x_2, \dots, x_n)$ be n observations of a random sample from the L-BXII(α, β, θ). The likelihood function of this sample using (2) can be written as

$$L(\alpha, \beta, \theta | \mathbf{x}) = \prod_{i=1}^n \left(\frac{\theta^2(1+x_i)(1+x_i^\alpha) + \alpha\beta x_i^{\alpha-1}(1+\theta+\theta x_i)}{1+\theta} \right) (1+x_i^\alpha)^{-(\beta+1)} e^{-\theta x_i}, \quad (7)$$

The MLEs of α , β and θ , say $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\theta}$, can be obtained through the solution of the three nonlinear equations which are obtained by setting the first partial derivatives of the log-likelihood function with respect to α , β and θ equal to zero (see Makubate *et al.* (2021)). These equations cannot be solved analytically and therefore, we have to solve the equations numerically. We can use iterative techniques such as a Newton-Raphson type algorithm to obtain the MLEs of the parameters α , β and θ . To find the standard error of the MLEs of the parameters α , β and θ , we can obtain the asymptotic variance-covariance matrix of the MLEs, which need to calculate the observed information matrix. The variance-covariance matrix \mathbf{V} can be approximated by the reciprocal of the observed information matrix, i.e., $\mathbf{V} = \mathbf{I}^{-1}$. Since \mathbf{V} involves the parameters α , β and θ , we replace the parameters by the corresponding MLEs in order to obtain an estimate of \mathbf{V} , which is denoted by

$$\hat{\mathbf{V}} = \begin{pmatrix} \hat{V}_{11} & \hat{V}_{12} & \hat{V}_{13} \\ \hat{V}_{21} & \hat{V}_{22} & \hat{V}_{23} \\ \hat{V}_{31} & \hat{V}_{32} & \hat{V}_{33} \end{pmatrix} = \begin{pmatrix} \hat{I}_{11} & \hat{I}_{12} & \hat{I}_{13} \\ \hat{I}_{21} & \hat{I}_{22} & \hat{I}_{23} \\ \hat{I}_{31} & \hat{I}_{32} & \hat{I}_{33} \end{pmatrix}^{-1}, \quad (8)$$

where \hat{I}_{ij} is the (i, j) th element of the observed information matrix I with α , β and θ replaced by $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\theta}$, respectively. The elements of the observed information matrix I are given in the Appendix.

For the prior distributions, we assumed that α , β and θ have independent gamma priors as $\text{Gamma}(\gamma_1, \lambda_1)$, $\text{Gamma}(\gamma_2, \lambda_2)$ and $\text{Gamma}(\gamma_3, \lambda_3)$, respectively, with the pdf's given by

$$\pi(\Theta | \gamma_i, \lambda_i) = \frac{\lambda_i^{\gamma_i} \Theta^{\gamma_i-1} e^{-\lambda_i \Theta}}{\Gamma(\gamma_i)}, \quad \Theta > 0, \quad \gamma_i, \lambda_i > 0, \quad (9)$$

where Θ can be each of the parameters α , β and θ and the hyperparameters (γ_i, λ_i) , $i = 1, 2, 3$, are assumed to be known. Now, using (7) and (9), the joint posterior density function of α , β and θ can be written as

$$\begin{aligned} \pi(\alpha, \beta, \theta | \mathbf{x}) &= \frac{\pi(\alpha)\pi(\beta)\pi(\theta)L(\alpha, \beta, \theta | \mathbf{x})}{\int_{\theta} \int_{\beta} \int_{\alpha} \pi(\alpha)\pi(\beta)\pi(\theta)L(\alpha, \beta, \theta | \mathbf{x}) d\alpha d\beta d\theta} \\ &\propto \prod_{i=1}^n \left(\frac{\theta^2(1+x_i)(1+x_i^\alpha) + \alpha\beta x_i^{\alpha-1}(1+\theta+\theta x_i)}{(1+\theta)(1+x_i^\alpha)^{\beta+1}} \right) \\ &\quad \times \alpha^{\gamma_1-1} \beta^{\gamma_2-1} \theta^{\gamma_3-1} e^{-\lambda_1 \alpha - \lambda_2 \beta - (\lambda_3 + \sum_{i=1}^n x_i) \theta}. \end{aligned} \quad (10)$$

Since the above multiple integrals can not be solved analytically, the expression for $\pi(\alpha, \beta, \theta | \mathbf{x})$ in (10) can not be written in a closed form. Therefore, we need a simulation technique to compute the Bayes estimate of the parameters. We adopt the Gibbs sampling technique which requires a decomposition of the joint posterior distribution into full conditional distributions for each parameter and then sampling from them. The full conditional distributions of α , β and θ can be obtained as follows

$$\pi(\alpha | \beta, \theta, \mathbf{x}) \propto \prod_{i=1}^n \left(\frac{\theta^2(1+x_i)(1+x_i^\alpha) + \alpha\beta x_i^{\alpha-1}(1+\theta+\theta x_i)}{(1+\theta)(1+x_i^\alpha)^{\beta+1}} \right) \alpha^{\gamma_1-1} e^{-\lambda_1 \alpha}, \quad (11)$$

$$\pi(\beta|\alpha, \theta, \mathbf{x}) \propto \prod_{i=1}^n \left(\frac{\theta^2(1+x_i)(1+x_i^\alpha) + \alpha\beta x_i^{\alpha-1}(1+\theta+\theta x_i)}{(1+\theta)(1+x_i^\alpha)^{\beta+1}} \right) \beta^{\gamma_2-1} e^{-\lambda_2\beta}, \quad (12)$$

$$\pi(\theta|\alpha, \beta, \mathbf{x}) \propto \prod_{i=1}^n \left(\frac{\theta^2(1+x_i)(1+x_i^\alpha) + \alpha\beta x_i^{\alpha-1}(1+\theta+\theta x_i)}{(1+\theta)(1+x_i^\alpha)^{\beta+1}} \right) \theta^{\gamma_3-1} e^{-(\lambda_3 + \sum_{i=1}^n x_i)\theta}. \quad (13)$$

The full conditional distributions for α , β and θ cannot be reduced analytically to well-known distributions and therefore it is not possible to sample directly by standard methods. In order to generate values of α , β and θ from (11)–(13), we use the Metropolis-Hastings algorithm into the Gibbs sampling algorithm as explained by Tierney (1994). Therefore, we can use the following algorithm for Gibbs sampling:

Step 1. Start with an initial guess $(\alpha^{(0)}, \beta^{(0)}, \theta^{(0)})$ and set $t = 1$.

Step 2. Using Metropolis-Hastings method, generate $\alpha^{(t)}$ from $\pi(\alpha|\beta^{(t-1)}, \theta^{(t-1)}, \mathbf{x})$ with the proposal distribution

$$q(\alpha) \propto N(\alpha^{(t-1)}, C_\alpha \hat{V}_{11}) I(\alpha > 0),$$

where C_α is a scaling factor, \hat{V}_{11} is given in (8) and $N(\mu, \sigma^2) I(\alpha > 0)$ denote the normal distribution $N(\mu, \sigma^2)$ truncated on $(0, \infty)$.

Step 3. Using Metropolis-Hastings method, generate $\beta^{(t)}$ from $\pi(\beta|\alpha^{(t)}, \theta^{(t-1)}, \mathbf{x})$ with the proposal distribution

$$q(\beta) \propto N(\beta^{(t-1)}, C_\beta \hat{V}_{22}) I(\beta > 0),$$

where C_β is a scaling factor and \hat{V}_{22} is given in (8).

Step 4. Using Metropolis-Hastings method, generate $\theta^{(t)}$ from $\pi(\theta|\alpha^{(t)}, \beta^{(t)}, \mathbf{x})$ with the proposal distribution

$$q(\theta) \propto N(\theta^{(t-1)}, C_\theta \hat{V}_{33}) I(\theta > 0),$$

where C_θ is a scaling factor and \hat{V}_{33} is given in (8).

Step 5. Set $t = t + 1$.

Step 6. Repeat Steps 2–5, N times, and obtain the posterior sample $(\alpha^{(t)}, \beta^{(t)}, \theta^{(t)})$, $t = 1, \dots, N$.

After generating a sample from the above algorithm, we can compute the Bayes estimate of the parameters α , β and θ under the SEL, LINEX and Stein loss functions as follow

$$\tilde{\Theta}_{BS} = \hat{E}(\Theta|X) = \frac{1}{N-M} \sum_{i=M+1}^N \Theta^{(i)}, \quad (14)$$

$$\tilde{\Theta}_{BL} = -\frac{1}{c} \log \hat{E}(e^{-c\Theta}|X) = -\frac{1}{c} \log \left(\frac{1}{N-M} \sum_{i=M+1}^N e^{-c\Theta^{(i)}} \right), \quad (15)$$

$$\tilde{\Theta}_{BST} = \left\{ \hat{E} \left(\frac{1}{\Theta} | X \right) \right\}^{-1} = \left\{ \frac{1}{N-M} \sum_{i=M+1}^N \frac{1}{\Theta^{(i)}} \right\}^{-1}, \quad (16)$$

respectively, where M is the burn-in period and Θ can be each of the parameters α , β and θ .

3 Real Data Analysis

In this section, we analyze the real data to illustrate the use of our proposed estimation methods. We consider a real data set consist of the failure times of 50 components (per 1000h) taken from Murthy *et al.* (2004). The TTT-plot presented by them for this data set exhibits a bathtub-shaped hrf. Makubate *et al.* (2021) has fitted the L-BXII distribution to this data with the corresponding MLEs as follows:

$$\hat{\alpha} = 0.7389, \quad \hat{\beta} = 0.7305, \quad \hat{\theta} = 0.2094.$$

To obtain Bayes estimates of the parameters α , β and θ , we used small values of the hyperparameters as $(\gamma_1, \lambda_1) = (0.5, 0.25)$, $(\gamma_2, \lambda_2) = (0.75, 0.5)$ and $(\gamma_3, \lambda_3) = (1, 0.5)$ to reflect little

prior information. We computed the Bayes estimates of the parameters based on $N = 10000$ MCMC samples and discard the first $M = 2000$ values as burn-in period. The simulated values and Histogram of the parameters α , β and θ generated by the algorithm of Gibbs sampling are plotted in Figure 1. Based on these simulated values, the Bayes estimate of the parameters α , β and θ under the SEL function using (14) are computed as

$$\tilde{\alpha}_{BS} = 0.7498, \quad \tilde{\beta}_{BS} = 0.7595, \quad \tilde{\theta}_{BS} = 0.2220.$$

From (15), the Bayes estimate of the parameters α , β and θ under the LINEX loss function for $c = 2$ are given as

$$\tilde{\alpha}_{BL} = 0.7313, \quad \tilde{\beta}_{BL} = 0.7352, \quad \tilde{\theta}_{BL} = 0.2180,$$

and for $c = -1$ are given as

$$\tilde{\alpha}_{BL} = 0.7594, \quad \tilde{\beta}_{BL} = 0.7721, \quad \tilde{\theta}_{BL} = 0.2240.$$

Also under the Stein loss function, the Bayes estimate of the parameters α , β and θ using (16) are given as

$$\tilde{\alpha}_{BST} = 0.7240, \quad \tilde{\beta}_{BST} = 0.7254, \quad \tilde{\theta}_{BST} = 0.2008.$$

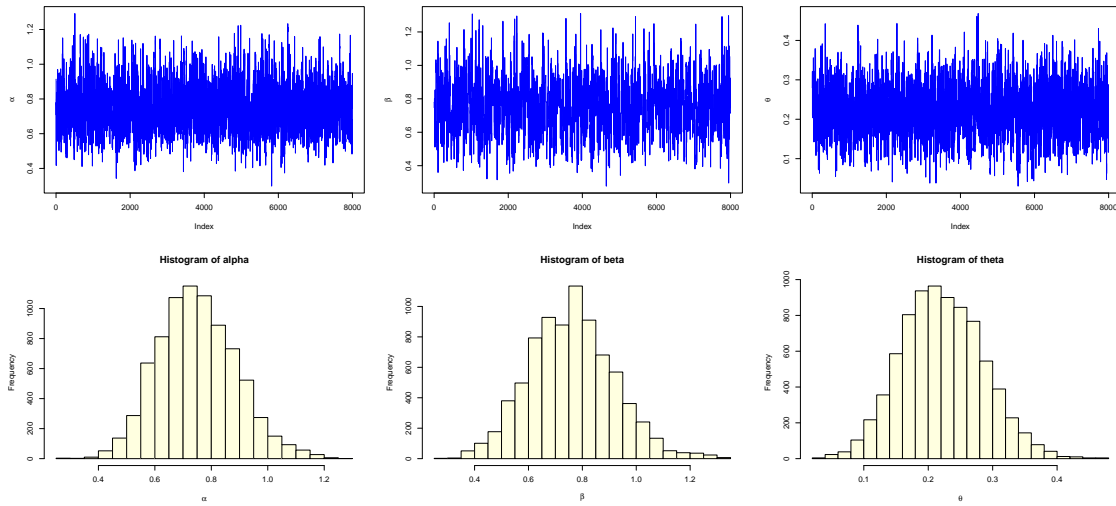


Figure 1: Simulated values and Histogram of the parameters α , β and θ

4 A Simulation Study

In this section, a Monte Carlo simulation study is conducted to investigate and compare the performance of the Bayes estimators presented in Section 2. In this simulation study we generate the samples of size $n = 20, 50, 100$ from the L-BXII distribution with different parameter combinations (α, β, θ) , namely $(0.25, 1, 0.75)$, $(1, 0.75, 0.5)$ and $(0.5, 0.25, 1)$, respectively. To compute different Bayes estimates, we use the small hyperparameter values as $(\gamma_1, \lambda_1) = (0.5, 0.75)$, $(\gamma_2, \lambda_2) = (1, 0.1)$ and $(\gamma_3, \lambda_3) = (0.25, 1)$. We generate $N = 10000$ MCMC samples and discard the first $M = 2000$ values as burn-in period as described in Section 2. The performance of the Bayes estimators is compared in terms of their estimated risk (ER). When θ is estimated by $\hat{\theta}$, the ER of θ under the SEL function is given by

$$ER_{BS}(\theta) = \frac{1}{T} \sum_{i=1}^T (\hat{\theta}_i - \theta)^2,$$

where T is the number of replications and $\hat{\theta}_i$ is the estimate of θ in i th replication. Moreover, the ER of θ under the LINEX and Stein loss functions are given by

$$ER_{BL}(\theta) = \frac{1}{T} \sum_{i=1}^T \left(e^{c(\hat{\theta}_i - \theta)} - c(\hat{\theta}_i - \theta) - 1 \right),$$

$$ER_{BST}(\theta) = \frac{1}{T} \sum_{i=1}^T \left(\frac{\hat{\theta}_i}{\theta} - \ln \frac{\hat{\theta}_i}{\theta} - 1 \right),$$

respectively. We report the average estimates and ER of the parameters in Tables 1–3. All the results are reported based on 2000 replications.

Table 1: Average estimates and estimated risk (in parentheses) of the parameters with $(\alpha = 0.25, \beta = 1, \theta = 0.75)$.

n	$\tilde{\alpha}_{BS}$	$\tilde{\alpha}_{BL}(c = 2)$	$\tilde{\alpha}_{BL}(c = -1)$	$\tilde{\alpha}_{BST}$
20	0.3024 (0.0536)	0.1879 (0.0405)	0.4703 (0.0794)	0.0621 (1.2069)
50	0.2803 (0.0188)	0.2207 (0.0117)	0.3842 (0.0207)	0.0853 (0.9805)
100	0.2651 (0.0093)	0.2429 (0.0084)	0.3180 (0.0123)	0.1104 (0.8394)
n	$\tilde{\beta}_{BS}$	$\tilde{\beta}_{BL}(c = 2)$	$\tilde{\beta}_{BL}(c = -1)$	$\tilde{\beta}_{BST}$
20	1.5203 (0.4626)	1.0542 (0.1903)	1.817 (0.3708)	0.7991 (0.1519)
50	1.2309 (0.1509)	1.0294 (0.1289)	1.4005 (0.1504)	0.8563 (0.1125)
100	1.1112 (0.1108)	0.9920 (0.1024)	1.1973 (0.0636)	0.8903 (0.0788)
n	$\tilde{\theta}_{BS}$	$\tilde{\theta}_{BL}(c = 2)$	$\tilde{\theta}_{BL}(c = -1)$	$\tilde{\theta}_{BST}$
20	1.0026 (0.2223)	0.6403 (0.1983)	1.1975 (0.2016)	0.4083 (0.3543)
50	0.8719 (0.1460)	0.6542 (0.1539)	0.9857 (0.0968)	0.4315 (0.2679)
100	0.7638 (0.0798)	0.6770 (0.1208)	0.8589 (0.0583)	0.4663 (0.1940)

Table 2: Average estimates and estimated risk (in parentheses) of the parameters with $(\alpha = 1, \beta = 0.75, \theta = 0.5)$.

n	$\tilde{\alpha}_{BS}$	$\tilde{\alpha}_{BL}(c = 2)$	$\tilde{\alpha}_{BL}(c = -1)$	$\tilde{\alpha}_{BST}$
20	0.5834 (0.9087)	0.2788 (0.9305)	1.4327 (0.4295)	0.2731 (1.0035)
50	0.6759 (0.4103)	0.3002 (0.4529)	1.3145 (0.2630)	0.3307 (0.9427)
100	0.8638 (0.2019)	0.5983 (0.3672)	1.1814 (0.1037)	0.4129 (0.7932)
n	$\tilde{\beta}_{BS}$	$\tilde{\beta}_{BL}(c = 2)$	$\tilde{\beta}_{BL}(c = -1)$	$\tilde{\beta}_{BST}$
20	1.4745 (0.6403)	0.9846 (0.2416)	1.8602 (0.6991)	0.9719 (0.1904)
50	1.3689 (0.5694)	0.9301 (0.1969)	1.6893 (0.6229)	0.8932 (0.1583)
100	1.2427 (0.4985)	0.8809 (0.1773)	1.5890 (0.5805)	0.8307 (0.0887)
n	$\tilde{\theta}_{BS}$	$\tilde{\theta}_{BL}(c = 2)$	$\tilde{\theta}_{BL}(c = -1)$	$\tilde{\theta}_{BST}$
20	0.9278 (0.3286)	0.6439 (0.1504)	1.2305 (0.4117)	0.2962 (0.5003)
50	0.6139 (0.0982)	0.5716 (0.0879)	0.8692 (0.1119)	0.3392 (0.3692)
100	0.5183 (0.0667)	0.4743 (0.0543)	0.5876 (0.0790)	0.3755 (0.1382)

From Tables 1–3, it can be seen that the ERs of all estimates decrease as the sample sizes increase in all cases, as expected. The results shows that the performances of the Bayes estimators under various loss functions are different. Also, it is clear that the Bayes estimates under the LINEX loss function are sensitive to the values of the shape parameter c . It can be observed that the performances of Bayes estimates under the LINEX loss function for small values of c become close to the performances of Bayes estimates under the SEL function, as expected. In most of the cases, the performances of Bayes estimates under Stein loss function are not suitable especially in the estimation of parameter α .

Table 3: Average estimates and estimated risk (in parentheses) of the parameters with ($\alpha = 0.5$, $\beta = 0.25$, $\theta = 1$).

n	$\tilde{\alpha}_{BS}$	$\tilde{\alpha}_{BL}(c = 2)$	$\tilde{\alpha}_{BL}(c = -1)$	$\tilde{\alpha}_{BST}$
20	0.2596 (0.1168)	0.1656 (0.2420)	0.2807 (0.0826)	0.0189 (1.4314)
50	0.2974 (0.0863)	0.1891 (0.2278)	0.3234 (0.0677)	0.0257 (1.0636)
100	0.3308 (0.0634)	0.1995 (0.1834)	0.3759 (0.0439)	0.0408 (0.9023)
n	$\tilde{\beta}_{BS}$	$\tilde{\beta}_{BL}(c = 2)$	$\tilde{\beta}_{BL}(c = -1)$	$\tilde{\beta}_{BST}$
20	0.4227 (0.0536)	0.3752 (0.0885)	0.4605 (0.0405)	0.2917 (0.1419)
50	0.3539 (0.0304)	0.3360 (0.0447)	0.3701 (0.0248)	0.2853 (0.0982)
100	0.3204 (0.0127)	0.2985 (0.0258)	0.3439 (0.0109)	0.2770 (0.0789)
n	$\tilde{\theta}_{BS}$	$\tilde{\theta}_{BL}(c = 2)$	$\tilde{\theta}_{BL}(c = -1)$	$\tilde{\theta}_{BST}$
20	1.0294 (0.1186)	0.8665 (0.1483)	1.1403 (0.0847)	0.6568 (0.1774)
50	1.0156 (0.0669)	0.8893 (0.0973)	1.0874 (0.0509)	0.7663 (0.0936)
100	0.9965 (0.0417)	0.9205 (0.0679)	1.0390 (0.0281)	0.7889 (0.0806)

Discussion and Results

In this paper, the problem of Bayesian inference on the parameters of a three-parameters distribution, called the Lindley-Burr XII distribution, is considered. Bayesian estimation of the parameters under SEL, LINEX and Stein loss functions are obtained by using the MCMC methods. A real lifetime data set is analyzed for illustrative purposes. Finally, to investigate and compare the performance of different types of estimators presented in this paper a Monte Carlo simulation study is conducted. Based on simulation results, we observed that the performance of the Bayes estimators are different under various loss functions and the Bayes estimates under the LINEX loss function are sensitive to the values of the shape parameter c . It is clear that the performance of the Bayes estimators under SEL and LINEX loss functions are better than the Bayes estimators under Stein loss function in most of the cases.

Appendix

In Section 2, we used the observed information matrix I to find the standard error of the MLEs of the parameters of the L-BXII distribution. The elements of this matrix are given by

$$\begin{aligned}
I_{11} &= -\sum_{i=1}^n \frac{\beta\theta^2 h_i(1+x_i)x_i^{\alpha-1} \log x_i(2+\alpha \log x_i) - \beta^2 x_i^{2\alpha-2} h_i^2 + \theta^4(1+x_i)^2 x_i^\alpha \log^2 x_i}{g_i^2} \\
&\quad + (\beta+1) \sum_{i=1}^n \frac{x_i^\alpha \log^2 x_i}{s_i^2}, \\
I_{12} &= -\sum_{i=1}^n \frac{\theta^2 x_i^{\alpha-1} h_i(1+x_i)(s_i + \alpha \log x_i)}{g_i^2} + \sum_{i=1}^n \frac{x_i^\alpha \log x_i}{s_i}, \quad I_{22} = \sum_{i=1}^n \left(\frac{\alpha x_i^{\alpha-1} h_i}{g_i} \right)^2, \\
I_{13} &= \sum_{i=1}^n \frac{\beta\theta x_i^{\alpha-1}(1+x_i)(s_i + \alpha \log x_i)(h_i + 1)}{g_i^2}, \quad I_{23} = \sum_{i=1}^n \frac{\alpha\theta x_i^{\alpha-1}(1+x_i)s_i(h_i + 1)}{g_i^2}, \\
I_{33} &= \sum_{i=1}^n \frac{2\alpha\beta x_i^{\alpha-1}(1+x_i)s_i(h_i - 2) + (1+x_i)^2[2\theta^2 s_i^2 - \alpha^2 \beta^2 x_i^{2\alpha-2}]}{g_i^2} - \frac{n}{(\theta+1)^2},
\end{aligned}$$

where

$$g_i(\alpha, \beta, \theta) = \theta^2(1+x_i)(1+x_i^\alpha) + \alpha\beta x_i^{\alpha-1}(1+\theta+\theta x_i), \quad h_i(\theta) = 1+\theta+\theta x_i, \quad s_i(\alpha) = 1+x_i^\alpha.$$

References

- Alizadeh, M., Ozel, G., Altun, E., Abdi, M. and Hamedani, G. G. (2017), The odd log-logistic Marshall-Olkin Lindley model for lifetime data. *Journal of Statistical Theory and Applications*, **16**, 382-400.
- Alsadat, N., Nagarjuna, B. V., Hassan, A. S., Elgarhy, M., Ahmad, H. and Almetwally, H. M. (2023), Marshall-Olkin Weibull-Burr XII distribution with application to physics data. *AIP Advances*, **13**, 095325.
- Guerra, R. R., Peña-Ramírez, F. A., Cordeiro, G. M. (2023), The logistic Burr XII distribution: properties and applications to income data. *Stats*, **6**, 1260-1279.
- Makubate, B., Oluyede, B. and Gabanakgosi, M. (2021), A new Lindley-Burr XII distribution: model, properties and applications, *International Journal of Statistics and Probability*, **10**, 33-51.
- Maya, R. and Irshad, M. R. (2017), New extended generalized Lindley distribution: Properties and applications. *Statistica*, **77**, 33-52.
- Murthy, D. P., Xie, M. and Jiang, R. (2004). *Weibull Models*, John Wiley and Sons.
- Tierney, L. (1994), Markov chains for exploring posterior distributions. *The Annals of Statistics*, **22**, 1701-1728.
- Varian, H. R. (1975), A Bayesian approach to real estate assessment. In: Finberg, S. E. and Zellner, A. *Studies in Bayesian Econometrics and statistics in honor of Leonard J. Savege*, Amsterdam: North Holland, pp. 195-208.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Some Estimators for the Residual Varextropy Function of Length-biased Data

R. Zamini*

Department of Mathematics, Faculty of Mathematical Sciences and Computer, Kharazmi University, Tehran, Iran.

Abstract

In reliability theory and survival analysis, the residual extropy is proposed to measure the residual uncertainty of lifetimes. In this paper under length-biased sampling, we introduce two estimators for the variance of the residual lifetimes, "residual varextropy" in short. To assess the performance of the proposed estimators, we conduct a finite-sample simulation study using bias and root mean squared error (RMSE). We illustrate the application of the proposed estimators using real data analysis.

Keywords: Residual varextropy, Length-biased data, Kernel density.

Mathematics Subject Classification (2020): 62G05, 62G07, 94A17.

1 Introduction

Consider the non-negative continuous random variable X with the distribution function $F(\cdot)$ and the density function $f(\cdot)$. [Shannon \(1948\)](#) introduced the entropy associated with X as

$$H(x) = - \int_0^{+\infty} f(x) \log(f(x)) dx. \quad (1)$$

$H(X)$ has applications in various fields, including physics, biology and computer sciences. [Lad and et al. \(2015\)](#) introduced the dual of entropy, extropy, as:

$$J(X) = E(-\frac{1}{2}f(X)) = -\frac{1}{2} \int_0^{+\infty} (f(x))^2 dx. \quad (2)$$

Among other authors who have applied extropy in statistical contexts, we can refer to the work of [Qiu \(2017\)](#), [Raqab and Qiu \(2019\)](#) and [Gneiting and Raftery \(2007\)](#). [Qiu and Jia \(2018\)](#) defined a dynamic metric of extropy, called residual extropy, as follows:

$$J_t(X) = J(X_t) = -\frac{1}{2(\bar{F}(t))^2} \int_t^{+\infty} (f(x))^2 dx, \quad (3)$$

where $\bar{F}(t) = 1 - F(t)$ is the survival function of X . $J_t(X)$ measures uncertainty in the residual lifetime $X_t = (X - t | X > t)$ with density function $f_t(x) = \frac{f(x+t)}{\bar{F}(t)}$, $x > 0$. [Vaselabadi and et](#)

*Corresponding author, zamini@khu.ac.ir

[al. \(2021\)](#) introduced the variance of $-\frac{1}{2}f(X)$ as another uncertainty measure:

$$VJ(X) = \text{Var} \left(-\frac{1}{2}f(X) \right) = \frac{1}{4} \int (f(x))^3 dx - \left(\frac{1}{2} \int (f(x))^2 dx \right)^2. \quad (4)$$

$VJ(X)$ is called varextropy measure. This measure indicates how the information content is scattered around the extropy. [Goodarzi \(2022\)](#) investigated the relationship between series and parallel systems and symmetric distributions with the help of varextropy. [Alizadeh Noughabi and Shafaei Noughabi \(2024\)](#) proposed some estimators for varextropy. [Zamini and et al. \(2025\)](#) proposed some non-parametric estimators for varextropy, residual and past varextropy under α -mixing dependent condition. [Vaselabadi and et al. \(2021\)](#) introduced variance of $-\frac{1}{2}f(X_t)$ as the residual varextropy:

$$VJ_t(X) = VJ(X_t) = \frac{1}{4(\overline{F}(t))^3} \int_t^{+\infty} (f(x))^3 dx - \left(\frac{1}{2(\overline{F}(t))^2} \int_t^{+\infty} (f(x))^2 dx \right)^2. \quad (5)$$

This measure indicates how the information content is scattered around the residual extropy. In this paper, we study some estimators for $VJ_t(X)$ under length-biased data.

If observations are selected from the interest population with probability proportional to their length, rather than randomly, we will encounter length-biased data. Examples of such data can be found in studies related to cancer screening trails, as explained by [Zelen and Feinleib \(1969\)](#), biomedicine ([Chakraborty and Rao \(2000\)](#)), astronomy, renewal processes and econometrics.

Let X denote a non-negative random variable with density function $f(\cdot)$ and distribution function $F(\cdot)$. The associated length-biased random variable Y , characterized by the distribution function $G(\cdot)$, is as follows:

$$G(y) = \frac{1}{\mu} \int_0^y xf(x)dx, \quad y \geq 0, \quad (6)$$

where $\mu = \int_0^\tau xf(x)dx < \infty$ is its expected value and $\tau = \sup\{x; F(x) < 1\} < \infty$.

Under length-biased sampling, some kernel estimators for extropy and residual extropy were proposed by [Rajesh and et al. \(2022\)](#) and [Pavithradas and et al. \(2025\)](#), respectively. [Zamini and et al. \(2023\)](#) presented non-parametric estimators for varextropy under length-biased sampling. They investigated the consistency and asymptotic normality of the proposed estimators under suitable regularity conditions. In this paper, we address the problem of kernel estimators of residual varextropy function under length-biased sampling.

The rest of the paper is organized as follows. In Section 2, we propose two kernel estimators of the residual varextropy for length-biased data and present the main theorems. We investigate the finite sample behaviour of the proposed estimators and a real dataset of length-biased lifetimes in Section 3. The proof of the main results are given in Section 4. We give briefly the conclusions of this study and suggestions for the future application in Section 5.

2 Main Results

In this section we propose two kernel estimators for the residual varextropy under length-biased scheme and investigate some of their asymptotic properties. In the following, we list some necessary assumptions in this paper.

Assumptions

A1 The symmetric density function K is bounded variation on $(-1, 1)$.

A2 $\int_{-1}^1 tK(t) dt = 0$.

A3 $\int_{-1}^1 t^2K(t) dt < \infty$.

A4 $\int_{-1}^{+1} |dK(t)| < \infty$.

A5 $\int_0^{+\infty} u^{-2}G^{1/r}(u)du < \infty$, for some $r > 2$.

A6 $f(t) = O(t)$ as $t \rightarrow 0$.

Let X be the interest random variable with density and distribution functions $f(\cdot)$ and $F(\cdot)$ respectively. The associated length-biased random variable Y has the distribution function G . Cox (1969) proposed the following empirical estimator for F :

$$F_n(t) = \frac{\mu_n}{n} \sum_{i=1}^n \frac{1}{Y_i} I(Y_i \leq t), \quad (7)$$

where Y_1, \dots, Y_n is a sample of independent and identically distributed random variables from G and $\mu_n^{-1} = \frac{1}{n} \sum_{i=1}^n Y_i^{-1}$. Chaubey and et al. (2010) proved that:

$$\lim_{n \rightarrow \infty} \sup_{0 < x < \infty} |F_n(x) - F(x)| = 0, \text{ a.s.} \quad (8)$$

Jones (1991) proposed the following kernel estimator for f based on F_n as follows:

$$f_n(x) = \frac{\mu_n}{nh_n} \sum_{i=1}^n \frac{1}{Y_i} K\left(\frac{x - Y_i}{h_n}\right), \quad (9)$$

where $\{h_n\}_{n \geq 1}$ is a bandwidth sequence, satisfying the conditions $\lim_{n \rightarrow \infty} h_n = 0$ and $\lim_{n \rightarrow \infty} nh_n = \infty$. Under some conditions, Ajami and et al. (2013) showed that

$$\lim_{n \rightarrow \infty} \sup_{0 < x < \tau} |f_n(x) - f(x)| = 0, \text{ a.s.} \quad (10)$$

Based on f_n and F_n we propose an estimator for $VJ_t(X)$ as follows:

$$VJ_t^1(X) = \frac{1}{4(\overline{F_n(t)})^3} \int_t^{+\infty} \left(\tilde{f}_n(x)\right)^2 dF_n(x) - \left(J_n^1(f, t)\right)^2, \quad (11)$$

where

$$\tilde{f}_n(Y_i) = \frac{\mu_{n-1}}{(n-1)h_{n-1}} \sum_{j=1, j \neq i}^n \frac{1}{Y_j} K\left(\frac{Y_i - Y_j}{h_{n-1}}\right), \quad (12)$$

and

$$J_n^1(f, t) = -\frac{\mu_n}{2n(\overline{F_n(t)})^2} \sum_{i=1}^n Y_i^{-1} \tilde{f}_n(Y_i) I(Y_i \geq t), \quad (13)$$

is the first estimator of $J_t(X)$ proposed by Pavithradas and et al. (2025).

We obtain another estimator for $VJ_t(X)$ as follows:

$$VJ_t^2(X) = \frac{1}{4(\overline{F_n(t)})^3} \int_t^{+\infty} (f_n(x))^3 dx - \left(J_n^2(f, t)\right)^2, \quad (14)$$

where

$$J_n^2(f, t) = -\frac{1}{2(\overline{F_n(t)})^2} \int_t^{\infty} (f_n(x))^2 dx, \quad (15)$$

is the second estimator of $J_t(X)$ proposed by Pavithradas and et al. (2025).

In the following theorems, consistency of the $VJ_t^1(X)$ and $VJ_t^2(X)$ are investigated.

Theorem 2.1. Let $\lim_{n \rightarrow \infty} \frac{\log n}{n^{\frac{1}{2} + \lambda} h_n} = 0$ for any $0 < \lambda < \frac{1}{2} - \frac{1}{r}$ and some $r > 2$. If f is bounded, then under Assumptions $A_1 - A_4$ we can write:

$$\lim_{n \rightarrow \infty} VJ_t^1(X) = VJ_t(X), \text{ a.s.} \quad (16)$$

Proof. See the Appendix □

Theorem 2.2. Under conditions of Theorem 2.1, we can write:

$$\lim_{n \rightarrow \infty} VJ_t^2(X) = VJ_t(X), \text{ a.s.} \quad (17)$$

Proof. See the Appendix □

Remark 2.3. Let $f \in \mathbb{F}$ where \mathbb{F} is the class of continuous density functions f defined on $[0, 1]$ with distribution function F . If there is a $0 < t_0 < 1$ for which $f_{t_0}(x) = \frac{f(x)}{F(t_0)} = \frac{1}{1-t_0}$, $x > t_0$, then $VJ_{t_0}(X) = \text{var}(-\frac{1}{2}f_{t_0}(X)) = 0$. On the other hand, if $VJ_{t_0}(X) = \text{var}(-\frac{1}{2}f_{t_0}(X)) = 0$, then we can see that $f_{t_0}(x) = \frac{f(x)}{F(t_0)} = c$, a.s., $\forall t_0 < x < 1$, where c is a constant. From the property of density function, $\int_{t_0}^1 f_{t_0}(x)dx = 1$, we can write $f_{t_0}(x) = \frac{f(x)}{F(t_0)} = \frac{1}{1-t_0}$, $\forall x > t_0$. This is equivalent to the hazard rate function $\lambda(x) = \frac{f(x)}{F(x)} = \frac{\bar{F}(t_0)}{(1-t_0)\bar{F}(x)}$ is increasing for $\forall x > t_0$. In other words, t_0 is the starting time that the unite in the system is becoming more and more prone to failure. Therefore, the residual varextropy can be used as a way to detect when systems start to fail.

3 Simulation Results

To demonstrate the performance of the proposed estimators, we conduct simulation studies in two following scenarios.

Scenario 1. Let X be a random variable having a beta distribution with density function:

$$f(x) = 6x(1-x), \quad 0 < x < 1.$$

Scenario 2. Let X be a random variable having a gamma distribution with density function:

$$f(x) = xe^{-x}, \quad 0 < x < \infty.$$

We use the notations $B(\alpha, \beta)$ and $G(\alpha, \beta)$ to indicate X has the beta and gamma distribution, respectively. We note that if X has $B(2, 2)$, then the corresponding length-biased form is $B(3, 2)$ and if X has $G(2, 1)$, then the corresponding length-biased form is $G(3, 1)$. We use 500 samples of various sample sizes $n = 50, 100, 500$ and the Epanechnikov density function $K(x) = \frac{3}{4}(1-x^2)I(|x| < 1)$ for these simulations. We select h_n by using the rule of thumb bandwidth selection technique for length-biased data. (See [Borrajao and et al. \(2017\)](#) for more details.) We compute the absolute values of bias (B_1 and B_2) and RMSE (RMSE1 and RMSE2) of the proposed estimators for various values of $u = F(t)$ and sample sizes. The results are given in Tables 1 and 2. From tables we can observe that as the sample sizes increase, B_1 , B_2 , RMSE1 and RMSE2 decrease. In beta distribution, VJ^1 generally has better performance than VJ^2 in terms of bias and RMSE. In gamma distribution, for $u = 0.2$, VJ^2 performs better in terms of bias and RMSE and for $u = 0.5, 0.8$, VJ^1 has better performance in terms of bias and RMSE. In the following, we assess the performance of the estimators using real data. [Helu and et al. \(2020\)](#) used 70 failure times of aircraft windshields with a unit of measurement of 1000h. They showed that these data have gamma distribution with parameters $\alpha = 7.75$ and $\beta = 0.285$. [Rajesh and et al. \(2022\)](#) modified these data and obtained the following length-biased sample:

3.000	2.83	2.224	1.505	3.166	1.568	1.281	2.135	2.194	3.478
3.117	3.344	2.038	2.661	1.899	2.385	1.981	0.557	2.688	1.876
1.303	3.114	2.097	3.443	2.962	2.632	3.595	2.610	1.615	2.646
2.300	2.089	2.229	3.103	1.652	1.914	1.757	3.385	1.911	2.481
1.506	2.085	1.652	2.223	2.324	1.619	3.578	2.349	1.866	1.912

In the second real dataset, we assess the performance of the estimators using real data provided by [Lawless \(2011\)](#). He used 98 lifetimes of brake pads with a unit of measurement of 1000h. He showed that these data have log-normal distribution with parameters $\mu = 4.109$ and $\sigma = 0.421$. Tables 3 and 4 show values of the residual varextropy and its estimators for the first and the second real data, respectively. From the Tables 3 and 4, we can observe that the residual values obtained using VJ_t^1 and VJ_t^2 are close to the residual varextropy VJ_t .

Table 1: Bias and RMSE of B(2,2).

n	u	VJ^1	B1	RMSE1	VJ^2	B2	RMSE2
50	0.2	0.0384	0.0081	0.0205	0.0551	0.0249	0.0364
50	0.5	0.1044	0.0016	0.0573	0.1371	0.0342	0.0819
50	0.8	0.4464	0.0278	0.3316	0.2404	0.2337	0.7234
100	0.2	0.0346	0.0044	0.0149	0.0497	0.0195	0.0262
100	0.5	0.1018	0.0010	0.0450	0.1328	0.0300	0.0632
100	0.8	0.4407	0.0334	0.2467	0.4139	0.0603	0.3035
500	0.2	0.0317	0.0015	0.0069	0.0407	0.0104	0.0131
500	0.5	0.1034	0.0006	0.0230	0.1226	0.0197	0.0332
500	0.8	0.4419	0.0322	0.1168	0.4953	0.0211	0.1207

Table 2: Bias and RMSE of G(2,1).

n	u	VJ^1	B1	RMSE1	VJ^2	B2	RMSE2
50	0.2	0.0024	0.0024	0.0027	0.0031	0.0017	0.0020
50	0.5	0.0063	0.0019	0.0035	0.0048	0.0034	0.0041
50	0.8	0.0146	0.0030	0.0091	0.0022	0.0093	0.0202
100	0.2	0.0027	0.0021	0.0023	0.0034	0.0014	0.0016
100	0.5	0.0068	0.0014	0.0025	0.0057	0.0024	0.0029
100	0.8	0.0138	0.0021	0.0056	0.0075	0.0040	0.0059
500	0.2	0.0036	0.0013	0.0014	0.0040	0.0008	0.0010
500	0.5	0.0076	0.0006	0.0012	0.0070	0.0012	0.0015
500	0.8	0.0127	0.0011	0.0026	0.0104	0.0011	0.0021

4 Appendix

The proofs of theorems are given in this section.

Proof of Theorem 2.1.

$$VJ_t^1(X) - VJ_t(X) = I'_n + II'_n, \quad (18)$$

where

$$I'_n = \left(\frac{1}{4(\vec{F}_n(t))^3} \int_t^\tau (f_n(x))^2 dF_n(x) - \frac{1}{4(\vec{F}(t))^3} \int_t^\tau (f(x))^2 dF(x) \right),$$

and

$$II'_n = \left((J_t(X))^2 - (J_n^1(f, t))^2 \right).$$

From Theorem 1 of Pavithradas and *et al.* (2025), we can observe that

$$\lim_{n \rightarrow \infty} II'_n = 0, \text{ a.s.} \quad (19)$$

On the other hand

$$I'_n = \frac{A_n}{4} \int_t^\tau (f_n(x))^2 dF_n(x) + \frac{1}{4(\vec{F}(t))^3} \left[\int_t^\tau (f_n(x))^2 dF_n(x) - \int_t^\tau (f(x))^2 dF(x) \right], \quad (20)$$

where

$$A_n = \frac{1}{(\vec{F}_n(t))^3} - \frac{1}{(\vec{F}(t))^3}. \quad (21)$$

Table 3: Values of estimators for aircraft windshields real data.

u	VJ_t	VJ^1	VJ^2
0.2	0.00005790913	0.00008251021	0.00006029210
0.5	0.0001446675	0.0001892130	0.0001952106
0.8	0.0002971429	0.0003312519	0.0004151892

Table 4: Values of estimators for brake pads real data.

u	VJ_t	VJ^1	VJ^2
0.2	0.00001049979	0.000009546706	0.0000136431
0.5	0.00002125407	0.00002558203	0.00002309441
0.8	0.00003112075	0.00009199287	0.00003524333

Using integration by parts we can write:

$$\begin{aligned}
 \left| \int_t^\tau (f_n(x))^2 dF_n(x) - \int_t^\tau (f(x))^2 dF(x) \right| &= \left| \int_t^\tau ((f_n(x))^2 - (f(x))^2) dF_n(x) \right. \\
 &\quad \left. - \int_t^\tau (f(x))^2 d(F_n(x) - F(x)) \right| \\
 &= \left| \int_t^\tau ((f_n(x))^2 - (f(x))^2) dF_n(x) \right. \\
 &\quad \left. - f^2(t)(F_n(t) - F(t)) - \int_t^\tau (F_n(x) - F(x)) d(f^2(x)) \right| \\
 &\leq \sup_{0 < x < \tau} |(f_n(x))^2 - (f(x))^2| \\
 &\quad + 3M^2 \sup_{0 < x < \infty} |F_n(x) - F(x)|, \tag{22}
 \end{aligned}$$

where $M = \sup_x f(x) < \infty$. (22), Equation (23) of [Zamini and et al. \(2023\)](#), (8) and (20) ensure that

$$\lim_{n \rightarrow \infty} I'_n = 0, \text{ a.s.} \tag{23}$$

Now, (18), (19) and (23) complete the proof. \square

Proof of Theorem 2.2.

$$VJ_t^2(X) - VJ_t(X) = I_n + II_n, \tag{24}$$

where

$$I_n = \left(\frac{1}{4(\vec{F}_n(t))^3} \int_t^\tau (f_n(x))^3 d(x) - \frac{1}{4(\vec{F}(t))^3} \int_t^\tau (f(x))^3 d(x) \right),$$

and

$$II_n = \left((J_t(X))^2 - (J_n^2(f, t))^2 \right).$$

From Theorem 3 of [Pavithradas and et al. \(2025\)](#), we can write:

$$\lim_{n \rightarrow \infty} II_n = 0, \text{ a.s.} \tag{25}$$

Next, to deal with I_n , observe that

$$I_n = \frac{An}{4} \int_t^\tau (f_n(x))^3 d(x) + \frac{1}{4(\vec{F}(t))^3} \left[\int_t^\tau (f_n(x))^3 d(x) - \int_t^\tau (f(x))^3 d(x) \right]. \tag{26}$$

It is easy to see that

$$\begin{aligned}
\left| \int_t^\tau (f_n(x))^3 d(x) - \int_t^\tau (f(x))^3 d(x) \right| &\leq \left| \int_t^\tau ((f_n(x)) - (f(x)))(f_n(x))^2 d(x) \right| \\
&+ \left| \int_t^\tau ((f_n(x))^2 - (f(x))^2) d(F(x)) \right| \\
&\leq \sup_{0 < x < \tau} |(f_n(x)) - (f(x))| \int_t^\tau (f_n(x))^2 dx \\
&+ \sup_{0 < x < \infty} |(f_n(x))^2 - (f(x))^2| \\
&\leq \sup_{0 < x < \tau} |(f_n(x)) - (f(x))| \\
&\times \left[\int_t^\tau |(f_n(x))^2 - (f(x))^2| dx + \int_t^\tau (f(x))^2 dx \right] \\
&+ \sup_{0 < x < \tau} |(f_n(x))^2 - (f(x))^2| \\
&< \sup_{0 < x < \tau} |(f_n(x)) - (f(x))| \\
&\times \left[\tau \sup_{0 < x < \tau} |(f_n(x))^2 - (f(x))^2| + \tau M^2 \right] \\
&+ \sup_{0 < x < \tau} |(f_n(x))^2 - (f(x))^2|, \tag{27}
\end{aligned}$$

where $M = \sup_x f(x) < \infty$. Equation (23) of Zamini and *et al.* (2023), (10), (27), (26) and (8) ensure that

$$\lim_{n \rightarrow \infty} I_n = 0, \text{ a.s.} \tag{28}$$

Thus, (24), (25) and (28) complete the proof. \square

Discussion and Results

We proposed some non-parametric kernel estimator for the residual varextropy under length-biased sampling and proved some asymptotic properties for the proposed estimators. Monte Carlo simulation studies were carried out to evaluate the performance of the estimators using the bias and RMSE. A real dataset was used to evaluate the performance of the residual varextropy estimators.

Acknowledgment

The author thanks the editor and referees for their comments and helpful suggestions which helped to improve the presentation.

References

- Ajami M., Fakoor V. and Jomhoori S. (2013), Some asymptotic results of kernel density estimator in length-biased sampling, *J.Sci. Islamic Republic of Iran*, **24**(1), 55-62.
- Alizadeh Noughabi H. and Shafaei Noughabi M. (2024), On the estimation of varextropy, *Statistics: A journal of Theoretical and Applied Statistics*, **58**(4), 1-15.
- Borrajó M. I., Gonzalez-Manteiga W. and Martinez-Miranda M. (2017), Bandwidth selection for kernel density estimation with length-biased data, *J. Nonparametric Stat.*, **29**(3), 636-668.
- Chakraborty R. and Rao C.R. (2000), 23 selection biases of samples and their resolutions, *Handbook of Statistics*, **18**, 675-712.
- Cox D. (1969), Some sampling problems in technology, In: *New developments in survey sampling*. Wiley Interscience, New York, 506-527.

- Chaubey Y., Sen P. and Li J. (2010), Smooth density estimation for length-biased data, *J. Indian Soc. Agri. Stat.*, **64**(2), 145-155.
- Gneiting T. and Raftery A.E. (2007), Strictly proper scoring rules, prediction, and estimation, *J. Am. Stat. Assoc.*, **102**(477), 359-378.
- Goodarzi F. (2022), Characterizations of some discrete distributions and upper bounds on discrete residual varextropy, *JIRSS*, **21**(2), 231-248.
- Helu A., Samawi H., Rochani H., Yin J. and Vogel R. (2020), Kernel density estimation based on progressive type-II censoring, *J. Korean Stat. Soc.*, **49**(2), 475-498.
- Jones M. C. (1991), Kernel density estimation for length-biased data, *Biometrika*, **78**(3), 511-519.
- Lad F., Sanfilippo G. and Agro G. (2015), Extropy: complementary dual of entropy, *Stat. Sci.*, **30**(1), 40-58.
- Lawless J. F. (2011), *Statistical models and methods for lifetime data*, Hoboken, NJ, USA: Wiley.
- Pavithradas V., Rajesh G. and Rajesh R. (2025), Nonparametric estimation of residual extropy function under length-biased sampling, *Ricerche di Matematica*, 1-17.
- Qiu G. (2017), The extropy of order statistics and record values, *Statistics and Probability Letters*, **120**(C), 52-60.
- Qiu G. and Jia K. (2018), The residual extropy of order statistics, *Statistics and Probability Letters*, **133**, 15-22.
- Raqab M. Z. and Qiu G. (2019), On entropy properties of ranked set sampling, *Statistics: A Journal of Theoretical and Applied Statistics*, **53**(1), 210-226.
- Rajesh R., Rajesh G. and Sunoj S. (2022), Kernel estimation of extropy function under length-biased sampling, *Stat. Prob. Lett.*, **181**, 8684-8693.
- Shannon C. (1998), A mathematical theory of communication, *Bell Syst. Tech. J.*, **27**(3), 379-423.
- Vaselabadi N. M., Tahmasebi S., Kazemi M. R. and Buono F. (2021), Results on varextropy measure of random variables, *Entropy*, **23**(3), 356.
- Zamini R., Goodarzi F. and Salimi M. (2025), Nonparametric estimators for varextropy under α -mixing condition with application in exponential $AR(1)$ model, *J. Mahani Math. Res.*, **14**(1), 45-61.
- Zamini R., Goodarzi F. and Hashemi F. (2023), Some kernel estimators for varextropy function under length-biased sampling, *Communications in Statistics-Simulation and Computation*, 1-21.
- Zelen M. and Feinleib M. (1969), On the theory of screening for chronic diseases, *Biometrika*, **56**(3), 601-614.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Pathwise Grid Valuation of Fixed-Income Portfolios with Applications to Risk Management

Shiva Zamani ^{*1}, Ali Chaghazardi ¹, Hamidreza Arian ¹

¹RiskLab, GSME, Sharif University of Technology

Abstract

Numerical calculation of Value-at-Risk (VaR) for large-scale portfolios poses great challenges to financial institutions. The problem is even more daunting for large fixed-income portfolios as their underlying instruments have exposure to higher dimensions of risk factors. This article provides an efficient algorithm for calculating VaR using a historical grid-based approach with volatility updating and shows its efficiency in computational cost and accuracy. Our VaR computation algorithm is flexible and simple, while one can easily extend it to cover other nonlinear portfolios such as derivative portfolios on equities and FX securities.

Keywords: Financial Risk Management, Value-at-Risk (VaR), Profit and Loss (PnL) Calculation, Scenario Analysis, Fixed-Income Portfolio.

^{*}Corresponding author, zamani@sharif.edu



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Modified Maximum Likelihood-Based Test for Nonlinear Time Series Model Selection

Sedigheh Zamani Mehreyan*

¹Department of Statistics, Imam Khomeini International University, Qazvin, Iran.

Abstract

We considered the symmetric and asymmetric types of the smooth transition autoregressive model which are suitable for modelling economic and financial time series. We derived modified maximum likelihood estimators of parameters and computed asymptotic distribution of modified maximum likelihood estimators. So that, we can derive asymptotic distribution of Vuong's test statistic for stationary smooth transition autoregressive models. A set of simulation results also leads strong support to the results presented in the paper. Using simulation, it shows that Vuong's test work to select the optimal stationary smooth transition autoregressive models.

Keywords: Model selection, Smooth transition autoregressive model, Vuong's test.

Mathematics Subject Classification (2020): 60Fxx, 62M10, 91B84.

1 Introduction

Often, in data analysis we seek to determine the underlying model or rule that is capable of describing the data. The objective of model selection is to select a model that is closest to true model. [Vuong \(1989\)](#) proposed model selection test for separated and non-nested models based on the likelihood ratio. [Shi \(2015\)](#) proposed a one-step nondegenerate test as an alternative to the classical Vuong's test. [Schneider et al. \(2019\)](#) applied Vuong's general approach of model selection to the comparison of nested and non-nested dimensional and multidimensional item response theory models. [Liao and Shi \(2020\)](#) proposed a new model selection test for the statistical comparison of semi/non-parametric models based on a general quasi-likelihood ratio criterion.

Nonlinear time series models have recently found widespread application in analysis of economics and finance time series. [Chan and Tong \(1986\)](#) used the cumulative function of the standard Normal variable as the transition function and introduced the smooth transition autoregressive, STAR, model as nonlinear time series model. The logistic autoregressive model and exponential autoregressive model are most popular specifications of the STAR model. Logistic smooth transition autoregressive, LSTAR, model is characterized by the asymmetric properties which make it suitable for modelling specific economics and finance time series, see

*Corresponding author, s.zamani@sci.ikiu.ac.ir

Yaya and Shitu (2016). The nature of numerical difficulties using Monte Carlo simulation is studied by Chan and Theoharakis (2011). They shown that the conventional optimization algorithms do not perform well in locating the global optimum of the associated likelihood function. Schleer (2016) studied the starting-values for estimation of vector STAR models based on a Monte Carlo method.

In this paper, we considered the smooth transition autoregressive model and estimated unknown parameters based on the modified maximum likelihood method and proposed Vuong's test as model selection test.

The rest of the paper is structured as follows: in Section 2 The unknown parameters of the smooth transition autoregressive model are estimated based on the modified maximum likelihood method and Vuong's test is proposed as model selection test. In Section 3, we study the obtained theoretical results by a simulation study.

2 Vuong's Test for Nonlinear Time Series Models

Consider the smooth transition autoregressive, STAR, model as:

$$y_t = \alpha y_{t-1} + \phi y_{t-1} G(s_t, \gamma, c) + \epsilon_t, t = 2, \dots, T \quad (1)$$

where error term ϵ_t distributed as independently and identically with mean 0 and variance σ^2 and $G(s_t, \gamma, c)$ is a transition function. It usually is a continuous function, bounded between 0 and 1 and thus allowing for a smooth transition between regimes. It is assumed that the transition variable s_t be a lagged endogenous variable, i.e. $s_t = y_{t-d}$. The value of d is varied in order to improve nonlinearity in the system when it is not known prior to model estimation. The transition function $G(s_t, \gamma, c)$ causes the nonlinear dynamics in the model. The U-shaped exponential function:

$$G(s_t, \gamma, c) = 1 - \exp \{ -\gamma (s_t - c)^2 \}, \gamma > 0 \quad (2)$$

and the first-order logistic function:

$$G(s_t, \gamma, c) = (1 + \exp \{ -\gamma (s_t - c) \})^{-1}, \gamma > 0 \quad (3)$$

are the most popular specifications of the transition function $G(\cdot)$. When $\gamma \rightarrow 0$ the STAR model with exponential transition function, ESTAR, and the STAR model with logistic transition function, LSTAR, converge to an autoregressive model. For large values of γ , the exponential function converges to one for values of s_t below or above threshold parameter c . Consider model (1), where error term ϵ_t distributed as Normal distribution, $N(0, \sigma^2)$. The log-likelihood of y_2, \dots, y_T conditional on y_1 is obtained as:

$$l(\theta) = -\frac{T}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=2}^T \epsilon_t^2 = -\frac{T}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=2}^T (y_t - \alpha y_{t-1} - \phi y_{t-1} G(s_t, \gamma, c))^2$$

To calculate the maximum likelihood estimators, we solve the estimating equations

$$\begin{aligned} \frac{\partial l(\theta)}{\partial \alpha} &= \sum_{t=2}^T \sigma^{-2} y_{t-1} (y_t - \alpha y_{t-1} - \phi y_{t-1} G(s_t, \gamma, c)) = 0 \\ \frac{\partial l(\theta)}{\partial \phi} &= \sum_{t=2}^T \sigma^{-2} G(s_t, \gamma, c) y_{t-1} (y_t - \alpha y_{t-1} - \phi y_{t-1} G(s_t, \gamma, c)) = 0 \\ \frac{\partial l(\theta)}{\partial \sigma^2} &= -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=2}^T \epsilon_t^2 = 0 \\ \frac{\partial l(\theta)}{\partial \gamma} &= \sum_{t=2}^T \sigma^{-2} (\phi y_{t-1}) \frac{\partial G(s_t, \gamma, c)}{\partial \gamma} (y_t - \alpha y_{t-1} - \phi y_{t-1} G(s_t, \gamma, c)) = 0 \end{aligned} \quad (4)$$

and

$$\frac{\partial l(\theta)}{\partial c} = \sum_{t=2}^T \sigma^{-2} (\phi y_{t-1}) \frac{\partial G(s_t, \gamma, c)}{\partial c} (y_t - \alpha y_{t-1} - \phi y_{t-1} G(s_t, \gamma, c)) = 0 \quad (5)$$

The likelihood equations (4 and 5) don't have explicit solution and have to be solved by iterative methods. To formulate modified likelihood equation, we define

$$M_1(s_t, \gamma, c) = \phi y_{t-1} \frac{\partial G(s_t, \gamma, c)}{\partial \gamma} (y_t - \alpha y_{t-1} - \phi y_{t-1} G(s_t, \gamma, c))$$

and

$$M_2(s_t, \gamma, c) = \phi y_{t-1} \frac{\partial G(s_t, \gamma, c)}{\partial c} (y_t - \alpha y_{t-1} - \phi y_{t-1} G(s_t, \gamma, c)).$$

The first two terms of Taylor expansion of $M_1(s_t, \gamma, c)$ and $M_2(s_t, \gamma, c)$ around $\gamma = 0$ are

$$M_1(s_t, \gamma, c) \cong M_1(\mu_s, \gamma, c) |_{\gamma=0} + M_1'(\mu_s, \gamma, c) |_{\gamma=0} (\gamma - 0)$$

and

$$M_2(s_t, \gamma, c) \cong M_2(\mu_s, \gamma, c) |_{\gamma=0} + M_2'(\mu_s, \gamma, c) |_{\gamma=0} (\gamma - 0),$$

where $M_1'(\mu_s, \gamma, c)$ and $M_2'(\mu_s, \gamma, c)$ are the derivative of $M_1(s_t, \gamma, c)$ and $M_2(s_t, \gamma, c)$ with respect to gamma, respectively. Substituted the Taylor expansion of $M_1(s_t, \gamma, c)$ and $M_2(s_t, \gamma, c)$ in (4 and 5) respectively and obtain modified maximum likelihood estimators by solving the estimating equations as:

$$\hat{\alpha} = \left(\sum_{t=2}^T y_{t-1}^2 \right)^{-1} \sum_{t=2}^T y_{t-1} (y_t - \hat{\phi} y_{t-1} G(s_t, \hat{\gamma}, \hat{c})) \quad (6)$$

$$\hat{\phi} = \left(\sum_{t=2}^T y_{t-1}^2 G^2(s_t, \hat{\gamma}, \hat{c}) \right)^{-1} \sum_{t=2}^T y_{t-1} G(s_t, \hat{\gamma}, \hat{c}) (y_t - \hat{\alpha} y_{t-1}) \quad (7)$$

and

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=2}^T \epsilon_t^2. \quad (8)$$

If the transition function be logistic function

$$G(s_t, \gamma, c) = (1 + \exp\{-\gamma(s_t - c)\})^{-1} - \frac{1}{2},$$

so

$$\begin{aligned} \frac{\partial l(\theta)}{\partial \gamma} &= \sum_{t=2}^T \sigma^{-2} \phi y_{t-1} (s_t - c) e^{-\gamma(s_t - c)} \left(\frac{y_t - \alpha y_{t-1} - \phi y_{t-1} \left[(1 + e^{-\gamma(s_t - c)})^{-1} - \frac{1}{2} \right]}{(1 + e^{-\gamma(s_t - c)})^2} \right) \\ &\cong \sum_{t=2}^T \sigma^{-2} (\phi y_{t-1}) \left(\frac{1}{4} \right) (s_t - c) \left(y_t - \alpha y_{t-1} - (\phi y_{t-1}) \left(-\frac{\gamma}{4} (s_t - c) \right) \right) \\ &= 0 \end{aligned}$$

and

$$\begin{aligned} \frac{\partial l(\theta)}{\partial c} &= -2\gamma \sum_{t=2}^T \sigma^{-2} \phi y_{t-1} e^{-\gamma(s_t - c)} \left(\frac{y_t - \alpha y_{t-1} - \phi y_{t-1} \left[(1 + e^{-\gamma(s_t - c)})^{-1} - \frac{1}{2} \right]}{(1 + e^{-\gamma(s_t - c)})^2} \right) \\ &\cong -\frac{\gamma}{2} \sum_{t=2}^T \sigma^{-2} \phi y_{t-1} \left(y_t - \alpha y_{t-1} - \phi y_{t-1} \left(-\frac{\gamma}{4} (s_t - c) \right) \right) \\ &= 0. \end{aligned}$$

Then the modified maximum likelihood estimator of γ and c are

$$\hat{\gamma} = \frac{4 \sum_{t=2}^T \hat{\phi} y_{t-1} (s_t - \hat{c}) (\hat{\alpha} y_{t-1} - y_t)}{\sum_{t=2}^T (\hat{\phi} y_{t-1})^2 (s_t - \hat{c})^2} \quad (9)$$

and

$$\hat{c} = \frac{4 \sum_{t=2}^T \hat{\phi} y_{t-1} \left(y_t - \hat{\alpha} y_{t-1} + \frac{\hat{\gamma} \hat{\phi}}{4} s_t y_{t-1} \right)}{\gamma \sum_{t=2}^T \left(\hat{\phi} y_{t-1} \right)^2}. \quad (10)$$

If the transition function be exponential function

$$G(s_t, \gamma, c) = 1 - \exp \left\{ -\gamma (s_t - c)^2 \right\},$$

so

$$\begin{aligned} \frac{\partial l(\theta)}{\partial \gamma} &= \sum_{t=2}^T \sigma^{-2} \phi y_{t-1} (s_t - c)^2 e^{-\gamma (s_t - c)^2} \left(y_t - \alpha y_{t-1} - \phi y_{t-1} \left(1 - e^{-\gamma (s_t - c)^2} \right) \right) \\ &\cong \sum_{t=2}^T \sigma^{-2} \phi y_{t-1} (s_t - c)^2 (y_t - \alpha y_{t-1} - \phi y_{t-1}) (1 - \gamma (s_t - c)^2) + \sum_{t=2}^T \sigma^{-2} (\phi y_{t-1})^2 (s_t - c)^2 (1 - 2\gamma (s_t - c)^2) \\ &= 0 \end{aligned}$$

and

$$\begin{aligned} \frac{\partial l(\theta)}{\partial c} &= \gamma \sum_{t=2}^T \sigma^{-2} \phi y_{t-1} (s_t - c) e^{-\gamma (s_t - c)^2} \left(y_t - \alpha y_{t-1} - \phi y_{t-1} \left(1 - e^{-\gamma (s_t - c)^2} \right) \right) \\ &\cong \sum_{t=2}^T \sigma^{-2} \phi y_{t-1} (s_t - c) (y_t - \alpha y_{t-1} - \phi y_{t-1}) (1 - \gamma (s_t - c)^2) + \sum_{t=2}^T \sigma^{-2} (\phi y_{t-1})^2 (s_t - c) (1 - 2\gamma (s_t - c)^2) \\ &= 0. \end{aligned} \quad (11)$$

Then the modified maximum likelihood estimator of γ is

$$\hat{\gamma} = \frac{\sum_{t=2}^T \hat{\phi} y_{t-1} (s_t - \hat{c})^2 (y_t - \hat{\alpha} y_{t-1})}{\sum_{t=2}^T \hat{\phi} y_{t-1} (s_t - \hat{c})^2 (y_t - \hat{\alpha} y_{t-1} + \hat{\phi} y_{t-1}) (s_t - \hat{c})^2} \quad (12)$$

and the estimator of c can be calculated by solving the cubic equation of (11). The modified maximum likelihood estimators follow the asymptotic properties of maximum likelihood estimators (such as consistency and asymptotic normal distribution).

Theorem 2.1. (*Asymptotic Distribution of the LR Statistic*): Given assumptions of [Vuong \(1989\)](#): If $f^{B*}(\cdot) = g^{\beta*}(\cdot)$ then

$$2LR_T(\hat{B}, \hat{\beta}) \xrightarrow{D} M_m(\cdot; \lambda_*),$$

where $M_m(\cdot; \lambda_*)$ is a weighted sum of chi-squares with some parameters m and λ_* . The parameter λ_* is the vector of eigenvalues of

$$W = \begin{pmatrix} -D_f(B_*)C_f^{-1}(B_*) & -D_{fg}(B_*, \beta_*)C_g^{-1}(\beta_*) \\ D_{gf}(\beta_*, B_*)C_f^{-1}(B_*) & D_g(\beta_*)C_g^{-1}(\beta_*) \end{pmatrix}.$$

If $f^{B*}(\cdot) \neq g^{\beta*}(\cdot)$ then

$$T^{-\frac{1}{2}} LR_T(\hat{B}, \hat{\beta}) - T^{\frac{1}{2}} E_h \left[\log \frac{f^{B*}(Y)}{g^{\beta*}(Y)} \right] \xrightarrow{D} N(0, \omega_*^2)$$

where $\omega_*^2 = Var_h \left[\log \frac{f^{B*}(Y)}{g^{\beta*}(Y)} \right]$.

3 Simulation Analysis

In this section, we examine by simulation the relative performance of the material presented so far in the paper. In particular, we examine more closely the performance of the modified maximum likelihood method for estimation and also the performance of the model selection test. We consider exponential and logistic smooth transition autoregressive models as true models. The observations are generated from true model $y_t = \alpha y_{t-1} + \phi y_{t-1} G(y_{t-1}, \gamma, c) + \epsilon_t$, where $\theta_0 = (\alpha, \phi, \gamma, c) = (0.3, -0.5, 2.0, 0.4)$ and error term ϵ_t distributed as Normal distribution, $N(0, 1)$. Here we assume that the true model is known and only the parameters need to

Table 1: The MMLE of LSTAR and ESTAR models.

	T	$\hat{\alpha}$	$\hat{\phi}$	$\hat{\gamma}$	\hat{c}
LSTAR:					
	50	0.2457	-0.6143	1.7423	0.3416
	150	0.2597	-0.6005	1.8696	0.3962
	300	0.2681	-0.5599	1.9089	0.3971
	500	0.2882	-0.5490	1.9487	0.4025
	1000	0.3068	-0.5181	1.9865	0.4067
ESTAR :					
	50	0.2199	-0.6616	1.6293	0.5278
	150	0.2437	-0.5463	1.7217	0.5192
	300	0.2623	-0.5454	1.8414	0.4835
	500	0.2826	-0.5318	1.9593	0.4673
	1000	0.3055	-0.5046	1.9901	0.4622

be estimated. The results for all estimation procedure are given for different sample sizes, of $n=50, 150, 300, 500, 1000$ and are summarized in Table 1. It shows that, as the sample size increase, the value of estimators become closer to the true parameters.

We perform 10^3 replications. The results for all estimation procedure and their root mean square error, $RMSE$, are given for different sample sizes of $n=50, 150, 300, 500, 1000$ and are summarized in Table 2 and Table 3. In Table 2 the mean of modified maximum likelihood estimation of parameters are given. In table 3, we present the corresponding root of mean squared error vis-a-vis the true parameters

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (\theta_0 - \hat{\theta}_i)^2}$$

where $m = 10^3$. It shows that, as the sample size increase the $RMSE$ decrease and the value of estimators are closed to the true parameters.

Table 2: The MMLE of LSTAR and ESTAR models with 10^3 replications.

	T	$\hat{\alpha}$	$\hat{\phi}$	$\hat{\gamma}$	\hat{c}
LSTAR:					
	50	0.3452	-0.5912	2.4118	0.4855
	150	0.3270	-0.5354	2.3457	0.4315
	300	0.3197	-0.5183	2.2189	0.4288
	500	0.3184	-0.5129	2.1865	0.4192
	1000	0.3037	-0.5083	2.0715	0.4108
ESTAR:					
	50	0.2645	-0.5407	2.5156	0.5342
	150	0.2754	-0.5395	2.4953	0.4841
	300	0.2859	-0.5154	2.3468	0.4384
	500	0.2932	-0.5023	2.1524	0.4208
	1000	0.3024	-0.5022	2.0654	0.4121

The results in Tables 1-3 are for the case of a well-specified model. But what happens when we consider model with a misspecified distribution? Consider therefore that the true model is a logistic smooth transition autoregressive model with $\theta_0 = (\alpha, \phi, \gamma, c) = (0.4, -0.4, 2, 0.4)$. We will ignore the true model and estimate models assuming a LSTAR and ESTAR to examine the impact of misspecification on estimation method. In Table 4, we report the results for the value of the estimated parameters, p-value of the associated Kolmogorov-Smirnov test, KS, and mean-squared error of the estimation of the one-step ahead prediction, $MSE = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2$. We do this to illustrate the potential problems that will arise in a misspecified model

Table 3: The *RMSE* of LSTAR and ESTAR models.

	T	$\hat{\alpha}$	$\hat{\phi}$	$\hat{\gamma}$	\hat{c}
ESTAR:					
	50	0.1791	0.1575	0.2973	0.0882
	150	0.1182	0.0784	0.2696	0.0840
	300	0.0742	0.0727	0.1598	0.0232
	500	0.0534	0.0656	0.1066	0.0224
	1000	0.0014	0.0050	0.0492	0.0113
LSTAR:					
	50	0.2392	0.1089	0.6810	0.1981
	150	0.1979	0.0854	0.4378	0.0852
	300	0.0548	0.0527	0.2713	0.0503
	500	0.0189	0.0151	0.1996	0.0298
	1000	0.0065	0.0072	0.0536	0.0039

when parameters that do not belong to the true model are estimated and then used to make predictions. The results in Table 4 show that, as the sample size increase the mean square error decrease and the estimated LSTAR model has the lowest value of the MSE_h .

Table 4: The value of MMLE and P-value of K-S test in misspecification case.

	T	$\hat{\alpha}$	$\hat{\phi}$	$\hat{\gamma}$	\hat{c}	KS	MSE_h
LSTAR:							
	50	0.3750	-0.3779	1.7265	0.3039	0.2365	0.1384
	150	0.3814	-0.3804	1.8042	0.3632	0.5166	0.0575
	300	0.3845	-0.3974	1.9369	0.3755	0.6529	0.0129
	500	0.3917	-0.3998	1.9857	0.4068	0.7966	0.0087
	1000	0.4047	-0.4003	2.0184	0.4050	0.8812	0.0035
ESTAR:							
	50	0.7548	-0.2815	2.4215	0.6780	0.0000	0.7654
	150	0.8095	-0.2653	1.5375	0.6759	0.0284	0.5153
	300	0.8473	-0.2452	1.4295	0.6487	0.0745	0.2074
	500	0.8938	-0.2120	1.4987	0.6234	0.0884	0.1426
	1000	0.9463	-0.2089	1.4534	0.6054	0.2540	0.1198

For theoretical results of model selection test, let us consider LSTAR model as true model and LSTAR and ESTAR models as competing models. Using obtained data and modified maximum likelihood method, the value of Vuong's statistic for paired of competing models are computed and compared with $z_{0.975} = 1.96$. The obtained value of Vuong's test statistic are given in Table 5. We observe that LSTAR model is optimum model. The Kolmogorov-Smirnov test presented in Table 4 confirms these results. For more illustration see Figure 1-2. Also the relative frequency of model selection test for each of rejection-acceptance regions is

Table 5: The values of Vuong's test.

T	$H_f : LSTAR$ $H_g : ESTAR$	Result
50	4.0943	LSTAR is better than ESTAR
150	6.5131	LSTAR is better than ESTAR
300	7.4175	LSTAR is better than ESTAR
500	11.8412	LSTAR is better than ESTAR
1000	12.4518	LSTAR is better than ESTAR

computed and the results are summarized in Table 6. For example, if $n=300$, when we test

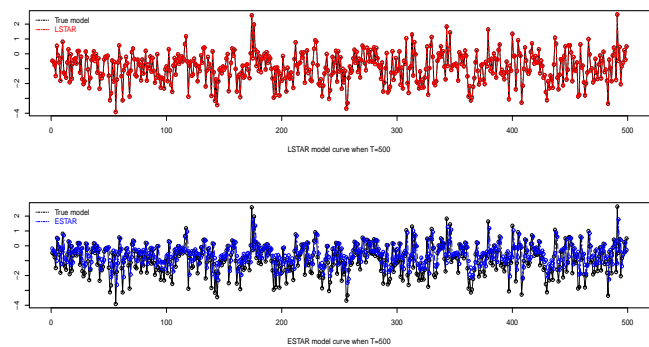


Figure 1: The estimated LSTAR and ESTAR models curve when $T=500$.

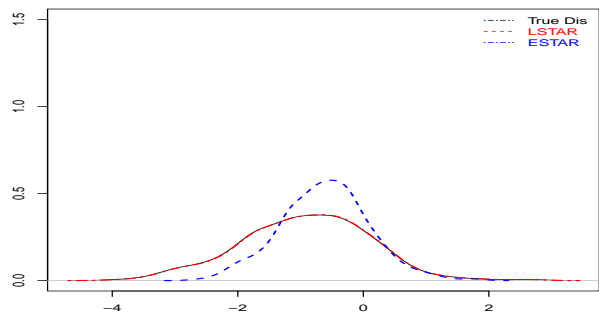


Figure 2: The competing model curve when $T=500$.

$H_f : LSTAR$ against $H_g : ESTAR$, the value of the relative frequency is 0.995, see column 2. It shows that Vuong's test select estimated LSTAR model as optimum model.

Table 6: The relative frequency of Vuong results.

T	f better	g better	f, g equi
50	0.854	0.062	0.084
150	0.971	0.023	0.006
300	0.995	0.005	0.000
500	0.999	0.001	0.000
1000	1.000	0.000	0.000

4 Conclusion

In this paper, we make a number of contributions to the literature that relates to nonlinear time series models. First, we proposed the modified maximum likelihood estimation for smooth transition autoregressive model. This is important, because explicit solutions from the likelihood equation cannot be obtained. The modified maximum likelihood estimators are consistent and asymptotically normally distributed. Also, the Vuong's test is proposed as model selection test based on the modified maximum likelihood estimator. Second, we examined the properties of the estimators and model selection statistic using simulations which validate our theoretical results. In summary, the theoretical derivations and the simulations support the use of the proposed model selection test as a useful tool in a practitioner's toolbox for empirical analysis.

References

- Chan, F. and Theoharakis, B. (2011). Estimating m-regimes STAR-GARCH model using QMLE with parameter transformation. *Journal of Mathematics and Computers in Simulation* **18**, 7, 1385-1396. <https://doi.org/10.1016/j.matcom.2010.05.023>.
- Chan, K. and Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis* **7**, 3, 179-190. <https://doi.org/10.1111/j.1467-9892.1986.tb00501.x>.
- Liao, Z. and Shi, X. (2020). A nondegenerate Vuong test and post selection confidence intervals for semi/nonparametric models. *Quantitative Economics* **11**, 983-1017.
- Schleer, F. (2016). Finding starting-values for the estimation of vector STAR models. *Journal of Econometrics* **3**, 65-90. <https://doi.org/10.3390/econometrics3010065>.
- Schneider, L., Chalmers, R.P., Debelak, R. and Merkle, E.C. (2019). Model selection of nested and non-nested item response models using Vuong tests. *Multivariate Behav Res* **55**, 5, 664-684. doi: 10.1080/00273171.2019.1664280.
- Shi, X. (2015). A nondegenerate Vuong test. *Journal of Quantitative Economics* **6**, 85-121.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Journal of Econometrica* **57**, 2, 307-333. <https://doi.org/10.2307/1912557>.
- Yaya, S. and Shitu, I. (2016). Symmetric variants of logistic smooth transition autoregressive models: Monte Carlo evidences. *Journal of Modern Applied Statistical Methods* **15**, 1 711-737.



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Estimation of count parameters in zero-inflated Poisson regression model based on Stein-Liu estimators

Zahra Zandi* and Hossein Bevrani

Department of Statistics, University of Kurdistan, Sanandaj, Iran

Abstract

The ZIP regression is applied for modelling count data with extra zeros. In this paper, we have improved the count parameter estimation in this model with correlated count predictor variables under the assumption of sparseness of the model using Stein-Liu estimators. We have conducted a Monte Carlo simulation study for various combinations of sample sizes, inactive count predictor variables, and correlation level between the count predictor variables in order to compare the performance of the suggested estimator with the unrestricted Liu estimator in terms of simulated relative efficiency criteria. Our results showed that the performance of the Stein-Liu estimators is superior to the unrestricted Liu estimator in any situations.

Keywords: Monte Carlo simulation, Multicollinearity problem, Stein-Liu estimators, zero-inflated Poisson model

Mathematics Subject Classification (2020): 62F10, 62F12, 62J07.

1 Introduction

The Poisson regression is widely used to model equivalent count data, i.e. the average and variance of data are equal. However, in many real applications, the count data contains an excess number of zeros which is known as zero-inflated count data. Zero-inflated Poisson (ZIP) regression model introduced by Lambert (1992) is a two-part model to analyze this data. The first part generates only structural zeros using a degenerate distribution at zero. This part is called zero-component. The other part is called count component which generates true zeros and positive count data from an ordinary Poisson distribution. The interested readers are referred to Hilbe (2011) for more details about ZIP model.

The multicollinearity problem that occurs in analysis count models includes correlated predictor variables. This problem leads to high variance of the maximum likelihood estimator (MLE) of the regression parameters. Thus, the interpretations based on this estimator can not be correct. The ridge regression estimator (Hoerl and Kennard (1970)) is a solution to overcome this problem. Also, the Liu estimator introduced by Liu (1993) is another method

*Corresponding author, zahrazandi163@yahoo.com

dealing with multicollinearity problem which has more advantages over the ridge estimator when there is a strong correlation between predictor variables.

When there is uncertain prior information (UPI) about some predictor variables which indicate that they are not significant and should be eliminated from regression model in order to improve parameters estimation, the Stein and positive Stein estimators introduced by Stein (1956) perform better than the MLE. The UPI appears on the parameter vector as a linear restriction (Ahmed (2014)).

In this paper, we consider a ZIP model in the presence of multicollinearity among the count predictors under a linear restriction on the count parameter vector β as

$$\mathbf{R}\beta = \mathbf{r} \quad (1)$$

where, $\mathbf{R}_{p_2 \times p}$ is a known matrix, $\beta_{p \times 1} = (\beta'_1, \beta'_2)'$ is the count parameters vector (β'_1 ($p_1 \times 1$) contains significant or active parameters and β'_2 ($p_2 \times 1$) includes non-significant or inactive parameters) and $\mathbf{r}_{p_2 \times 1}$ is a vector with constant elements. Our mean goal is to improve count parameters in the ZIP model using Stein-Liu (SL) and positive Stein-Liu (PSL) estimators. Recently these estimators are proposed in the multicollinear Poisson model by Zandi and Bevrani (2024).

This paper is organized as follows. The ZIP model, unrestricted Liu (UL), restricted Liu (RL) and Stein-Liu estimators are introduced in Section 2. The SL estimators are compared to the UL estimator using a Monte Carlo simulation study in Section 3, and conclusions are presented in the final section.

2 Model specifications and estimators

In the current section, we introduce the ZIP model, the SL, and PSL estimators to improve the count parameters estimation in this model.

2.1 Zero-inflated Poisson model

Suppose Y be dependent variable in the ZIP model. The probability mass function of Y for a random sample of size n is as follows:

$$f_{Y_i}(y_i | \tau_i, \mu_i) = \begin{cases} \tau_i + (1 - \tau_i) e^{-\mu_i} & y_i = 0 \\ (1 - \tau_i) \left(\frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} \right) & y_i = 1, 2, 3, \dots \end{cases}$$

where $\log(\mu_i) = \mathbf{x}'_i \beta$ is the link function of the Poisson regression model in the count component of ZIP model where, $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$ is the i th row of design matrix $\mathbf{X}_{n \times p}$ with p count predictor variables and $\text{logit}(\tau_i) = \mathbf{z}'_i \gamma$ is the link function related to the zero component, where $\mathbf{z}_i = (z_{i1}, z_{i2}, \dots, z_{iq})'$ is the i th row of design matrix \mathbf{Z} and γ is a $q \times 1$ vector of parameters related to the zero component. The log-likelihood function of the ZIP model can be written as follows:

$$\mathcal{L}(\beta, \gamma) = \sum_{i=1}^n \left\{ I_i \ln \left(e^{\mathbf{z}'_i \gamma} + e^{-\exp(\mathbf{x}'_i \beta)} \right) + (1 - I_i) \left(y_i \mathbf{x}'_i \beta - e^{\mathbf{x}'_i \beta} - \ln(y_i!) \right) - \ln(1 + e^{\mathbf{z}'_i \gamma}) \right\}, \quad (2)$$

where I_i is an indicator function defined as:

$$I_i = \begin{cases} 1 & y_i = 0 \\ 0 & y_i = 1, 2, 3, \dots \end{cases}$$

We denote the MLE of the parameter vector (β, γ) by $(\hat{\beta}^{MLE}, \hat{\gamma}^{MLE})$ that can be obtained by maximizing Equation (2) with respect to β and γ using numerical methods.

2.2 Unrestricted Liu and restricted Liu estimators

When there is near linear dependence among the count predictor variables in the ZIP model, the interpretations based on $\hat{\beta}^{MLE}$ are not true. One solution is to use Liu estimator (or unrestricted Liu estimator) as follows (Liu (1993)):

$$\hat{\beta}^{UL} = (\mathbf{X}'\hat{\mathbf{W}}\mathbf{X} + \mathbf{I}_p)^{-1}(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X} + d\mathbf{I}_p)\hat{\beta}^{MLE}, \quad 0 \leq d \leq 1 \quad (3)$$

where $\hat{\mathbf{W}}$ is a diagonal matrix such that the i th diagonal element is equal to $\hat{\mu}_i = \exp(\mathbf{x}_i'\hat{\beta}^{MLE})$, \mathbf{I}_p is an identity matrix of order p and d is the shrinkage parameter. The MSE of the UL estimator is defined as follows (Gelman and Golam Kibria (2020)):

$$MSE(\hat{\beta}^{UL}) = \sum_{j=1}^p \frac{(\lambda_j + d)^2}{\lambda_j(\lambda_j + 1)^2} + (d - 1)^2 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + 1)^2} \quad (4)$$

where λ_j is the j th eigenvalue of matrix $\mathbf{C} = \mathbf{X}'\hat{\mathbf{W}}\mathbf{X}$ and α_j is the j th element of $\mathbf{H}'\beta$, where \mathbf{H} is a $p \times p$ matrix whose j th column is the corresponding eigenvector of λ_j where, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. The estimation of shrinkage parameter d is as follows (Månsson (2013)):

$$\hat{d} = \max\left(0, \text{median}\left\{\frac{\hat{\alpha}_j^2 - 1}{\lambda_j^{-1} + \hat{\alpha}_j^2}\right\}\right),$$

where $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p)' = \mathbf{H}'\hat{\beta}^{MLE}$.

Under the linear restriction (1) and in the presence of multicollinearity, the RL estimator of β denoted by $\hat{\beta}^{RL}$ is defined as follows (Gelman and Golam Kibria (2020)):

$$\hat{\beta}^{RL} = \mathbf{A}\hat{\beta}^{MLE} - \mathbf{A}\mathbf{C}^{-1}\mathbf{R}'(\mathbf{R}\mathbf{C}^{-1}\mathbf{R}')^{-1}(\mathbf{R}\hat{\beta}^{MLE} - \mathbf{r}), \quad (5)$$

where $\mathbf{A} = (\mathbf{X}'\hat{\mathbf{W}}\mathbf{X} + \mathbf{I}_p)^{-1}(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X} + d\mathbf{I}_p)$.

2.3 Stein-Liu estimators

Based on the idea of Zandi and Bevrani (2024), the Stein-Liu estimator of β denoted by $\hat{\beta}^{SL}$ is defined by combining the $\hat{\beta}^{UL}$ and $\hat{\beta}^{RL}$ as follows:

$$\hat{\beta}^{SL} = \hat{\beta}^{RL} + \left(1 - \frac{p_2 - 2}{T_n}\right)(\hat{\beta}^{UL} - \hat{\beta}^{RL}), \quad p_2 = 3, 4, \dots,$$

where T_n is the test statistic for testing $H_0 : \mathbf{R}\beta = \mathbf{r}$ versus $H_1 : \mathbf{R}\beta \neq \mathbf{r}$ which is defined as follows:

$$T_n = 2\left\{\mathcal{L}(\hat{\beta}^{UL}) - \mathcal{L}(\hat{\beta}^{RL})\right\},$$

where $\mathcal{L}(\hat{\beta}^{UL})$ and $\mathcal{L}(\hat{\beta}^{RL})$ are, respectively, the log-likelihood function (2) values for the unrestricted Liu and restricted Liu estimators. The test statistic T_n under H_0 has $\chi^2(p_2)$ when $n \rightarrow \infty$.

The Positive Stein-Liu estimator of β denoted by $\hat{\beta}^{PSL}$ is defined as an adjusted version of the SL estimator:

$$\hat{\beta}^{PSL} = \hat{\beta}^{RL} + \max\left\{0, 1 - \frac{p_2 - 2}{T_n}\right\}(\hat{\beta}^{UL} - \hat{\beta}^{RL}), \quad p_2 = 3, 4, \dots$$

3 Simulation experiment

We compare the performance of the Stein-Liu estimators with respect to the UL in ZIP model in the presence of multicollinearity via a Monte Carlo simulation using the simulated relative efficiency (SRE) criteria which is defined as:

$$SRE(\hat{\beta}^{UL}, \hat{\beta}^\diamond) = \frac{SMSE(\hat{\beta}^{UL})}{SMSE(\hat{\beta}^\diamond)},$$

where $\hat{\beta}^\diamond$ is any one of the $\hat{\beta}^{SL}$ and $\hat{\beta}^{PSL}$, and $SMSE(\hat{\beta}^\diamond)$ is the simulated mean squared error of $\hat{\beta}^\diamond$ which is defined as:

$$SMSE(\hat{\beta}^\diamond) = \frac{1}{2000} \sum_{t=1}^{2000} (\hat{\beta}^\diamond - \beta)_t' (\hat{\beta}^\diamond - \beta)_t.$$

For the value of SRE is greater than one, $\hat{\beta}^\diamond$ dominates the UL estimator. The correlated count predictor variables in the ZIP model were generated as:

$$x_{ij} = (1 - \rho^2)^{1/2} u_{ij} + \rho u_{ip}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, p,$$

here x_{ij} s are elements of design matrix \mathbf{X} , ρ represents the correlation level between the count predictor variables and u_{ij} s are generated independently from standard normal distribution. Hence, the response variable Y_i in the ZIP model is generated with $\mu_i = e^{\mathbf{x}_i' \beta}$ and $\tau_i = 0.5$. The simulation is conducted in *R* statistical software under different sample sizes, $n = 35, 50$, various correlation level between the predictor variables, $\rho = 0.90, 0.93, 0.95$, and the inactive predictor variables, $p_2 = 3, 5, 7$. We also set the number of active predictors $p_1 = 3$, thus $p = p_1 + p_2 = 6, 8, 10$. The simulation is replicated 2000 times for each case. We consider the linear restriction as $\mathbf{R}\beta = \mathbf{0}$, where $\mathbf{R} = [\mathbf{0}_{p_1 \times p_2}, \mathbf{I}_{p_2 \times p_2}]$ such that $\mathbf{0}_{p_1 \times p_2}$ is a matrix with zero elements and $\mathbf{I}_{p_2 \times p_2}$ is an identity matrix. Therefore, the count parameter is partitioned as $\beta = (\beta_1', \beta_2')'$, where $\beta_1 = (0.09, 0.27, 0.34)'$ contains the active parameters and $\beta_2 = (\mathbf{0}_{p_2})'$ includes the inactive parameters. To compare the performance of the Stein-Liu estimators to the UL estimator in different parts of the parameter space, we define a distance between the proposed sub-model and the simulated model as $\Delta = \|\beta - \beta^0\|^2$, where $\beta^0 = (\beta_1', \mathbf{0}_{p_2}')$ is the true parameter and $\|\cdot\|$ is the Euclidean norm. We set $\Delta = 0, 0.4, 0.6, 0.8, 1, 2, 3, 4$ in the simulations.

The simulation results reported in Tables 1-3 confirm that the SREs of the RL and Stein-Liu estimators increase by increasing the number of inactive count predictors p_2 and also as the correlation level ρ increases. All these estimators have better performance than the UL estimator when the sample size n decreases. The restricted Liu estimator has the best performance at $\Delta = 0$. When Δ moves away from zero, the SRE of this estimator sharply decreases. The positive Stei- Liu estimator outperforms better than the Stein-Liu estimator at $\Delta = 0$ and near it in any situation. An important result of the proposed estimators is that the Stein-Liu estimators uniformly outperform the unrestricted Liu estimator.

Discussion

In this article, we proposed Stein-Liu and positive Stein-Liu estimators to estimate count parameters in Zero-inflated Poisson model in the presence of multicollinearity under a linear restriction on the count parameters. The performance of the suggested estimators was investigated through a Monte Carlo simulation study. The simulation results revealed that the performances of the Stein-Liu estimators were better than the unrestricted Liu estimator in various parts of the parameter space.

Acknowledgment

The authors are thankful to the reviewers for the insightful comments and suggestions that have resulted in a much improved version of this paper.

References

Ahmed S. E. (2014), *Penalty, shrinkage and pretest strategies – Variable selection and estimation*. Heidelberg, Springer.

- Hilbe J. M. (2011), *Negative binomial regression*, Cambridge UK: Cambridge University Press.
- Hoerl A. E. and Kennard R. W. (1970), Ridge regression: Biased estimation for non-orthogonal problems. *Technometrics* **12** (1), 55–67
- Gelman K. and Golam Kibria B. M. (2020), Estimating the unrestricted and restricted Liu estimators for the Poisson regression model: method and application. *Computational Economics* **58**, 311–326.
- Lambert D. (1992), Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics* **34** (1), 1–14.
- Liu K. (1993), A new class of biased estimate in linear regression. *Communications in Statistics-Theory and Methods* **22**, 393–402.
- Månsson K. (2013), Developing a Liu estimator for the negative binomial regression model. *Journal of Statistical Computation and Simulation* **83** 1773–1780.
- Stein C. (1956), The admissibility of Hotelling’s T^2 -test. *Mathematical Statistics* **27**, 616–623.
- Zandi Z. and Bevrani H. (2024), Improved parameters estimation in the multicollinear Poisson regression model based on Stein-Liu estimators. *Journal of Statistical Modelling: Theory and Applications* **5** (2), 77–95.

Table 1: The $SREs$ of the proposed estimators with respect to $\hat{\beta}^{UL}$ for $p_2 = 3$

ρ	Δ	$n = 35$			$n = 50$		
		\hat{d}			\hat{d}		
		RL	SL	PSL	RL	SL	PSL
0.90	0.0	4.946	1.456	1.627	3.139	1.198	1.490
	0.4	3.125	1.374	1.439	2.001	1.267	1.319
	0.6	2.115	1.259	1.303	1.313	1.172	1.197
	0.8	1.453	1.196	1.201	0.879	1.111	1.119
	1.0	1.035	1.136	1.137	0.614	1.078	1.078
	2.0	0.303	1.037	1.037	0.173	1.019	1.019
	3.0	0.139	1.017	1.017	0.078	1.008	1.008
	4.0	0.079	1.010	1.010	0.044	1.005	1.005
0.93	0.0	5.627	1.532	1.661	3.365	1.184	1.493
	0.4	3.946	1.450	1.512	2.469	1.285	1.383
	0.6	2.836	1.350	1.385	1.756	1.213	1.269
	0.8	2.030	1.264	1.275	1.238	1.156	1.175
	1.0	1.485	1.192	1.195	0.894	1.109	1.115
	2.0	0.456	1.053	1.053	0.264	1.029	1.029
	3.0	0.211	1.024	1.024	0.121	1.012	1.012
	4.0	0.121	1.014	1.014	0.068	1.007	1.007
0.95	0.0	5.764	1.464	1.595	3.534	1.386	1.507
	0.4	4.418	1.417	1.501	2.853	1.295	1.433
	0.6	3.381	1.343	1.408	2.178	1.245	1.335
	0.8	2.540	1.223	1.317	1.619	1.176	1.243
	1.0	1.922	1.232	1.241	1.210	1.158	1.172
	2.0	0.632	1.070	1.070	0.382	1.044	1.044
	3.0	0.298	1.031	1.031	0.177	1.019	1.019
	4.0	0.171	1.018	1.018	0.101	1.010	1.010

Table 2: The $SREs$ of the proposed estimators with respect to $\hat{\beta}^{UL}$ for $p_2 = 5$

ρ	Δ	$n = 35$			$n = 50$		
		\hat{d}			\hat{d}		
		RL	SL	PSL	RL	SL	PSL
0.90	0.0	5.661	2.025	2.450	5.035	1.871	2.440
	0.4	3.374	1.787	1.960	2.684	1.715	1.829
	0.6	2.312	1.588	1.650	1.667	1.433	1.510
	0.8	1.614	1.425	1.437	1.086	1.310	1.322
	1.0	1.166	1.301	1.303	0.749	1.216	1.216
	2.0	0.356	1.075	1.075	0.208	1.057	1.057
	3.0	0.166	1.030	1.030	0.094	1.026	1.026
	4.0	0.095	1.015	1.015	0.053	1.014	1.014
0.93	0.0	5.889	2.028	2.461	5.282	1.803	2.465
	0.4	3.911	1.892	2.076	3.274	1.833	1.977
	0.6	2.845	1.641	1.786	2.194	1.577	1.662
	0.8	2.073	1.525	1.560	1.498	1.409	1.443
	1.0	1.542	1.393	1.401	1.063	1.302	1.307
	2.0	0.498	1.107	1.107	0.310	1.084	1.084
	3.0	0.235	1.044	1.044	0.142	1.038	1.038
	4.0	0.136	1.022	1.022	0.081	1.021	1.021
0.95	0.0	6.098	2.071	2.340	5.518	1.987	2.457
	0.4	4.379	1.860	2.046	3.844	1.956	2.105
	0.6	3.354	1.717	1.819	2.762	1.702	1.820
	0.8	2.544	1.578	1.622	1.978	1.495	1.586
	1.0	1.949	1.457	1.472	1.448	1.401	1.420
	2.0	0.672	1.148	1.148	0.447	1.118	1.118
	3.0	0.323	1.065	1.065	0.207	1.052	1.052
	4.0	0.188	1.035	1.035	0.118	1.029	1.029

Table 3: The $SREs$ of the proposed estimators with respect to $\hat{\beta}^{UL}$ for $p_2 = 7$

ρ	Δ	$n = 35$			$n = 50$		
		\hat{d}			\hat{d}		
		RL	SL	PSL	RL	SL	PSL
0.90	0.0	8.802	2.783	3.203	6.337	2.661	3.269
	0.4	5.482	2.470	2.730	3.680	2.202	2.414
	0.6	3.751	2.156	2.331	2.359	1.860	1.920
	0.8	2.605	1.910	1.992	1.563	1.596	1.609
	1.0	1.871	1.714	1.741	1.088	1.423	1.424
	2.0	0.560	1.238	1.238	0.306	1.120	1.120
	3.0	0.259	1.111	1.111	0.139	1.056	1.056
	4.0	0.148	1.064	1.064	0.079	1.033	1.033
0.93	0.0	9.044	2.756	3.091	7.187	2.770	3.305
	0.4	6.285	2.550	2.822	4.680	2.385	2.621
	0.6	4.593	2.326	2.509	3.188	2.052	2.149
	0.8	3.342	2.096	2.201	2.196	1.773	1.803
	1.0	2.478	1.894	1.944	1.565	1.567	1.573
	2.0	0.788	1.333	1.333	0.458	1.167	1.167
	3.0	0.369	1.156	1.156	0.210	1.077	1.077
	4.0	0.212	1.089	1.089	0.119	1.044	1.044
0.95	0.0	9.157	2.638	2.878	1.044	2.779	3.216
	0.4	6.822	2.490	2.694	5.228	2.466	2.709
	0.6	5.274	2.315	2.474	3.804	2.178	2.299
	0.8	4.018	2.121	2.238	2.745	1.899	1.960
	1.0	3.083	1.944	2.020	2.019	1.691	1.712
	2.0	1.058	1.405	1.406	0.626	1.223	1.223
	3.0	0.507	1.203	1.203	0.291	1.105	1.105
	4.0	0.294	1.119	1.119	0.166	1.061	1.061



پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان



Seminar On Probability
and Stochastic Processes
August 30-31, 2025
University of Kurdistan



Value at Risk Estimation Using the ARMA–GARCH Framework

Somaye Mohebbi ^{*1}, Ali Mohammadian Mosammam²

¹Department of Mathematics, University of Zanjan, Zanjan, Iran.

²Department of Statistics, University of Zanjan, Zanjan, Iran.

Abstract

This study presents a comprehensive framework for modeling volatility in financial data and estimating Value at Risk (VaR). It begins with quantile regression to examine the influence of macroeconomic variables on the returns of Bank M.E. The residual structure of this model is then analyzed for dependence and autocorrelation, for which Autoregressive Moving Average (ARMA) models are fitted to capture linear patterns in the errors. Conditional variance is subsequently modeled using various members of the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) family, classified and tailored to the data. The resulting volatility estimates are employed to compute VaR, and the forecasting performance is assessed via statistically robust backtesting procedures. ARCH tests conducted before and after the modeling process demonstrate that the proposed framework effectively removes autocorrelation and improves the modeling of volatility dynamics.

Keywords: Time Series, ARMA models, GARCH models and Value at Risk.

Mathematics Subject Classification (2020): 62M10.

1 Introduction

In today's world, data has become a foundational element of scientific, industrial, and social decision-making. Any observable, measurable, or storable phenomenon generates data that, when properly analyzed, can provide deep and actionable insights. Data is not only a reflection of reality; it serves as a powerful tool for pattern discovery, behavior prediction, and process optimization.

With the rapid growth of digital technologies, the volume of generated data is increasing at an unprecedented rate, a phenomenon referred to as Big Data. These data sources are often highly complex in structure, frequently exhibit temporal dependencies and volatility, and are especially prominent in domains such as financial markets and climate systems.

Among the various types of data, those recorded systematically over time, known as **time series** data, hold particular significance. This form of data is characterized by a unique

*Corresponding author, so.mohebbi@znu.ac.ir

structure in which each observation is dependent on its temporal context and often influenced by previous values. Examples include stock prices, exchange rates, weather measurements, and daily patient counts.

Modeling time series data plays a crucial role not only in understanding historical behavior but also in forecasting future outcomes. Through specialized statistical tools, researchers can uncover hidden patterns, detect volatility, and identify conditional dependencies embedded within temporal datasets.

Given the importance of volatility forecasting in financial data and its critical role in risk management, volatility models such as GARCH have increasingly attracted attention as effective tools for Value at Risk (VaR) assessment. This study aims to develop a statistical framework for more precise risk analysis by utilizing GARCH-based models and evaluating their performance in estimating VaR for selected financial datasets.

2 Methodology

In this section, the theoretical framework for statistical time series modeling is presented over the probability space $(\Omega, \mathcal{F}_t, \Pr)$. From a statistical perspective, a time series X_t can be modeled as follows:

$$X_t = \mu_t + \varepsilon_t \quad (1)$$

$$\varepsilon_t = h_t^{1/2} z_t, \quad (2)$$

According to Equation (1), X_t the return at time t is decomposed into a conditional mean component μ_t , given the sigma-field \mathcal{F}_{t-1} , and a residual term ε_t . Equation (2) defines the residuals as the product of the conditional volatility $\sqrt{h_t}$ and a stochastic innovation process z_t . Here, h_t represents the conditional variance of the series given \mathcal{F}_{t-1} , and z_t is an independent and identically distributed (*i.i.d.*) process with zero mean and unit variance. The conditional mean μ_t is modeled using an Autoregressive Moving Average (ARMA) structure, while the conditional variance h_t is captured through GARCH models.

2.1 Autoregressive Moving Average

Time series analysis often begins with two fundamental models: the autoregressive (AR) model and the moving average (MA) model. Both aim to explain the internal structure of the data by relying on past observations, though the nature of dependency in each is different.

2.1.1 Autoregressive

Just as linear regression aims to model a relationship between explanatory and response variables, time series analysis—at its simplest—assumes that the current value of a time series depends on its own past values. When this relationship is linear, it can be represented using an autoregressive model of order p ($AR(p)$), expressed mathematically as:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t,$$

where x_t is the value of the time series at time t , $\phi_1, \phi_2, \dots, \phi_p$ are the autoregressive coefficients and w_t is a white noise error term. This formulation captures the principle that past observations can serve as predictors for present (and potentially future) values.

2.1.2 Moving Average

In the Moving Average model of order q ($MA(q)$), the current value depends on previous shocks (noise terms) expressed as:

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q},$$

where w_t is a white noise with mean zero and variance σ_w^2 . This model captures dependency on the structure of past errors and shocks and is suitable for modeling rapid and unpredictable fluctuations.

The combination of the AR and MA models leads to the Autoregressive Moving Average of orders p and q ($ARMA(p, q)$) as:

$$x_t = \mu + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q}, \quad (3)$$

here, μ denotes the intercept. This compact model allows us to capture both the persistent structure of the series and irregular fluctuations. In practice, $ARMA$ models are usually applied to stationary time series. $ARMA$ models are also considered effective for modeling the first-order moment, particularly the conditional mean of time series data (Shumway and *et al.*, 2000).

2.2 Generalized Autoregressive Conditional Heteroskedasticity

Empirical studies have demonstrated that the assumption of constant variance over time is inadequate in financial time series. In response to the observed time-varying behavior of volatility, Engle (1982) introduced the Autoregressive Conditional Heteroskedasticity ($ARCH$) model, which allows the conditional variance to be modeled dynamically as a function of past shocks. The general $ARCH(p)$ model is specified as:

$$h_t = \alpha_0 + \sum_{j=1}^p \alpha_j \epsilon_{t-j}^2,$$

where, h_t denotes the conditional variance at time t , ϵ_{t-j} represents past residuals (shocks) and α_0, α_j are positive model parameters.

In the $ARCH$ structure, the conditional variance increases following large shocks, i.e., larger absolute values of ϵ_{t-j} yield higher volatility forecasts. This reflects the characteristic behavior of volatility clustering, where periods of high fluctuations tend to be followed by further volatility. To ensure finite variance, the condition $h_t < \infty$ must be satisfied, which places restrictions on the parameter space to maintain weak stationarity.

Despite their theoretical foundation, $ARCH$ models rely solely on past squared residuals, which can lead to over parameterization and inefficiency in capturing the persistence of volatility. To overcome these limitations, the Generalized $ARCH$ ($GARCH$) families are proposed, which enriches the structure by incorporating both past innovations and previous conditional variances. This recursive framework allows for more parsimonious modeling of volatility clustering. In general, $GARCH$ models can be classified into two main categories:

- Symmetric Models
 - The classical $GARCH$ model introduced by Bollerslev and Engle (1986):

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j},$$
 - The Absolute Value $GARCH$ ($AVGARCH$) model proposed by Poon and Taylor (1992):

$$h_t^{1/2} = \alpha_0 + \alpha_1 |\epsilon_{t-1}| + \beta_1 h_{t-1}^{1/2}$$
- Asymmetric Models
 - Exponential $GARCH$ ($EGARCH$) introduced by Nelson (1991):

$$\log(h_t) = \alpha_0 + \alpha_1 \frac{|\epsilon_{t-1}|}{\sqrt{h_{t-1}}} + \gamma \frac{\epsilon_{t-1}}{\sqrt{h_{t-1}}} + \beta_1 \log(h_{t-1})$$
 - Zakoian $GARCH$ ($ZARCH$) introduced by Zakoian (1994):

$$h_t^{1/2} = \alpha_0 + \alpha_1 |\epsilon_{t-1}| + \gamma I\{\epsilon_{t-1} < 0\} |\epsilon_{t-1}| + \beta_1 h_{t-1}^{1/2}$$
 - GJR $GARCH$ proposed by Glosten *et al.* (1993):

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \gamma \{\epsilon_{t-1} < 0\} \epsilon_{t-1}^2 + \beta_1 h_{t-1} \quad (4)$$

Symmetric models assume that the sign of an innovation does not influence volatility; only its magnitude matters. In other words, a positive shock and a negative shock of equal size are assumed to have the same impact on the conditional variance.

However, early empirical evidence challenged this assumption. [Black \(1976\)](#) was the first to observe that negative shocks tend to cause greater increases in volatility than positive shocks of similar magnitude—a phenomenon now known as the leverage effect.

This asymmetry in volatility response led to the development of asymmetric GARCH models, which explicitly account for the direction of shocks in the conditional variance equation. These models are designed to allow for differential reactions to positive and negative innovations. [Mohebbi and Mosammam \(2025\)](#) employed GARCH family models to assess systemic risk within the Iranian banking network.

3 Value at Risk

Value at Risk (VaR) is one of the fundamental tools in financial risk management, designed to estimate the potential loss of a portfolio over a specified time horizon and confidence level.

Definition 3.1. *Value at Risk (VaR) is formally defined as the worst expected loss within a given confidence level over a specified time horizon. Mathematically, it is expressed as:*

$$VaR_\alpha(X_t) = \inf \{x_t \geq 0 \mid P(X_t \leq x_t) \geq \alpha\}$$

where, X_t represents the portfolio value at time t , α is the probability that losses exceed the VaR threshold and $1 - \alpha$ is the confidence level.

Its importance lies in its ability to summarize risk exposure into a single interpretable figure, which can serve as a basis for strategic decision-making, capital allocation, and regulatory compliance. Given that financial data often exhibit features such as volatility clustering, skewness, and leverage effects, the use of GARCH models—especially asymmetric variants—is highly effective in producing accurate and dynamic VaR estimates. By incorporating the conditional structure of volatility, these models enable more realistic risk forecasting and play a key role in reducing uncertainty in financial environments.

Discussion and Results

In this study, financial time series data from Bank of Middle East (M.E.) spanning the period from March 25, 2014, to February 1, 2025 are analyzed. Missing observations were addressed using Kalman filtering, and a preprocessing stage was conducted to ensure data quality. A linear quantile regression model is then employed to estimate Value at Risk (VaR), in accordance with Definition 3.1, using the logarithmic return of bank stocks as the dependent variable and macroeconomic indicators—including exchange rate, interbank interest rate, and inflation rate—as the independent variables. Since the true value of the τ -th quantile is not observable, the error term is defined as the deviation between the actual observed value and the estimated quantile:

$$y_t = Q_\tau(y_t|x_t) + e_t, \quad (5)$$

$$e_t = \mu_t + \varepsilon_t, \quad \mu_t = \text{ARMA Process} \quad (6)$$

$$\varepsilon_t = h_t^{1/2} z_t, \quad h_t = \text{GARCH Process} \quad (7)$$

To ensure that the residuals from the quantile estimation are free of autocorrelation and to properly capture conditional heteroskedasticity, a combined ARMA-GARCH modeling approach is applied. The ARMA component addresses serial dependence in the mean equation, while the GARCH structure captures the dynamic behavior of conditional variance. This

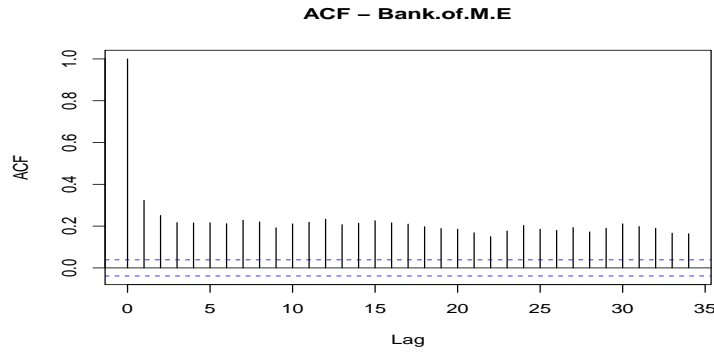


Figure 1: Autocorrelation function (ACF) of residuals prior to ARMA-GARCH model estimation.

integrated framework allows for a more realistic representation of financial time series and improves the accuracy of risk forecasts.

In the initial stage of analysis for the bank, the residuals obtained from the quantile regression model were evaluated using the autocorrelation function (ACF). The ACF plot indicated substantial serial correlation within the time series, suggesting that the baseline model was insufficient in capturing the underlying temporal dependencies. This diagnostic result is visualized in Figure 1.

Table 1: ARCH LM Test Result for residuals of Bank of M.E. prior to ARMA-GARCH model estimation.

χ^2 Statistic	Degrees of Freedom	p-value
76.367	2	$< 2.2 \times 10^{-16}$

To investigate conditional heteroskedasticity in the residuals, we performed the ARCH LM test with lag = 2 which is reported in Table 1. The null hypothesis of no ARCH effects was strongly rejected, indicating the presence of volatility clustering. As a result, a GARCH-family model was deemed appropriate for modeling conditional variance.

Based on the minimum value of the Bayesian Information Criterion (BIC), the ARMA(1,2)–GARCH(1,1) model was selected to capture the conditional mean and volatility dynamics. Due to the presence of asymmetric responses to positive and negative shocks commonly observed in financial time series, the GJR-GARCH specification was employed. The estimated parameters of this model—reported in Table 2—are obtained based on Equation (3) and Equation (4).

Table 2: Estimated parameters of the ARMA(1,2)–GJR-GARCH(1,1) model for Bank of M.E.

Parameter	Estimate	Std. Error	t-value	p-value
μ	0.01099	0.00322	3.41	0.0007
ϕ_1	0.97391	0.00531	183.37	0.0000
θ_1	-0.49084	0.02271	-21.61	0.0000
θ_2	-0.15993	0.02127	-7.52	0.0000
ω	0.000016	0.000003	6.33	0.0000
α_1	0.46832	0.04475	10.47	0.0000
β_1	0.55631	0.03039	18.31	0.0000
γ	-0.05144	0.05995	-0.86	0.3909
Skew	0.99429	0.02258	44.03	0.0000
Shape	3.37873	0.13548	24.94	0.0000

Table 2 presents the estimated parameters of the model fitted to the financial time series data for Bank of Meadle East (M.E.). The intercept term (μ) is statistically significant ($p < 0.01$), indicating a positive unconditional mean in the series. The autoregressive coefficient (ϕ_1) is very close to unity and highly significant, suggesting strong persistence in returns. The moving average coefficients (θ_1, θ_2) are both negative and significant, confirming the presence of short-term correction mechanisms in shock propagation. Regarding volatility dynamics, the coefficients α_1 and β_1 are both statistically significant, validating the existence of conditional heteroskedasticity and volatility clustering. The leverage effect parameter (γ), though negative as expected, is not statistically significant at conventional levels, implying that the asymmetry in shock response may be mild in this case. Finally, the skewness and shape parameters of the skewed Student-t distribution are both highly significant, highlighting the need to accommodate heavy tails and asymmetry in the innovation process.

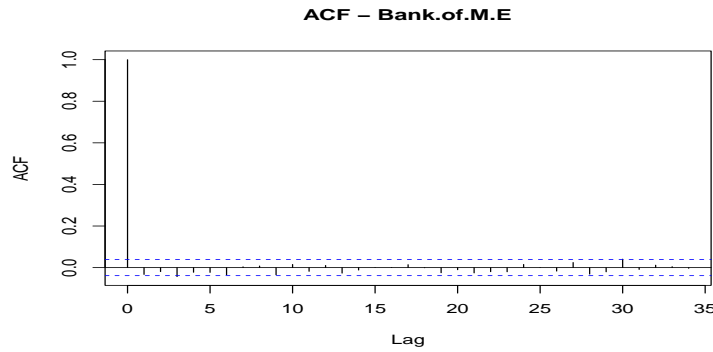


Figure 2: Autocorrelation function (ACF) of residuals following ARMA-GARCH model estimation.

Table 3: ARCH LM Test Result for residuals of Bank of M.E following ARMA-GARCH model estimation.

χ^2 Statistic	Degrees of Freedom	p-value
0.035	2	0.9825

Figure 2 illustrates that following model fitting, the autocorrelation effect is noticeably reduced. Furthermore, the ARCH test results reported in Table 3 confirm that the null hypothesis is accepted, indicating no remaining ARCH effects in the residuals.

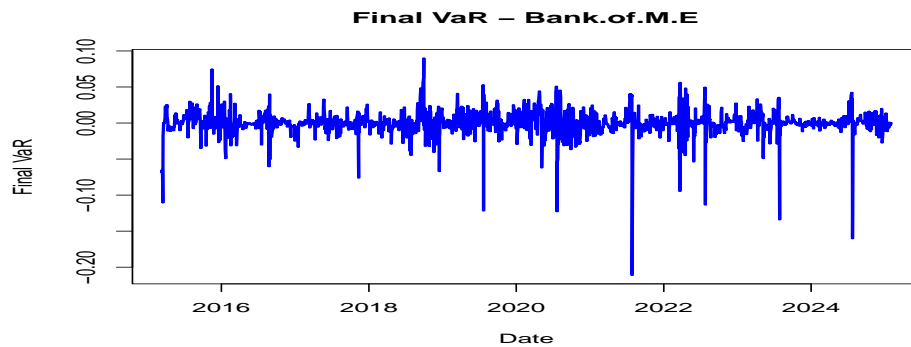


Figure 3: Time series of estimated Value-at-Risk (VaR) for Bank M.E at the 95

Finally, Figure 3 displays the estimated Value at Risk (VaR) for Bank M.E. over the

specified time period with 95 % of confidence level, revealing a consistently elevated level of potential risk throughout the observed period.

Acknowledgment

The authors gratefully acknowledges the **Monetary and Banking Research Institute of the Central Bank of the Islamic Republic of Iran** for providing access to the data and for financially supporting this doctoral dissertation.

References

- Black, F. (1976), *Studies of Stock Price Volatility Changes*, Proceedings of the 1976 Meeting of the Business and Economic Statistics Section, American Statistical Association, Washington DC, pp. 177–181.
- Engle, R. F. (1982), *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*, *Econometrica*, **50**(4), 987–1007.
- Engle, R. F., and Bollerslev, T., (1986), *Modelling the Persistence of Conditional Variances*, *Econometric Reviews*, **5**(1), 1-50 .
- Glosten L. R., Jagannathan R., and Runkle D. E. (1993), *On the relation between the expected value and the volatility of the nominal excess return on stocks*, *The Journal of Finance*, **48**(5), pp. 1779–1801.
- Mohebbi S., and Mosammam A. (2025), *Evaluating Systemic Risk with Conditional Value at Risk and Vine Copulas in the Iranian Banking Network*, *Journal of Statistical Studies*, **19**(1), pp. 183–204.
- Nelson D. B. (1991), *Conditional heteroskedasticity in asset returns: A new approach*, *Econometrica*, **59**(2), pp. 347–370.
- Poon, S. H., and Taylor, S. J., (1992), *Stock Returns and Volatility: An Empirical Study of the UK Stock Market*, *Journal of Banking and Finance*, **16**(1), 37-59.
- Shumway, R. H., Stoffer, D. S., and Stoffer, D. S., (2000), *Time series analysis and its applications*, **3**, New York: springer.
- Zakoian J.-M. (1994), *Threshold heteroskedastic models*, *Journal of Economic Dynamics and Control*, **18**(5), pp. 931–955.