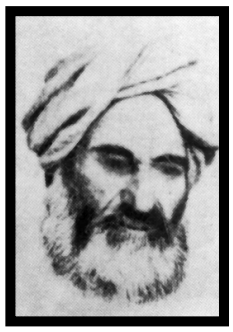


JOURNAL OF SCIENCES

University of Tehran

Abu Reyhan al-Biruni
(973-1048)
Iranian Scientist

ISLAMIC REPUBLIC OF IRAN

ISSN 1016-1104

Vol. 35, No. 3, Summer 2024

CONTENTS

بسم الله الرحمن الرحیم

In the Name of Allah, the Beneficent, the Merciful

JOURNAL OF SCIENCES

ISLAMIC REPUBLIC OF IRAN

Journal of Sciences, Islamic Republic of Iran is published quarterly by The University of Tehran,
16th St., North Kargar Ave, Tehran, Islamic Republic of Iran
Tel.: (0098-21)88334188, 88012080-212
Fax: (0098-21) 88334188
P.O. Box 13145-478
http://jsciences.ut.ac.ir
E-mail: jsciences@ut.ac.ir
E-mail: nooridaloii@sina.tums.ac.ir

ISSN 1016-1104

Instructions to Authors

1. General Policy

The *Journal of Sciences, Islamic Republic of Iran* (*J. Sci. I. R. Iran*) is published quarterly by the University of Tehran. Contributions from all fields of basic sciences may be submitted by scientists from all over the world.

The papers submitted to this journal should not have been published previously, except in the form of a brief preliminary communication, nor submitted to another journal. The decision to accept a contribution rests with the Editorial Committee of the *J. Sci. I. R. Iran.* Manuscripts will be considered for publication in the form of articles, preliminary communications, notes and review articles. The work should be original or a through review by an authoritative person in a pertinent field.

2. Copyright

Submission of a manuscript implies that the author(s) agree to transfer copyright to the *J. Sci. I. R. Iran* when the contribution is accepted for publication. Reproduction of the text, figures, or tables of this journal is allowed only by permission of the Editorial Committee.

3. Preparation of Manuscripts

3.1. General considerations. Manuscripts must be submitted in English according to Journal Instructions (It is necessary to submit at least 2 files including "Title page" and "Main file"). They **must be** typewritten in **Microsoft Word** (all versions). Authors are requested to reserve margins of at least 3 cm at the top and bottom of each page and at least 4 cm on the left-hand side.

Tables and illustrations (both numbered in Arabic numerals) should be prepared on separate pages. Tables require a heading and figures a legend, also prepared on a separate page. In Electronic submission, figures should be with the following caveats: all figures should be submitted at a minimum of 300 dpi and saved as TIFF files (avoid submitting JPEG files) after your text file.

Manuscripts should be kept to a minimum length and should be subdivided into labeled sections (**Introduction**, **Materials and Methods**, **Results**, **Discussion**, **Acknowledgement**, **References**). A current issue of *J. Sci. I. R. Iran* should be consulted.

3.2. ***Title page.*** The title of a manuscript should reflect concisely the purpose and findings of the work in order to provide maximal information for a computerized title search. Abbreviations, symbols, chemical formulae, references, and footnotes should be avoided.

The authors' full first names, middle initials and last names should be given, followed by the address(es) of the contributing department(s). (e.g. Department, Faculty, University, City, Country).

Telephone, Fax and Email of corresponding author should be footnoted on the bottom of the first page of each manuscript.Footnotes may be added to indicate the present mailing address(es) of the author(s). (e.g. *Corresponding author, Tel: 00982188012080, Fax: 00982188012081, E-mail: jsciences@ut.ac.ir).

Special types of print should be indicated as follows:

Type	Mark	Symbol	Example	Example (Printed)
Boldface[a]	Single underline	———	Introduction	**Introduction**
Italic[b]	Wave-like underline	~~~	In vivo	*In vivo*
Small capital[c]	Double underline	═══	0.2 m	0.2 M
Boldface italic[d]	Underline with wavy line	~~~	R	***R***

a) Headings, designated numbers of chemical compounds, subheadings.

b) Configurational prefixes ((R)-, (S)-, cis-, trans-, tert-, etc); Latin words or abbreviations, words in languages other than English; trade names of compounds (first letter should be capitalized); names of authors if mentioned in the text.

c) Symbols of molar and normal concentrations (M and N), D- and L-, the names or initials of the nomenclature of species.

d) Italicized terms and prefixes in headings.

The total number of pages (including references, tables, copies of formula collections (if any), schemes and figures should be marked in the upper left-hand corner of the first page of each copy. The complete address, including phone number, and E-mail address of the correspondence author should also be given.

3.3. ***Main File:***

a) The title of the article (The first letter of each word must be capital).

b) Abstract should be self-explanatory and intelligible without references to the text and titles, it must not exceed 250 words.

c) At least between three to five keywords should be chosen by the author(s).

d) Introduction

e) Materials and Methods

f) Results and Discussion (they can be separate section too)

g) Figures and Tables (if There are any)

Theoretical articles must have atleast two main title : Introduction and Results.

3.4. References. References may be numbered alphabetically or sequentially in the order they are cited in the text. References typed with double spacing are to be listed in numerical order at the end of the main text. They should be addressed according to the following examples:

Journals:

Noori-Daloii M.R., Swift R.A., Kung H.J., Crittenden L.B., and Witter R.L. Specific integration of REV proviruses in avian bursal lymphomas. *Nature*, **294** (5841): 574-576 (1981).

Noori Daloii M.R., Saffari M., Raoofian R., Yekaninejad M., Saydi Dinehkabodi O., and Noori-Daloii A.R. The multidrug resistance pumps are inhibited by silibinin and apoptosis induced in K562 and KCL22 leukemia cell lines. *Leukemia Research*, **38** (5): 575-580 (2014).

Books:

Rang G.M. and Petrocelli S.R. *Fundamentals of Aquatic Toxicology.* Hemisphere Publishing Corporation, New York, 1129 p. (1991).

Chapters in Books:

Walsh J.H. Gastrointestinal hormones. In: Johnson L., Christensen J., Jackson M., Jacobson E., and Walsh J.H. (Eds.), *Physiology of the Gastrointestinal Tract*, 2nd Ed., Raven, New York, pp. 181-254 (1987).

Thesis:

Kossir A. Extraction liquid-liquid du Zinc (II) en milieu cyanure. Application ala valorization des minerais de zinc oxides, Ph.D. *Thesis*, University of Paris (VI), 116 p. (1991).

Please note that papers with incorrect formatted references will be returned.

4. Forwarding Address

Manuscripts should be submitted to the journal site (http://jscinces.ut.ac.ir).

M.R. Noori-Daloii, Ph.D.
J. Sci. I. R. Iran
University of Tehran
P.O. Box 13145-478 Tehran, I. R. Iran
http://jsciences.ut.ac.ir
E-mail: jsciences@ut.ac.ir

Journal of Sciences, Islamic Republic of Iran 35(3): 205 - 219 (2024) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Bayesian Clustering of Spatially Varying Coefficients Zero-Inflated Survival Regression Models

S. Asadi, M. Mohammadzadeh*

Department of Statistics, Tarbiat Modares University, Tehran, Islamic Republic of Iran

Received: 15 October 2024 / Revised: 1 January 2025 / Accepted: 13 January 2025

Abstract

The study addresses the challenges of analyzing time-to-event data, particularly emphasizing the discrete nature of durations, such as the number of years until divorce. This frequently results in zero-inflated survival data characterized by a notable frequency of zero observations. To address this, the study employs the zero-inflated discrete Weibull regression (ZIDWR) model, which serves as a suitable framework for evaluating the impact of explanatory variables in survival analysis. However, challenges such as nonstationarity in the relationship between variables and responses and spatial heterogeneity across geographical regions can result in a model with too many parameters To mitigate this, we propose a spatial clustering approach to summarize the parameter space. This Paper leverages nonparametric Bayesian methods to explore the spatial heterogeneity of regression coefficients, focusing on the geographically weighted Chinese restaurant process (gwCRP) for clustering the parameters of the ZIDWR model. Through simulation studies, the gwCRP method outperforms unsupervised clustering algorithms clustering K-means and the standard Chinese restaurant process (CRP), exhibiting superior accuracy and computational efficiency, particularly in scenarios with imbalanced cluster sizes. This improved performance is quantitatively demonstrated through higher Rand indices, lower average mean squared error (AMSE) in parameter estimation and superior log pseudo-marginal likelihood (LPML) values. Applying this methodology to Iranian divorce data reveals distinct spatial clusters characterized by varying covariate effects on the probability of divorce within the first five years of marriage and the subsequent time to divorce.

Keywords: Survival Analysis; Varying Coefficient; Spatial Clustering.

Introduction

Survival analysis is a statistical technique used to evaluate time-to-event data. While survival time is generally treated as a continuous random variable, it is often recorded at discrete intervals (e.g., 0, 1, 2, 3…). This discretization may result in zero observations, indicating events that occurred before the first time recording unit (e.g., daily, monthly, or yearly). These zero values, sometimes referred to as "sampling zeros"(1), arise from events that take place right at the)commencement of the study. Such occurrences are

* Corresponding Author: Tel:+989122066712; Fax:+982182883483: Email:mohsen_m@modares.ac.ir

prevalent across various domains. For example, in the healthcare sector, pregnant women might spend less than a day in the hospital before delivery. Likewise, in studies related to job placement, a zero survival time may signify an immediate job placement. Traditional survival models often struggle to accommodate these instances. As a result, researchers have turned to "zero-inflated survival models" to tackle these challenges more effectively. Applications of these models include zero-inflated Cox models for analyzing rat sleep time following ethanol exposure (2), Weibull models for investigating time until banking fraud occurs (3-4), and zero-inflated cure models employed in studies of labor duration and cervical cancer (5-6). The choice of the baseline distribution is crucial in zero-inflated discrete models. While the Poisson distribution is frequently used for its intuitive interpretation of count data, its inherent assumption of equality between mean and variance often fails in practice. This limitation leads to over- or under-dispersion, resulting in inaccurate inferences and underestimated standard errors. Although the negative binomial distribution effectively addresses over-dispersion, it is unsuitable for under-dispersed data. Furthermore, by modeling the probability of a specific number of events within a defined period and assuming independence, the Poisson distribution is not directly analogous to time-to-event distributions. Consequently, generalizing discrete distributions in survival analysis is necessary to accommodate all types of dispersion and relax the independence assumption, mainly when dealing with correlated data. The Type I Discrete Weibull distribution proposed (7) is designed to mirror its continuous counterpart, is well-suited for discrete survival data and effectively handles both over- and under-dispersion. The Zero-Inflated Discrete Weibull (ZIDW) regression model is ideal for zero-inflated discrete survival data as it captures dispersion in zero and non-zero modes (8). This model includes two regression relationships:one for the effect of explanatory variables on the rate of non-zero responses and another for the probability of zero, allowing each explanatory variable to have two regression coefficients. Considering the spatial references of survival data, known as survival spatial analysis, enables the estimation and comparison of survival across different geographical areas, revealing spatial patterns. This helps identify areas with the highest and lowest survival rates.

One notable aspect of spatial variability is the difference in the influence of explanatory variables on survival time across different locations, a phenomenon known as spatial heterogeneity. Spatial heterogeneity refers to how the relationship between explanatory and response variables alters with geographical displacement. This variation arises because different locations exhibit different properties or values. Consequently, the values of regression coefficients can differ significantly from region to region. As a result, traditional regression models may fail to accurately capture the nature of these relationships in the context of spatial data analysis. Two main methods exist for estimating regression coefficients in models with spatially variable coefficients. The first is geographically weighted regression, a local method that estimates model parameters by weighting them at any point in the examined space. Unlike conventional regression, which describes general relationships between variables, geographically weighted regression provides spatial information on the variations in these relationships. The second method treats regression coefficients as random variables following spatial distributions. The spatial distribution can be assessed by selecting appropriate prior probability functions for the parameters. (9) examined the application of a geographically weighted regression model for accelerated failure time in spatial survival data. (10) investigated the influence of explanatory variables through a geographically weighted regression model on the Cox survival models, explicitly applying the Weibull distribution to handle the data.

The second method, Spatial Variable Coefficient (SVC), addresses this spatial heterogeneity by treating regression coefficients as spatial random variables. This method considers regression coefficients as spatial random variables that follow spatial distributions. By selecting suitable prior probability functions for the parameters, it is possible to assess the spatial variability of the parameters at different locations in the Bayesian spatially-varying coefficient (BSVC) model and to estimate the regression coefficients from their posterior probability (11). Simulation studies indicate that SVC processes outperform GWR by accurately estimating regression coefficients, so GWR must be considered a purely exploratory tool (12). In survival data analysis, the spatial variable coefficients (SVCs) method in the Cox model with a frequency-oriented perspective has been suggested by (13). (14) have suggested an AFT model with prior spatial distributions for spatially variable regression coefficients. (15) have suggested a geographically weighted Cox regression model for sparse spatial survival data.

Although these methods enhance survival prediction accuracy by considering the spatial variability of regression coefficients, they also raise model fitting complexity for ZIDWR models, assuming spatial variability of parameters for both regression coefficient vectors. This is because not only can the effect of explanatory variables on non-zero response have spatial

variability, but their effect on the probability of zeroing can also be different in different places. Thus, all regression coefficients have spatial variability.

Moreover, there are often censored observations in these data due to limited follow-up time that cannot be overlooked. Clustering the model parameters with similar spatial features is a proper method for efficiently reducing model dimensions and summarizing data. Spatial clustering methods, such as the K-means method, can be used to summarize data efficiently. Hence, Bayesian nonparametric processes, such as the Dirichlet mixture process, are used to investigate the spatial heterogeneity of regression coefficients (16). This process simultaneously considers intra-cluster correlation and heterogeneity between clusters in the spatial clustering structure. These processes turn the model into a simple parametric form by clustering with complex data on the model parameter space. Given its computational ease, the Dirichlet Process is one of the best random processes among nonparametric processes (for instance, the Gaussian process, Pólya tree process (17), and so on) for clustering parameters in models with SVCs. In this method, we can cluster coefficients into homogeneous groups by choosing prior probabilities on the distribution of discrete partitions, where several parameters get the same value simultaneously.

As a representation of the infinite mixture Dirichlet process, the Chinese Restaurant Process (CRP) introduced by (18,19) allows for dividing model parameters into homogeneous clusters without predetermined assumptions about the cluster count. As our purpose of spatial clustering of parameters is to reduce the spatial heterogeneity in the data, it is necessary to consider the geographical location of the units in the allocation of clusters because the presence of common factors in close areas causes the parameters in them to be similar. In other words, each member's allocation within each cluster will be such that if that member is closer to the other cluster members regarding geographical distance, it has a better chance of being in that cluster. Hence, the distance between regions has a significant role in this clustering. (20) offers a compelling alternative:the distance-dependent Chinese restaurant process (ddCRP). This model directly incorporates the probability of assigning data points to existing clusters, making the assignment dependent on the distance between data points. An excellent way to do this is to make this function one of the weighting functions in geographically weighted regression models. Recently (21) introduced a Geographically Weighted Chinese Restaurant Process (gwCRP) to analyze the spatial heterogeneity of regression coefficients. This method simultaneously considers intra-cluster correlation and spatial clustering structure heterogeneity and estimates the number of clusters using a nonparametric Bayesian approach. While recent studies have explored the spatial heterogeneity of regression coefficients in count data models (22) and zero-inflated models (23), the spatial clustering of coefficients within survival models incorporating both zero-inflated and right-censoring remains an uncharted area of research.

Here, we demonstrate the adaptability of the geographically weighted Chinese restaurant process (gwCRP) clustering method for zero-inflated and right-censored survival models and show that compared to traditional CRP and k-means methods, gwCRP consistently estimates the number of clusters regarding distances while maintaining precise parameter estimation of each component of our two-part generalized linear regression model. To our knowledge, we are the first to introduce the spatial varying coefficients in the ZIDW regression model. Finally, we demonstrate how a Zero-Inflated Discrete Weibull (ZIDW) model, incorporating covariates such as the husband's employment status, wife's financial autonomy, age gap, and spousal similarity, could best fit the data. By spatial analysis, we will reveal significant regional variations in the effects of these covariates on both the probability of early divorce and the duration of marriage when divorce occurs later. Our novel approach, combining survival analysis with spatial clustering, provides a more nuanced understanding of divorce than traditional CRP and k-mean methods and offers valuable insights for targeted policy interventions.

The remainder of the paper is organized as follows. Section 2 summarizes ZIDW regression models. Section 3 defines the variability of SVC regression coefficients in ZIDW survival data and provides an overview of CRP and gwCRP methods. Section 4 presents the Bayesian analysis with a Gibbs sampling algorithm for clustering parameters of the ZIDW model with spatial variability of regression coefficients. Section 5 compares the existing methods in a simulation study. Then, numerical results on divorce data are presented in Section 6.

Materials and Methods

Let the random variable T has a discrete Weibull distribution $T \sim DW(q, \beta)$ with probability mass function $f(t) = P(T = t) = q^{t^\beta} - q^{(t+1)^\beta}, t = 0,1,2, \ldots$ One uses the discrete Weibull regression model with some link functions of the parameters q or β to consider the effects of some covariates on T. To define a ZIDW regression model, let the survival time T be a non-negative random count variable with the probability mass function

$$P(T=t \mid X,Z) = \begin{cases} p(Z)+(1-p(Z))(1-q(X)), & t=0 \\ (1-p(Z))\left(q(X)^{t^{\beta}}-q(X)^{(t+1)^{\beta}}\right), & t=1,2,\dots \end{cases}$$

denoting by $T \mid X,Z \sim \text{ZIDW}(p(Z),q(X),\beta)$, where the parameters $q \equiv q(X)$ and $p \equiv p(Z)$ depend on the covariates $X_{n\times(m_1+1)} = \left(1,X_1,\dots,X_{m_1}\right)$ and $Z_{n\times(m_2+1)} = \left(1,Z_1,\dots,Z_{m_2}\right)$, respectively, through the link functions (24):

$$\log\left(-\log(q(X))\right) = X'\boldsymbol{\alpha},\ \Rightarrow\ q \equiv q(X) = e^{-e^{X'\alpha}} \quad (1)$$

$$\text{logit}(p(Z)) = Z'\boldsymbol{\gamma},\ \Rightarrow\ p \equiv p(Z) = \frac{e^{Z'\gamma}}{1+e^{Z'\gamma}} = \left(1+e^{-Z'\gamma}\right)^{-1} \quad (2)$$

where $\alpha = \left(\alpha_0,\dots,\alpha_{m_1}\right)$ and $\boldsymbol{\gamma} = \left(\gamma_0,\dots,\gamma_{m_2}\right)$ are the vectors of regression coefficients. The ZIDW regression models assume that the effect of explanatory variables on the response variable is the same in different places. However, other conditions in each region may cause spatial heterogeneity. Here, we consider the spatial variability for all regression coefficients. Let $T_{\ell i}$, for $i=1,\dots,n$, and $\ell=1,\dots,n_i$ denote the survival time for the case ℓ at site $s_i = (u_i,v_i)$, n_i denotes the number of subjects at site s_i, and $X_\ell(s_i), Z_\ell(s_i)$ are the vectors of covariates. Let $T_{\ell i} \mid X,Z \sim ZIDW(p_{\ell i}(Z),q_{\ell i}(X),\beta)$, then the equations (1) and (2) considering the spatial variability of regression coefficients $\alpha_{\ell i}$ and $\gamma_{\ell i}$ will be as follows:

$$p_{\ell i}(Z) = \frac{e^{Z_\ell(s_i)\gamma(s_i)}}{1+e^{Z_\ell(s_i)\gamma(s_i)}},\ q_{\ell i}(X) = e^{-e^{X_\ell(s_i)\alpha(s_i)}}$$

Where $\gamma(s_i) = \left(\gamma_0(s_i),\dots,\gamma_p(s_i)\right)$ and $\boldsymbol{\alpha}(s_i) = \left(\alpha_0(s_i),\dots,\alpha_p(s_i)\right)$ are the model components that can be estimated by fitting two separate models. So

$$P\left(T_{i\ell}=t \middle| X,Z,\beta\right) = \begin{cases} \frac{1}{1+e^{Z_\ell(s_i)\gamma(s_i)}}\left[1+e^{Z_\ell(s_i)\gamma(s_i)}-e^{-e^{X_\ell(s_i)\alpha(s_i)}}\right] & t=0 \\ \frac{1}{1+e^{Z_\ell(s_i)\gamma(s_i)}}\left[\left(e^{-e^{X_\ell(s_i)}}\right)^{t^{\beta}}-\left(e^{-e^{X_\ell(s_i)}}\right)^{(t+1)^{\beta}}\right] & t=1,2\dots \end{cases}$$

Zero-Inflated Discrete Weibull (CZIDW) model, if $T_{\ell i}$ is the survival time of the ℓ-th unit and $C_{\ell i}$ the censored from the right that is independent of $T_{\ell i}$, then for a censored unit, the only available information is $C_{\ell i} < T_{\ell i}$. By defining $Y_{\ell i} = \min(T_{\ell i}, C_{\ell i})$, $\delta_{i\ell} = 1$ if $T_{\ell i} \geq C_{\ell i}$ and $J_{\ell i} = 1$, if $Y_{\ell i} = 0$, we can divide all the data, $\boldsymbol{D} = \left\{\left(T_{\ell i},\delta_{\ell i},X_\ell(s_i)\right), i=1,\dots,n,\ell=1,\dots,n_i\right\}$, as follows

$$\begin{cases} J_{\ell i}=1,\delta_{\ell i}=0 & Y_{\ell i} \text{ is zero and not right-censored} \\ J_{\ell i}=0,\delta_{\ell i}=0 & Y_{\ell i} \text{ is non-zero and not right-censored} \\ J_{\ell i}=0,\delta_{\ell i}=1 & Y_{\ell i} \text{ is non-zero and right-censored} \end{cases}$$

In this case, the likelihood of the CZIDW model can be defined as follows

$$L(\beta,\boldsymbol{\alpha},\gamma \mid n,Y,X,Z) = \Pi_{i=1}^{n}\Pi_{\ell=1}^{n_i}[F_{\ell i}+(1-F_{\ell i})(1-G_{\ell i})]^{J_{\ell i}(1-\delta_{\ell i})} \times\left[(1-F_{\ell i})\left(G_{\ell i}^{Y_{\ell i}^{\beta}}-G_{\ell i}^{(Y_{\ell i}+1)^{\beta}}\right)\right]^{(1-J_{\ell i})(1-\delta_{\ell i})}\left[1-F_{\ell i}-(1-F_{\ell i})\left(1-G_{\ell i}^{C_{\ell i}^{\beta}}\right)\right]^{\delta_{\ell i}} \quad (3)$$

where $F_{\ell i} = \left(1+e^{-Z_\ell(s_i)\gamma(s_i)}\right)^{-1}$ and $G_{\ell i} = e^{-e^{X_\ell(s_i)\alpha(s_i)}}$.

1. Clustering of Model Coefficients

For each particular location $s_i, i=1,\dots,n$, we define $\boldsymbol{\theta}(s_i) = (\boldsymbol{\alpha}(s_i)^{\top},\boldsymbol{\gamma}(s_i)^{\top})^{\top}$ the collection of parameters. CRP assumes n customers enter a Chinese restaurant with unlimited tables (5). In our setting, we assume that the n parameter vectors can be clustered into k groups, i.e., $\boldsymbol{\theta}(s_i) = \boldsymbol{\theta}_{\lambda_i} \in \{\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_k\}$, where $\lambda_i \in \{1,\dots,k\}$, with k being the total number of clusters. One popular way to model the joint distribution of $\boldsymbol{\lambda} = (\lambda_1,\dots,\lambda_k)$ is the CRP, which is an essential representation of the Dirichlet process and defines a series of conditional distributions as

$$P(\lambda_i = c \mid \boldsymbol{\lambda}_{-i}) \propto \begin{cases} \frac{n_{i,c}}{\alpha^*+i-1} & \text{existing cluster} \\ \frac{\alpha^*}{\alpha^*+i-1} & \text{new cluster} \end{cases} \quad (4)$$

where $\boldsymbol{\lambda}_{-i} = (\lambda_1,\dots,\lambda_{i-1})$ and $n_{i,c}$ is the number of elements in cluster c, and α^* is the concentration parameter of the underlying Dirichlet process. Equation (4) expresses the conditional probability of placing the i^{th} unit in the c^{th} cluster, given that the $i-1$ of the previous unit is clustered. (15,20) introduced the "geographically weighted Chinese Restaurant Process" (gwCRP) clustering method based on the weight functions of distances. So in equation (3), we have $n_{i,c} = \sum_{j=1}^{i-1} w_{ij} I(\lambda_j = c)$, where w_{ij} s are elements of the weight matrix W. Spatial weights are accommodated using a Stochastic Neighborhood Conditional Autoregressive (SNCAR) model (25), extending the conventional Conditional Autoregressive (CAR) model (26) to account for areal data. (27) defined a weight matrix based on graph distance. Assume that the whole area we are considering is a graph A with a set of vertices $V(A) = \{v_1,\dots,v_n\}$ and a set of edges $E(A) = \{e_1,\dots,e_m\}$ then the matrix elements are $w_{ij} = 1$ if $d_{v_iv_j} \leq 1$, otherwise $w_{ij} = \exp\left(-\frac{d_{v_iv_j}}{h}\right)$ where

$$d_{v_iv_j} = \begin{cases} |V(e)|, & \text{if } e \text{ is the shortest path connecting } v_i \text{ and } v_j \\ \infty, & \text{if } v_i \text{ and } v_j \text{ are not connected} \end{cases}$$

is the distance graph between A_i and A_j, and h is the bandwidth (28). Moreover, $|V(e)|$ is the cardinality of the $V(e)$ set, where e is the shortest path to the two vertices. It is evident that when $h=0$, the suggested gwCRP technique is identical to the traditional CRP technique. In this particular situation, the CRP technique tends to cluster excessively. Another significant pattern is that as h rises, the estimated number of clusters decreases before rising again. Simultaneously, the Rand

index demonstrates an initial increase followed by a decrease as h becomes excessively large. This pattern emerges because, starting from $h = 0$, the gwCRP technique effectively begins to capture the inherent spatial relationships in the data. Nevertheless, as $h \to \infty$, the geographic weights w_{ij} for spatial-discontinuous areas decrease to zero. As a result, only neighboring areas are categorized within the identical cluster, bringing back the problem of excessive clustering.

2. Bayesian Analysis

Suppose for the CZIDW model for the set of parameters $\Theta = (\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\pi}, k)$, we have separated the model parameters by to $k \leq n$. In that case, we expect that each member of the parameter space $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ where $\boldsymbol{\theta}(s_i) = \boldsymbol{\theta}_{\lambda_i}$ is equal to one of the k separate values of the separation set $\boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_{K^*}^*$ If K^* denotes the number of clusters excluding the i-th observation $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{i-1}$. Thus, if G_0 is a continuous distribution Polya Urn scheme, the conditional distribution of θ_i given $\boldsymbol{\theta}_{-i} = \{\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n\}$ will be as follows:

$$P(\theta_i \mid \boldsymbol{\theta}_{(-i)}, \alpha^*, G_0) \propto \begin{cases} \frac{1}{\alpha^* + i - 1} \sum_{k^*=1}^{K^*} \sum_{j=1}^{i-1} w_{ij}^* I(\boldsymbol{\theta}(s_j) = \boldsymbol{\theta}_{k^*}) \delta_{\theta_k^*}(\boldsymbol{\theta}(s_i)) & \text{existing cluster} \\ \frac{\alpha^* G_0(\boldsymbol{\theta}(s_i))}{\alpha^* + i - 1} & \text{new cluster.} \end{cases}$$

Where $\delta(\cdot)$ is the indicator function. Then by defining Prior hierarchically as follows:

$$T \mid X, Z, U \sim ZIDW\big(p_{\lambda_i}(Z), q_{\lambda_i}(X), \beta\big),\ i = 1, \dots, n,$$
$$\boldsymbol{\alpha}_h \sim \mathrm{N}(\mathbf{0}, \Sigma_\alpha),\ \gamma_h \sim \mathrm{N}\big(\mathbf{0}, \Sigma_\gamma\big),\ h = 1, \dots, k,$$
$$G_0(\alpha, \gamma) \propto P(\alpha)P(\gamma) = MVN(0, \Sigma_0),$$
$$\lambda_i \mid \boldsymbol{\pi}, k \sim \mathrm{Multinomial}(\pi_1, \cdots, \pi_k),$$
$$\boldsymbol{\pi} \sim \mathrm{gwCRP}(\alpha^*, h),\ k \sim P(\cdot).$$

For data $D = (Y, X, Z, J, \delta)$, with $L(\Theta \mid D)$, our goal is to sample from the posterior distribution of the parameters $k, \boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n) \in \{1, \dots, k\}, \boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_k)$, and $\gamma = (\gamma_1, \dots, \gamma_k)$. In nonparametric Bayesian models with the prior Dirichlet Processes (8), due to the unavailability of the analytical form for the posterior distribution of Θ, we employ the Gibbs sampling (27) to repeatedly draw values for each θ_i from its conditional distribution given both the data and the θ_j for $j \neq i$. Then, we combine this result with the likelihood and derive the full conditional distribution for θ_i for use in Gibbs sampling:

$$\theta_i \mid \theta_{-i}, Y \sim Q\Big[\sum_{i \neq r} L(\theta_i \mid n, Y, X_1, X_2)\delta_{\theta_r}(\theta_i) + \alpha^*\Big(\int L(\theta_i \mid n, Y, X_1, X_2) dG_0(\theta)\Big) H_i(\theta_i)\Big],$$

Where Q is the normalizer constant, $H_i(\theta)$ is the posterior distribution of θ obtained by combining information from the prior distribution G_0 and observed data D_i.

3. Cluster Configurations

Using Dahl's method introduced by (28) allows for obtaining posterior estimates of cluster memberships $\lambda_1, \dots, \lambda_n$ and other model parameters γ and α. This method selects an "average" clustering using all posterior clusterings in the three below steps:

Step 1. Define membership matrices $\mathcal{A}^{(b)} = \big(\mathcal{A}^{(b)}(i,j)\big)_{i,j \in \{1,\dots,n\}} = \Big(I\big(\lambda_i^{(b)} = \lambda_j^{(b)}\big)\Big)_{n \times n}$, where $b = 1, \dots, B$ is the index for the retained MCMC draws after burn-in, and $I(\cdot)$ is the indicator function.

Step 2. Calculate the element-wise mean of the membership matrices over MCMC draws $\overline{\mathcal{A}} = \frac{1}{B}\sum_{b=1}^{B} \mathcal{A}^{(b)}$.

Step 3. Identify the most representative posterior $\overline{\mathcal{A}}$ draw based on minimizing the element-wise Euclidean distance $\sum_{i=1}^{n}\sum_{j=1}^{n}\big(\mathcal{A}^{(b)}(i,j) - \overline{\mathcal{A}}(i,j)\big)^2$ among the retained $b = 1, \dots, B$ posterior draws.

The algorithm accuracy can be evaluated using the Rand index (29) for comparing cluster configurations obtained with different methods to the actual clusters. The Rand index computes a similarity measure between two clusterings by considering all sample pairs and counting pairs assigned in the same or different clusters in the predicted and true clusterings. This index allows us to measure the similarity between different clustering results, providing valuable insights into the match ability of these configurations. To measure the agreement between $\boldsymbol{\lambda}^{(C_{LS})}$ and the true clustering configuration. The Rand index of two partitions, $\mathcal{S}_1 = \{U_1, \dots, U_r\}$ and $\mathcal{S}_2 = \{V_1, \dots, V_s\}$, of a set of n objects $S = \{o_1, \dots, o_n\}$, is defined as

$$RI = \frac{a + b}{a + b + c + d} = \frac{a + b}{\binom{n}{2}}$$

where a represents the number of pairs of objects in set S that are in the same cluster in $\mathcal{S}_1$ and the same cluster in $\mathcal{S}_2$, b represents the number of pairs of objects in set S that are in different clusters in $\mathcal{S}_1$ and different clusters in $\mathcal{S}_2$, c represents the number of pairs of objects in set S that are in the same cluster in $\mathcal{S}_1$ and different clusters in $\mathcal{S}_2$ and d represents the number of pairs of objects in set S that are in different clusters in $\mathcal{S}_1$ and the same cluster in $\mathcal{S}_2$. The Rand index varies from 0 to 1, where a higher value signifies more excellent agreement between the two partitions. When the partitions are in complete agreement, the Rand index equals 1.

For model selection, the decaying effect parameter h for geographical weights needs to be tuned, and we use the logarithm of the Pseudo-Marginal Likelihood (30) based on conditional predictive ordinate to select h. The LPML is defined as LPML $\sum_{i=1}^{N} \log(\mathrm{CPO}_i)$, where CPO_i is the $i - the$ conditional predictive ordinate. The

Table 1. Comparison of LPML for different h values in both scenarios

	h-values				
Scenario	1	1.6	2	2.6	3
Balanced	-22402	-155942	-12658	-8360	-9653
Imbalanced	-20188	-21070	-13181	-5133	-6671

Monte Carlo estimate of the CPO, within the Bayesian framework, can be obtained as $\widehat{\mathrm{CPO}_i}^{-1} = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{f\left(D_i|\boldsymbol{\theta}_{\lambda_i}^b\right)}$, where B is the total number of Monte Carlo iterations, $\boldsymbol{\theta}_{\lambda_i}^b$ is the $b-the$ posterior sample, and $f(\cdot)$ is the likelihood function defined in (3). An estimate of the LPML can subsequently be calculated as $\widehat{\mathrm{LPML}} = \sum_{i=1}^{N}\log\left(\widehat{\mathrm{CPO}_i}\right)$. A model with a more considerable LPML value is preferred.

4. Simulation Study

A simulation study compares the K-means and the CRP clustering methods with the proposed gwCRP clustering for zero-inflated discrete time-to-event data with spatially varying covariates. The study will examine two balanced and imbalanced scenarios for data geographical clustering patterns. Under the balanced scenario, each group contains an equal number of units. Under the imbalanced scenario, the group sizes differ, and we have three to four clusters over two scenarios. The number of sites is set to the number of provinces in Iran, i.e., 31. We then generated a sample of size $n_i = 5$ for each province with center s_i, so the total number of observations is $n = 155$. We assumed X is equivalent to Z and a similar set of covariates affect q and p parameters. Then, we generated spatial covariates $X_\ell(s_i)$ from Normal distribution $N(0,1)$. The temporal component pdf for the $\ell^{th}, \ell = 1, \dots, n_i$ observation in province s_i, follows the distribution $T_{\ell i} \mid X_{\ell i} \sim \mathrm{ZIDW}(p(X_{\ell i}), q(X_{\ell i}), \beta)$, with a fixed value of $\beta = 1.2$. So, two related responses were controlled under two generalized linear models, $\mathrm{logit}(p)$ and $\log(-\log(q))$. We set initial values for model coefficient parameter α_{real}, $(-2,0.5)$, $(1.5,0.6)$, $(2.1,-0.4)$, $(1.1,0.3)$, and for γ_{real}, $(0.95,1.1)$, $(-0.4,0.6)$, $(0.5,0.8)$, $(1,1.5)$ corresponding to each of the 4 clusters, respectively. Then, to investigate the right censoring, we considered the quantile 93% of data as the censored point $C_{i\ell}$ and as a threshold to cut the simulated sample, such that all values $y_{\ell i} \geq C_{\ell i}$ were re-valued to be equal to $C_{\ell i}$. Also, if $T_{\ell i}$ is not greater than the generated censored time $C_{i\ell}$, we set $\delta_{\ell i} = 1$, otherwise, it is considered zero. To add a zero-inflated feature for each response, first, a random vector from a uniform distribution $U = (u_1, \dots, u_n) \sim U(0,1)$ is generated if $u_{\ell i} \leq p_{\ell i}$, set $J_{\ell i} = 0$ and $Y_{\ell i} = 0$ otherwise, we considered $J_{\ell i} = 1$ and generated $Y_{\ell i}$ from DW distribution. We generated the outcome data under the following two generalized linear models

$$\mathrm{logit}(p_{i\ell}) = \gamma_{0\ell}(s_i) + \gamma_{1\ell}x_\ell, \quad (5)$$

$$\log(-\log(q_{i\ell})) = \alpha_{0\ell}(s_i) + \alpha_{1\ell}(s_i)x_\ell \quad (6)$$

We used Normal prior distributions $N(0,\sigma_\alpha^2)$ for regression coefficients α_0 and α_1, with precision parameters, $\sigma_\alpha^{-2} \sim T(10^{-5}, 10^{-5})$. Similarly, for γ_0 and γ_1, the Normal priors $N\left(0,\sigma_\gamma^2\right)$, are considered, respectively, with $\sigma_\gamma^{-2} \sim T(10^{-5}, 10^{-5})$. To assess gwCRP's clustering performance across a range of h values, we will evaluate it from 1 to 3 in a grid of 0.2. The optimal value of h will be determined using LPML (Table 1). We fixed the concentration parameter $\alpha^* = 1$. We provide information on estimating the number of clusters and the compatibility of clustering configurations. The maximum distance in the spatial structure of the 31 regions is 10k. m, so yielding an optimal bandwidth $\left(h_{\text{opt}} = 2.6\right)$ induces a weighting scheme that ensures relative weights are assigned appropriately. Each replicate involves running an MCMC chain of length 10,000 with a thin of one and burn-in of 2,000 samples.

Results

After meticulously examining the MCMC chain length, we run our proposed algorithm in 100 separate data replicates. A vital part of this process is obtaining 100 RI values, which we then compare with the real values to validate our results. We calculated the mean in the 100 replicates and the posterior means of the parameters. Each replicate runs a total of MCMC iterations. We calculated the cover rate for each scenario, which equals the percentage of replicates in which our proposed algorithm accurately recovers the number of clusters. In our gwCRP model for scenario 2, we observe that the correct number of clusters is inferred in at least 25 out of 100 instances. Specifically, for model 1 under scenario 2, the final estimate of the number of clusters consistently reaches five across 90 replicates. However, in scenario 1, 75 cases underestimate the number of clusters by 10. We also provide a detailed comparison of our method with the K-means Algorithm. As the K-means algorithm cannot infer the number of clusters,

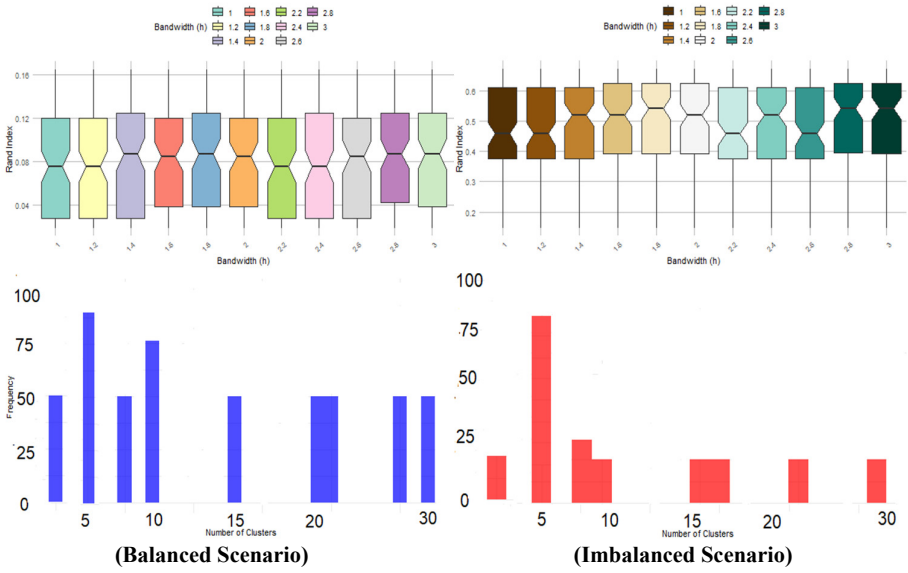


Figure 1. Histogram of estimates of k under h-optimal and box plot of Rand index under different h and LPML selection.

Table 2. RI indexes for different clustering methods ($h_opt = 2.6$)

Type	CRP	K-means	gwCRP
Balanced	0.8931	0.6865	0.9344
Imbalanced	0.8122	0.7624	0.9581

such values must be pre-specified. We supplied them with the number of clusters inferred by our method in each replicate, providing a comprehensive understanding of their differences. We also present the histogram of the final number of clusters inferred for each cluster scenario and data generation model combination in (Figure 1).

As the Gibbs sampler does not directly yield the posterior distribution of k, we employed Dahl's method to estimate it. The RIs for each scenario and data generation model are reported in Table 2. We also thoroughly compared our method to the K means algorithm. As the K-means algorithm cannot infer the number of clusters, such values need to be pre-specified, and we supplied them with the number of clusters inferred by our method in each replicate. As the Gibbs sampler does not directly yield the posterior distribution of k, we employed Dahl's method to estimate it. Table 2 demonstrates the significant improvement in computational efficiency offered by our proposed gwCRP model with vectorization for both scenarios. This enhancement, coupled with the model's highest RI, underscores its innovative approach and high accuracy in clustering. The K-means model, while having an RI greater than 0.6, does not match the performance of our proposed model. Furthermore, using the optimal value of h determined by LPML has resulted in excellent clustering performance. In addition to assessing clustering performance, we also evaluate the estimation performance of covariate coefficients.

Let $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$ be the actual clustering label vector, $\boldsymbol{\theta}_r(s_i)$ be the true parameter value of cluster $j, \kappa_r = \sum_{i=1}^{31} I(\lambda_i = r)$ be the number of provinces in cluster r (where $r = 1, \dots, k$ and $\sum_{r=1}^{k} \kappa_r = n$). For the simulated dataset t, let $\widehat{\boldsymbol{\theta}}_{(t)}(s_i)$ be the estimate of Dahl's method at location s_i. Then, the average of mean squared error (AMSE) is calculated as

$$\text{AMSE} = \frac{1}{k}\sum_{r=1}^{k}\frac{1}{\kappa_r}\sum_{i|\lambda_i=r}\frac{1}{100}\sum_{b=1}^{100}\left(\widehat{\boldsymbol{\theta}}_{(b)}(s_i) - \boldsymbol{\theta}_r(s_i)\right)^2$$

Which calculates mean squared errors for each cluster first and then averages across clusters. Table 3 presents the AMSE results for parameter estimation of gwCRP using optimal values of h in two different scenarios. Table 3 presents the AMSE results for parameter estimation of gwCRP using optimal values of h in two different scenarios. Generally, the K-mean method has a higher AMSE than other methods. Our research Identifies a pattern in clustering performance, showing that gwCRP exhibits a lower AMSE than traditional

Table 3. Performance of parameter estimates under the two true cluster scenarios with AMSE (h = 2.6)

	Balanced				Imbalanced			
Method	α_0	α_1	γ_0	γ_1	α_0	α_1	γ_0	γ_1
gwCRP	0.0115	0.0266	0.0029	0.0350	0.0023	0.0023	0.0143	0.0009
CRP	0.0268	0.0214	0.1059	0.0901	0.0137	0.0319	0.0269	0.0401
K-mean	0.1810	0.0815	0.0297	0.3112	0.0997	0.1069	0.1018	0.1704

CRP. This result indicates the importance of selecting the optimal h based on LPML for accurate estimation. The AMSE fluctuates more in the balanced scenario than the imbalanced scenario;in this scenario, AMSE values are lower overall due to being mis-clustered.

In conclusion, our simulation studies clearly show that the gwCRP models outperform the standard CRP models in terms of clustering accuracy and parameter estimation. Our proposed model selection criterion, the LPML, effectively identifies the optimal h value, yielding superior results for clustering and parameter estimation tasks. These conclusions should convince the audience of the strength of our research findings.

The computational costs of our different clustering methods vary significantly. K-means has a time complexity of 37,200 units and is faster when the number of clusters is pre-defined, but it cannot automatically determine the optimal number of clusters. The Chinese Restaurant Process (CRP) has a more complex time complexity of around 24,025 units due to its iterative evaluation of potential cluster assignments, making it less efficient for larger datasets. In contrast, the proposed method, gwCRP, utilizes vectorization and optimized techniques, achieving a time complexity of approximately 930 units per iteration, leading to faster convergence. Additionally, the use of the LPML criterion helps identify the optimal value for parameter (h), further enhancing efficiency. In conclusion, while K-means is computationally efficient for fixed clusters but lacks flexibility, CRP is more adaptable but computationally intensive. The gwCRP method offers a balance of robust clustering performance and improved efficiency. Simulation studies confirm that gwCRP outperforms standard CRP in clustering accuracy and parameter estimation, with carefully designed parameters reflecting realistic scenarios in geographical data, highlighting the strengths of the proposed research.

1. Analysis of Divorce Data

Understanding the dissolution of marriages is crucial in addressing the social issue of divorce through Survival analysis. Recent studies show a worrying inflation of divorce in the first five years of marriage. To further investigate, we have partitioned the time axis into six 5-year periods, $[0,5), [5,10), \ldots, [30,35)$. The starting points of these intervals, namely $0,1, \ldots,6$, define the discrete survival times. Fifty couples who had experienced one or more marriages between 1989 and 2019 were selected from each of Iran's 31 provinces. The final dataset comprised 1,550 couples, of which 874 had experienced divorce. Other couples who did not experience divorce by the end of 2019 were considered

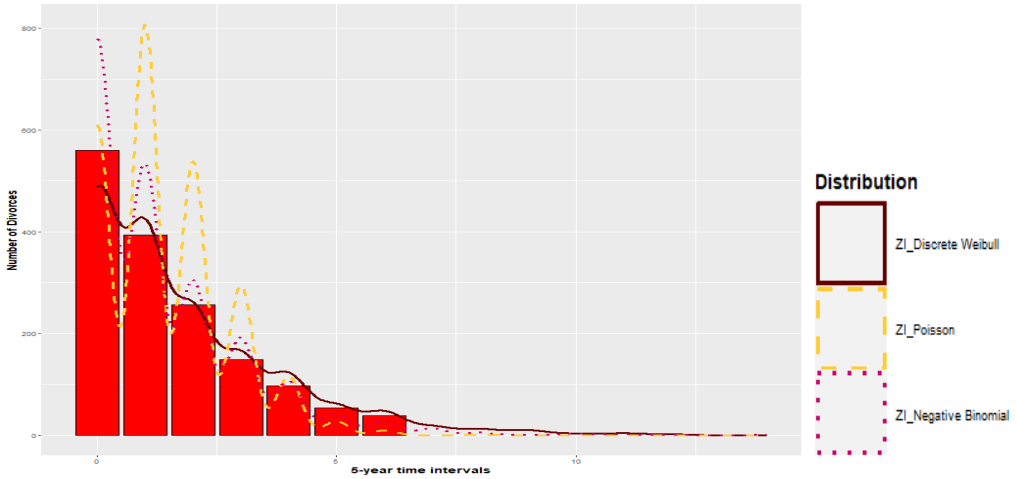


Figure 2. Histogram and Zero-inflated distributions of marriage duration among couples.

Table 4. Demographic characteristics.

Variable	Group	Number	Percent
Husband's employment status	Fixed	514	8.3
	Temporary	907	58.5
	Unemployed	129	33.2
Wife's Financial Autonomy	Independent	1128	72.8
	Dependent	422	27.2
Age gap	Less than fifteen years	1128	72.8
	More than fifteen years	422	27.2
Similarity	No	1283	82.8
	Yes	267	17.2

right-censored data. Approximately %36 of divorces occurred within the initial five years of marriage. Consequently, it is imperative to employ censored zero-inflated discrete distributions to model the data. The dispersion index represents the ratio of the observed variance from the data to the observed mean. In this case, the dispersion index equals 1.69, indicating over-dispersion in the data. Three distributions, namely Zero-Inflated Discrete Weibull (ZIDW), zero-inflated negative Binomial, and zero-inflated Poisson, have been fitted to the time data to reach the divorce event. Based on the observation in (Figure 2), it is evident that the ZIDW is better suited for these data than the other two distributions.

Due to the multidimensional nature of the divorce issue and the existence of various economic, social, cultural, demographic, etc. factors influencing the risk of divorce during the marriage period and also the probability of divorce less than five years, the demographic information of people such as the age difference of spouses, employment status are included in the CZIDW model as auxiliary variables according to (Table 4). In this study, we also examine the effect of spousal similarity.

To fit the distribution $T_{\ell i} \mid X, Z(s_i) \sim ZIDW\big(p_{\ell i}\big(X(s_i)\big), q_{\ell i}\big(X(s_i)\big), \beta\big)$ to the data, first, it is necessary to build the $5 \times n$ scenario matrix $X = (1, X_1, \dots, X_4)$, including the covariates "Husband's employment status" X_1, "Similarity" X_2, "Age gap" X_3, and "Wife's Financial Autonomy" X_4. Then we have:

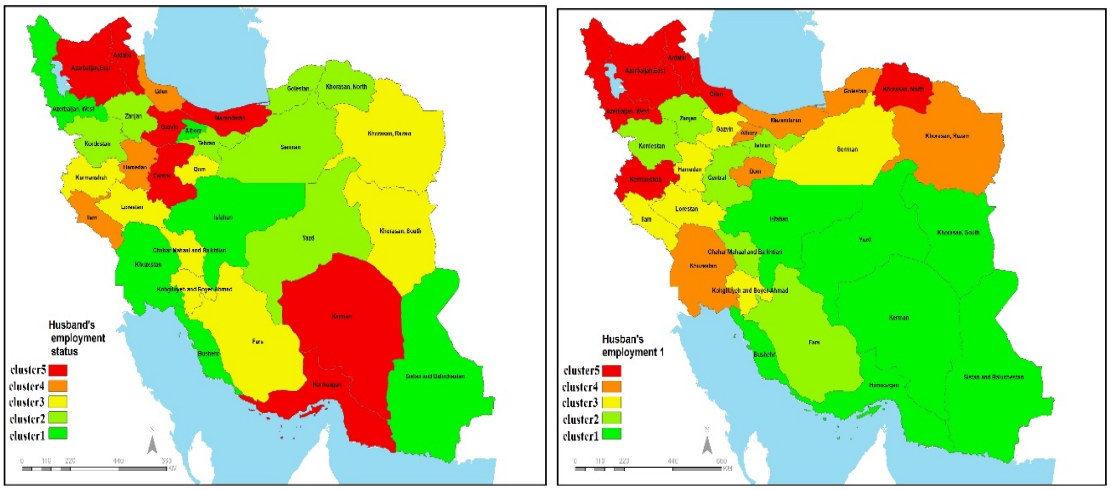
$$\log\left(-\log\left(q_{i\ell}\right) = \alpha_{0\ell}(s_i) + \sum_{m=1}^{4} \alpha_{m\ell}(s_i) x_{m\ell}(s_i)\right. \quad (7)$$

$$\operatorname{logit}(p_{i\ell}) = \gamma_{0\ell}(s_i) + \sum_{m=1}^{4} \gamma_{m\ell}(s_i) x_{m\ell}(s_i) \quad . \quad (8)$$

We first fit the two-part ZIDW model for each area using the covariates selected. Before being visualized, the covariates are adjusted to have a mean of 0 and a standard deviation of 1. According to the geographical patterns specified in (Figure 2-5) for each of the four covariates in both models, the probability of divorce in less than five years (zeroing the marriage survival time) and the duration of cohabitation provided that the couple has lived together for at least five years (non-zero count values), emphasizes the necessity of using SVC model. Also, it is seen that some provinces have similar characteristics, not limited to only adjacent counties, indicating possibilities of globally discontiguous clusters. In more detail, (Figure 3) shows significant spatial variation in divorce rates across Iranian provinces, strongly influenced by the husband's employment status (temporary, permanent, or unemployed). This variation reflects substantial socioeconomic disparities, including unemployment rates, job security, access to social services, and cultural and religious factors. These factors affect the relationship between a husband's employment and divorce probability, leading to stronger associations in some provinces than others. This is demonstrated by the varying regression coefficients for the husband's employment status across the country, as mapped in (Figure 3) for both models (6 and 7).

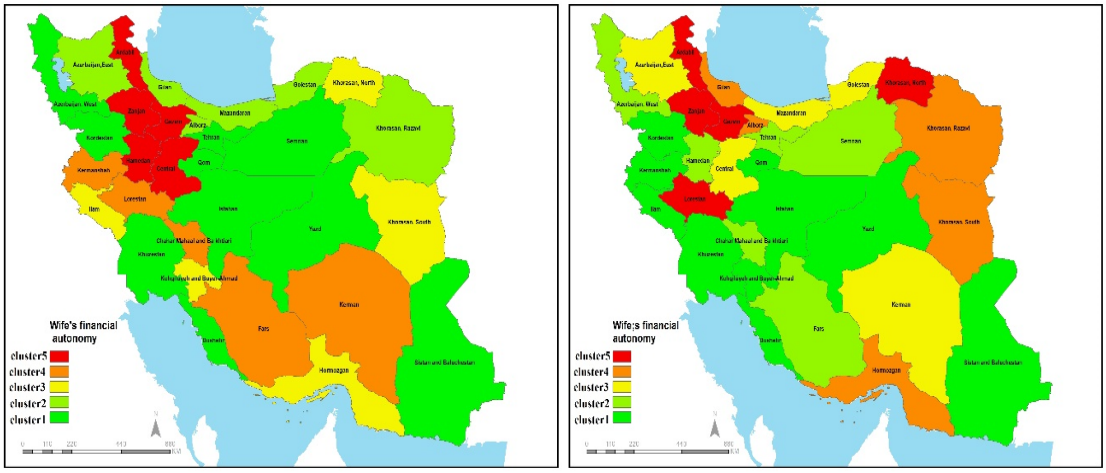
Additionally, according to spatial disparities shown in (Figure 4), the regression coefficient for both models in (7) and (8) for the wife's financial autonomy covariate in Iran is expected to vary spatially due to significant regional differences in socioeconomic development and cultural norms. More developed provinces with higher female education and employment may show weaker links between financial autonomy and divorce, while less developed, more conservative regions might exhibit stronger negative correlations, reflecting societal pressures and differing views on gender roles and responsibilities.

Moreover, As shown in (Figure 5) the impact of spousal similarity (education, socioeconomic status,



(a) (b)

Figure 3. The spatial varying covariate effects of Husband's employment on the two-part ZIDW model of provinces in Iran a:the probability of marriage survival becoming zero, b:the duration of time to divorce.



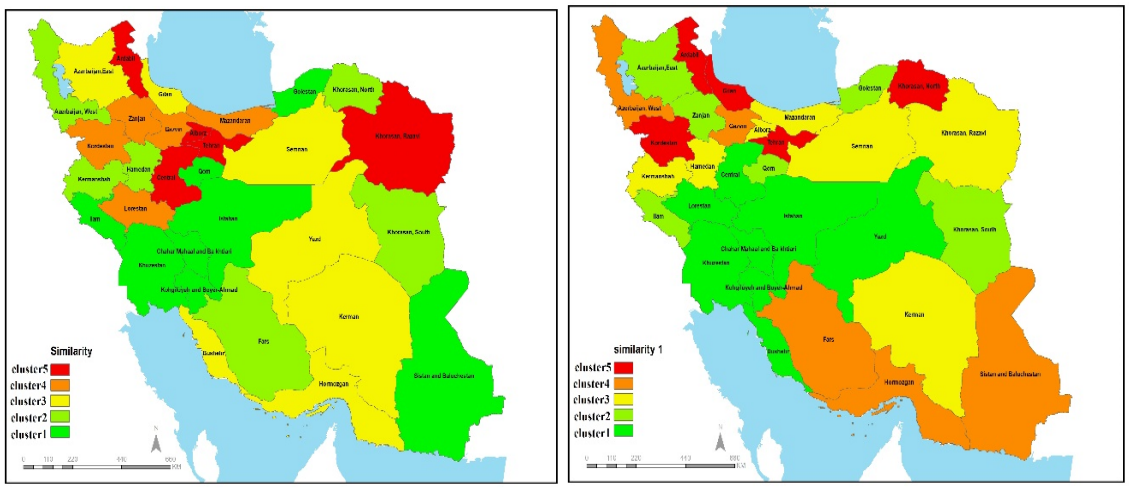
(a) (b)

Figure 4. The spatial varying covariate effects of Wife's Financial Autonomy on the two-part ZIDW model of provinces in Iran a:the probability of marriage survival becoming zero, b:the duration of time to divorce.

religious observance, ethnicity, and attitudes/personality) on time until the divorce event occurs in Iran varies significantly across provinces. For example, educational similarity is greater in provinces with higher literacy rates, while socioeconomic disparity's negative impact is stronger in provinces with high-income inequality. Similarly, religious similarity matters more in religiously conservative provinces, and ethnic similarity is more impactful in ethnically diverse regions.

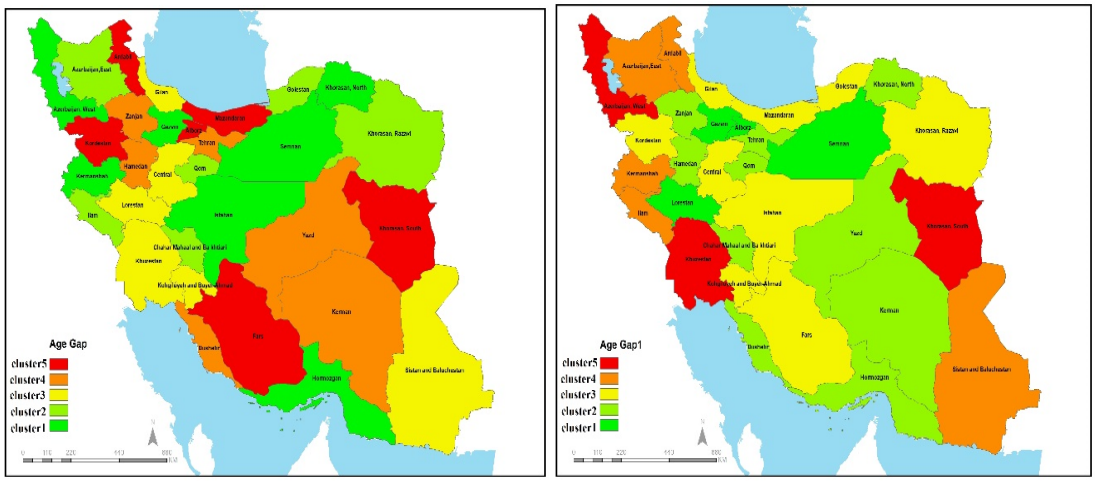
Finally, we visualize how the impact of age differences in couples varies significantly across provinces of Iran (Figure 6). Societies with traditional values or limited opportunities may show less adverse effects from larger age gaps than those with more liberal views or better opportunities.

We run 10,000 MCMC iterations, dropping the first 2000 as burn-in. We retained every fifth observation to reduce autocorrelation. We adopted a non-informative prior for the bandwidth and estimated the optimal bandwidth by the LPML method, choosing an optimal value of h at 4.2. The maximum distance between any two points is 10. The result from Dahl's method for the gwCRP model suggests that all couples are to be classified into five groups. However, our proposed



(a) (b)

Figure 5. The spatial varying covariate effects of Similarity on the two-part ZIDW model of provinces in Iran a:the probability of marriage survival becoming zero, b:the duration of time to divorce.



(a) (b)

Figure 6. The spatial varying covariate effects of the Age gap on the two-part ZIDW model of provinces in Iran a:the probability of marriage survival becoming zero, b:the duration of time to divorce.

gwCRP model, with its unique features, presents a different perspective. The sizes of the five groups in our model are7,9,4,6 and 5, respectively. The arrangement of these cluster assignments is based on Dahl's method, and the mode is depicted in (Figure 7), which illustrates their spatial distribution.

From (Figure 7), our gwCRP approach effectively identifies spatially connected and disconnected clusters. Provinces in the "light green" cluster exhibit spatial contiguity, and provinces in the "dark green" cluster display spatial discontinuity. Several interesting observations can be made from (Figure 6 and Table 5):

1. WestAzarbaijan, Kermanshah, Ilam, Khuzestan, Isfahan, Qom, Semnan, Khorasan North, Sistan, and Baluchestan all four covariates have moderate hazard effects compared with other counties.

2. East Azarbaijan, Golestan, Bushehr, Hormozgan, Kohgiluyeh, Buyer Ahmad, Chahar Mahall, and Bakhtiari starkly contrast in risk effects. Husband's employment status demonstrates significantly higher risk effects than Wife's Financial Autonomy status.

3. North Khorasan, Razavi Khorasan, Yazd, Gilan,

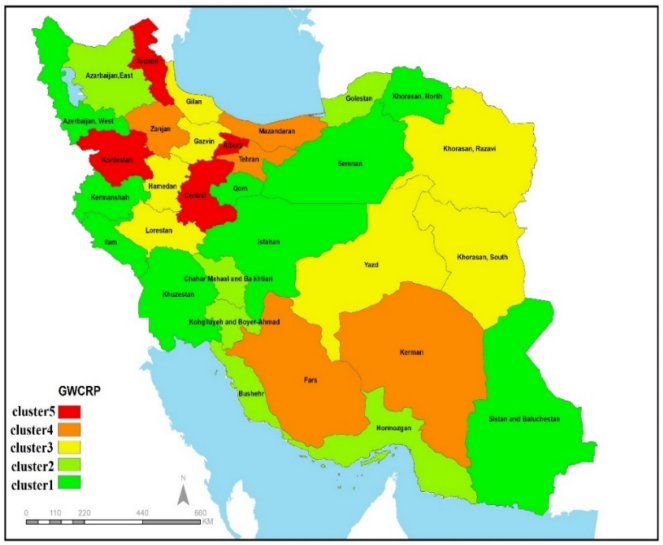


Figure 7. Clustering by gwCRP for the mean of estimated coefficients of the model

Table 5. Dahl's method estimates regression coefficients by gwCR

Cluster	$\widehat{\alpha_0}$	$\widehat{\alpha_1}$	$\widehat{\alpha_2}$	$\widehat{\alpha_3}$	$\widehat{\alpha_4}$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$
1	0.216	0.135	0.535	1.419	1.063	1.052	0.975	0.988	0.999	1.1103
2	2.492	0.364	0.785	1.668	1.110	0.965	0.981	0.065	1.000	0.3042
3	0.304	0.066	0.428	1.056	1.066	1.013	0.936	0.072	0.152	0.2587
4	1.975	-0.280	0.578	2.011	1.197	-0.259	0.753	-0.588,	-1.633	1.089
5	-0.801	0.135	0.047	-0.441	0.583	-0.417	1.539	2.444	-0.821	0.876

Ghazvin, Hamedan, and Lorestan are similar and have the highest risk effects in both model parts.

4. Mazandaran, Tehran, Kerman, and Fars:The spouses' Age differences have a negative risk effect in the model of non-zero count values and a positive effect in the probability of survival time becoming zero. The husband's employment has the most impact compared to other provinces.

5. Ardebil, Kordestan, Markazi, Alborz:The Wife's Financial Autonomy has the least hazard on the average duration of cohabitation, provided that the couple has lived together for at least 5 years.

Table 5 shows that the Bayes estimates of our spatial varying regression covariates coefficients through the gwCRP approach are quite different across different clusters.

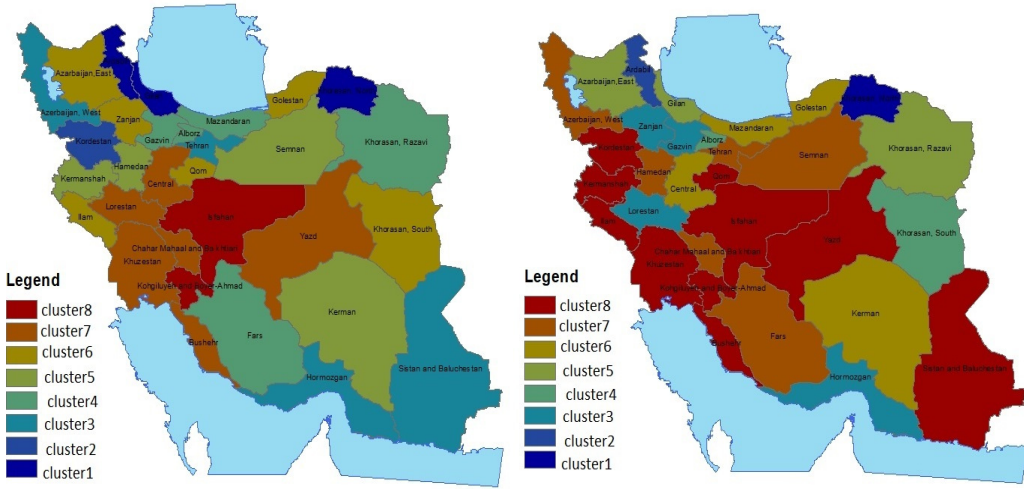
$\hat{\alpha}$, and $\hat{\gamma}$ are respectively the estimated regression coefficients for the models (7) and (8) in each cluster. They represent the effect of predictor variables (Husband's employment status, Wife's Financial Autonomy, Age gap, Similarity) on the duration of cohabitation provided that the couple has lived together for at least five years, within each cluster. For example, in Cluster 1, the estimated coefficient for Husband's employment status is 0.135. This means, that within Cluster 1, a one-unit increase in this covariate is associated with a 0.135 unit increase in c-log-log(q) in (1), holding all other variables constant. The intercept $\widehat{\alpha_0}$ represents the value of c-log-log(q) when all predictor variables are zero within that cluster. Also, in this cluster, the coefficient $\hat{\gamma}_1$ is 0.975 for the same covariate Husband's employment status in the logit model (8), suggesting a positive relationship between the Husband's employment status and the probability of divorce in less than five years represented by p. The effect is more substantial here than in the c-log-log model.

Finally, to show that our proposed clustering method, gwCRP, performs better than the two methods, traditional CRP and K-mean, in clustering regression coefficients in models 7 and 8 and determine which method yields estimation that best suits the data, the LPML values are calculated. As a more considerable LPML value indicates a better fit, we base our conclusion on the gwCRP results (Table 6).

It can be seen in (Figure 8) that the traditional K-mean method, high dimensional supervised

Table 6. LPML values for different methods in modeling divorce data

Method	gwCRP	CRP	K-mean
LPML	-367700.24	-406588.20	-416597.30



(a) (b)

Figure 8. Clustering by the k-mean with 8 clusters for the mean of a:$\hat{\alpha}$ and b:$\hat{\gamma}$

classification, and clustering, categorizes the provinces into 8 clusters, which leads to over-clustering.

To determine the optimal number of clusters (k), the Elbow method is used in k-means clustering, an empirical approach. This method is based on examining a graph that shows the value of the Within-Cluster Sum of Squares (WCSS) in terms of the number of clusters. WCSS is the sum of the squares of the distances of each data point to the center of its corresponding cluster. In (Figure 9), the horizontal axis represents the number of clusters (k), and the vertical axis represents WCSS. In this graph, the value of WCSS usually decreases as k increases. This decrease is significantly rapid at first, but after reaching a certain point (number 8), this decrease loses its speed and the downward trend becomes slower. This point, which resembles an elbow, indicates the optimal number of clusters.

Also, Comparing the clustering results using the traditional CRP method shown in (Figure 10) with the proposed method, we see that our proposed method successfully detects both spatially continuous clusters and discontinuous clusters simultaneously, however in the traditional CRP clustering method, neighboring provinces are more likely to be in the same cluster.

Discussion

In the present study, we propose an innovative Bayesian clustered coefficients regression model that employs a gwCRP to capture the spatial homogeneity of the regression coefficient proficiently. Our gwCRP models effectively address the intricate challenges associated with spatially varying coefficients in datasets characterized by right censoring and zero inflation. Through a combination of theoretical foundations and empirical evaluations, we provide compelling evidence that our methodologies yield precise parameter estimates within the ZIDW model while adeptly identifying the number of clusters and their configurations, even amidst varying proportions of zero counts. Furthermore, a comparative analysis with established clustering methodologies, such as K-means and traditional Chinese restaurant processes, illustrates that our approach achieves superior clustering concordance without additional tuning parameters, as indicated by higher Rand indexes, lower average mean squared error (AMSE), and improved log pseudo-marginal likelihood (LPML). Extensive simulation results are carried out using R version 4.3.3., to show that our proposed method has better clustering performance than the others. No issues with likelihood calculation were encountered in the simulations or the application to Iranian divorce data, however, the existence of two indicator functions often leads to extremely small likelihood functions, complicating the modeling process and requiring careful

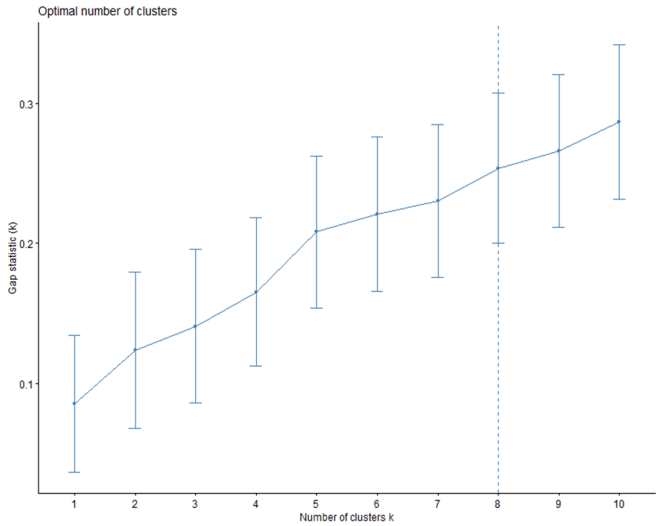
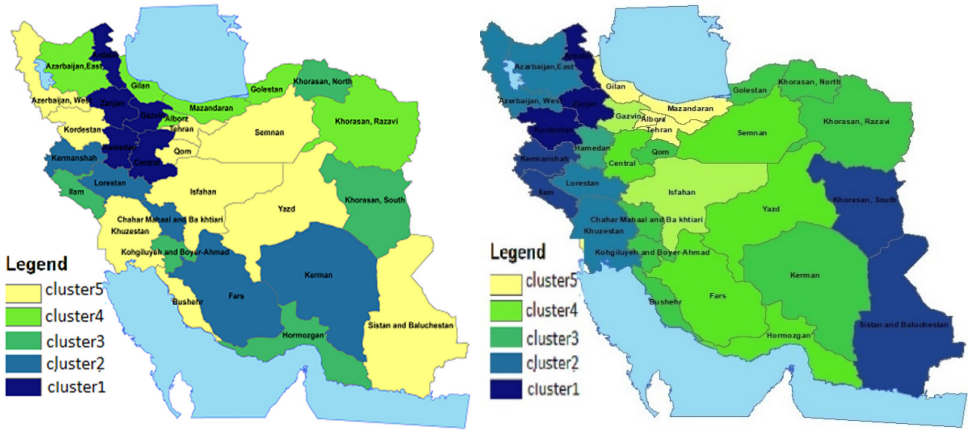


Figure 9. a: Elbow curve to determine the number of optimal clusters (k = 8), b:visualize the clustering results in the K-mean method



(a) (b)

Figure 10. Clustering by the CRP with 5 clusters for the mean of a:$\hat{\alpha}$ and b:$\hat{\gamma}$

consideration. While the discrete Weibull distribution proved beneficial for simulation data generation, using two-part regression models increased computational demands due to high-dimensional parameter spaces, resulting in extended convergence times. Furthermore, spatial heterogeneity and the inherent complexity of Bayesian hierarchical models contributed to substantial computational costs, particularly when analyzing the Iranian divorce dataset. Despite these computational challenges, our gwCRP model provides a robust and superior approach for analyzing spatially varying coefficients in complex datasets.

There are several possible directions for further investigation. The current model needs to be adapted to handle other related data (e.g., number of events) and longitudinal data (repeated measurements over time). Additionally, in this paper, our posterior sampling is based on the Chinese restaurant process, allowing for the inference of the number of clusters based on the unique latent cluster labels. To enhance the model, we suggest using a Mixed Finite Mixture (MFM) prior, allowing for the joint estimation of both regression coefficients and the probabilities of zero inflation (23) along with their associated clustering information. Finally, research is needed to improve computational efficiency, particularly for handling high-dimensional and sparse datasets, which

can be challenging to analyze.

Acknowledgments

The authors thank the editorial team and the anonymous reviewers for their comments, helpful suggestions, and encouragement, which helped improve the final version of this paper. Receiving support from the Center of Excellence in Analysis of Spatio-Temporal Correlated Data at Tarbiat Modares University is acknowledged.

References

1. Eissa S. Diagnostic biosensors for coronaviruses and recent developments. Advanced Biosensors for Virus Detection. 2022;8:261-278.
2. Moattari G, Izadi Z, Shakhsi-Niaei M. Development of an electrochemical genosensor for detection of viral hemorrhagic septicemia virus (VHSV) using glycoprotein (G) gene probe. Aquaculture. 2021; 536:736451.
3. Li J, Jin X, Feng M, Huang S, Feng J. Ulteransensitive and highly selective electrochemical biosensor for HIV gene detection based on Amino-reduced graphene oxide and b-cyclodextrin modified glassy carbon electrode. International Journal of Electrochemistry Science. 2020;15:2722-2738.
4. Zhao F, Bai Y, Cao L, Han G, Fang C, Wei S, Chen Z. New electrochemical DNA sensor based on nanoflowers of Cu3 (PO4) 2-BSA-GO for hepatitis B virus DNA detection. Journal of Electroanalytical Chemistry. 2020;867:114184.
5. Chowdhury AD, Takemura K, Li T-C, Suzuki TM, Park EY. Electrical pulse-induced electrochemical biosensor for hepatitis E virus detection. Nat, Commun. 2019;10:3737.
6. Lee T, Park SY, Jang H, Kim GH, Lee Y, Park C, Mohammadniaei M, Lee MH, Min J. Fabrication of electrochemical biosensor consisted of multi-functional DNA structure/porous au nanoparticle for avian influenza virus (H5N1) in chicken serum. Material Science Engineering. 2019;99:511-519.
7. Faria HAM, Zucolotto V. Label-free electrochemical DNA biosensor for zika virus identification. Biosensors and Bioelectronics. 2019;131:149-155.
8. Sabzi RE, Sehatnia B, Pournaghi-Azar MH, Hejazi MS. Electrochemical detection of human papilloma virus (HPV) target DNA using MB on pencil graphite electrode. Journal of the Iranian Chemical Society. 2008;5:476–483.
9. Campos-Ferreira DS, Souza EVM, Nascimento GA, Zanforlin DML, Arruda MS, Beltrão MFS, et al. Electrochemical DNA biosensor for the detection of human papillomavirus E6 gene inserted in recombinant plasmid. Arabian Journal of Chemistry. 2016;9:443–450.
10. Balvedi RPA, Castro ACH, Madurro JM, Brito-Madurro AG. Detection of a specific biomarker for epstein-barr virus using a polymer-based genosensor. International Journal of Molecular Sciences. 2014;15:9051–9066.
11. Dong S, Zhao R, Zhu J, Lu X, Li Y Qiu S, et al. Electrochemical DNA Biosensor Based on a Tetrahedral Nanostructure Probe for the Detection of Avian Influenza A (H7N9) Virus. ACS Applied Material Interfaces. 2015;7:8834–8842.
12. Shakoori Z, Salimian S, Kharrazi S, Adabi M, Saber R. Electrochemical DNA biosensor based on gold nanorods for detecting hepatitis B virus. Analytical and Bioanalytical Chemistry. 2015;407:455–461.
13. Manzano M, Viezzi S, Mazerat S, Marks RS, Vidic J. Rapid and label-free electrochemical DNA biosensor for detecting Hepatitis A virus. Biosensors and Bioelectronics. 2018;100:89-95.
14. Shawky SM, Awad AM, Allam W, Alkordi MH, EL-Khamisy SF. Gold aggregating gold: A novel nanoparticle biosensor approach for the direct quantification of hepatitis C virus RNA in clinical samples. Biosensors and Bioelectronics. 2017;92:349–356.
15. Mohammadi J, Moattari A, Sattarahmady N, Pirbonyeh N, Yadegari H, Heli H. Electrochemical biosensing of influenza A subtype genome based on meso/macroporous cobalt (II) oxide nanoflakes-applied to human samples. Analytica Chimica Acta. 2017;979:51–57.
16. Ilkhani H, Farhad S. A novel electrochemical DNA biosensor for Ebola virus detection. Analytical Biochemistry. 2018;557:151–155.
17. Marrazza G, Ramalingam M, Jaisankar A, Cheng L, Selvolini G, Vitale IA. Advancements and emerging technologies in biosensors for rapid and accurate virus detection. TrAC Trends in Analytical Chemistry. 2024;172:117609.
18. Rezaei B, Jamei HR, and Ensafi AA. An ultrasensitive and selective electrochemical aptasensor based on RGO-MWCNTs/Chitosan/carbon quantum dot for the detection of lysozyme. Biosensors and Bioelectronics. 2018;115:37-44.
19. Izadi Z, Sheikh-Zeinoddin M, Ensafi AA, Soleimanian-Zad S. Fabrication of an electrochemical DNA-based biosensor for Bacillus cereus detection in milk and infant formula. Biosensors and Bioelectronics. 2016;80:582–589.
20. Mansor NA, Zain ZM, Hamzah HH, Noorden MSA, Jaapar SS, Beni V, Ibupoto ZH. Detection of Breast Cancer 1 (BRCA1) Gene Using an Electrochemical DNA Biosensor Based on Immobilized ZnO Nanowires. Open Journal of Applied Biosensors. 2014;3:9–17.
21. Izuan J, Rashid A, Azah N. The strategies of DNA immobilization and hybridization detection mechanism in the construction of electrochemical DNA sensor : A review. Sensing and Bio-Sensing Research. 2017;16:19–31.
22. Chun L, Kim SE, Cho M, Cheo WS, Nam J, Lee DW, Lee Y. Electrochemical detection of HER2 using single stranded DNA aptamer modified gold nanoparticles electrode. Sensors and Actuators B. 2013;186:446–450.
23. Kavita V J. DNA Biosensors-A Review. Journal of Bioengineering Biomedical Science. 2017;7:1-5.
24. Gunasekaran B M, Srinivasan S, Ezhilan M, Nesakumar N. Nucleic acid-based electrochemical biosensors. Clinica Chimica Acta. 2024;559:119715.
25. Eksin E, Yildirim A, Bozoglu A, Zor E, Erdem, A. Paper-based nucleic acid biosensors. TrAC Trends in Analytical Chemistry. 2024;171:117511.

Journal of Sciences, Islamic Republic of Iran 35(3): 221 - 231 (2024) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Modeling Some Repeated Randomized Responses

M. Tarhani, M. R. Zadkarami*, S. M. R. Alavi

Department of Statistics, Faculty of Mathematical Sciences and Computer, Shahid Chamran University of Ahvaz, Ahvaz, Islamic Republic of Iran

Received: 25 September 2024 / Revised: 2 January 2025 / Accepted: 1 February 2025

Abstract

Some social surveys address sensitive topics for which respondents do not report reliable responses. Randomized response techniques (RRTs) are employed to increase privacy levels and provide honest answers. However, estimates obtained from this method tend to exhibit increased variances. Repeating randomized responses for each individual increases the sample size, and the mean of observations for each individual reduces the variance of the parameter's estimator, bringing them closer to reality. In this study, considering continuous additive repeated randomized responses (RRRs), we apply the averaged RR of each individual using the linear regression model for sensitive variable mean. Data on the income of family heads were collected from students, and each respondent was asked to randomize their responses five times. The maximum likelihood estimators of parameters are obtained by two methods. In the first method, the response variable is the first reported observation, and in the second method, we considered the averaged RR for each individual. The results emphasize that the estimators from the second method are closer to reality and have lower variance.

Keywords: Randomized Response; Repeated Randomized Response; Linear Regression Model; Continuous Sensitive Variable; Repeated Individual Observations.

Introduction

In many social sampling surveys, some questions may be sensitive to respondents, leading to insecurity in providing honest answers. A sensitive variable has a high level of social privacy or pertains to individuals' private lives. For example, research related to addiction, bribery, specific political views, socially undesirable behaviors, or income. The RR technique is a sampling process in which respondents are more willing and confident in providing honest answers to questions.

As the pioneer paper, the RR technique for sensitive binary questions was introduced in 1965 (1). It included answering a sensitive question or its complement using a Bernoulli trial (tossing a coin). Considering this trick, the sensitive answer remains hidden from the researcher, preserving the respondent's privacy. Afterward, many methods were proposed to examine sensitive qualitative data, including the unrelated response method or Simon's method (2). Many authors extended this method (3-5). Another method is the forced response technique introduced in 1971 (6). An estimate of the sensitive proportion through the maximum likelihood method was obtained using the proposed RRR (7). The RRR technique increases privacy protection and provides a truthful answer by reporting different responses by an individual. The logistic regression parameters for RR data gathered using Warner's method were estimated (8,

* Corresponding Author: Tel: 06133369509; Email: zadkarami@yahoo.co.uk, zadkarami_m@scu.ac.ir

9). Subsequently, many researchers estimated the sensitive proportion and regression parameters using various RRTs for univariate or multivariate logistic regression models (10-16). The additive RR method was used when the sensitive attribute is a discrete quantitative variable, and the mean of the item-sum technique was estimated (17-18).

For continuous sensitive variables, the mean estimate is obtained using the systematic random sampling design in the presence of a non-sensitive auxiliary variable (19). Additionally, methods for estimating the mean of sensitive variables in the presence of measurement error have been developed (20-21), and variance estimators for sensitive variables using RRT have also been proposed (22). Quantitative RRTs were investigated to enhance respondent trust (23- 24). The effect of the initial non-response on the regression estimator in panel surveys was reduced using RRT (25). By assuming truthful responses about domain membership, non-sensitive quantitative variables were estimated for specific sensitive domains (26). RRs can shorten the length of certain confidence intervals with a conditional coverage guarantee (27).

Modeling for continuous RRs is a less explored topic. In many cases, the sensitive variable is continuous. For example, income, tax evasion, expenses for election campaigns, drug or alcohol consumption during a week, student grade point averages, and financial or ethical corruption. The unrelated question design was employed in 1971 (28) to estimate the mean of the quantitative sensitive variable. The sensitive response was added to a random number of the scramble variable (a variable with known-finite mean and variance) (29). The multiplicative method was introduced by multiplying the sensitive variable by the scramble variable (30). Additive and multiplicative approaches, the optional and mixture RR methods, increase reliability and reduce bias in the reported responses (31, 32). Regression-cum-ratio estimator estimates the sensitive variable mean (33).The authors apply several estimation methods. The regression parameters using forced RRT and the EM algorithm in a Poisson distribution were estimated (34). Multiplicative RR regression parameters were estimated using the least squares method (35). The regression coefficients were estimated using the maximum likelihood method for the multiplicative design when the scramble variable was distributed as uniform (36). Later, a multiplicative RR design was applied as the dependent variable, and the regression parameters were estimated using the least squares method (37). The regression parameters for the model introduced in (39) were estimated (38). The model parameters were estimated for a generalized linear mixed effects model employing the forced-response technique (40). There are some reasons for the limited research on modeling based on RRs, including the complexity of the model and the limited packages in commonly used software. Furthermore, changing the method of randomizing responses also affects the modeling, making it more complex (40).

A privacy criterion was introduced (41). The larger this criterion, the more confidential the RRT becomes, and respondents are expected to be more willing to participate in the study. A measure for comparing quantitative RR methods based on the variance-to-privacy was proposed (42). The smaller the value, the greater the privacy for the RRT. In this paper, we use this criterion to evaluate the privacy of quantitative RRTs.

The main focus of this article is to study models for continuous RRRs. Using RRs gets the parameter estimators closer to reality and improves efficiency; however, it increases the variance of the estimators. We consider the RRR model for the mean individual observations, which can remedy the variance growth by increasing the number of responses for each respondent. It is worth mentioning that the scramble variable with a known mean should be chosen so that the true sensitive value cannot be discerned from the participant's reported value. Otherwise, they may lack confidence in providing honest answers.

We studied repeated additive RR responses from 512 students in 2018. The information included the number of family members, education, occupation, age of the family head, and the monthly income in millions of the family head. The monthly income of the family head was added to an existing random number of the scramble variable, and the result was reported, and this process was repeated five times. Regression parameter estimators were obtained using the first response of each respondent and the average of each respondent's responses, which was reported.

The remainder of the paper is structured as follows. In the second section, the parameters for the additive, multiplicative, mixture, and optional techniques are estimated when considering the normal sensitive and scramble variable(s). The third section explains the real data application. In the fourth section, simulations are performed to evaluate the parameter estimates. Their privacy is compared using thecriteria above. The discussion is in the last section.

1. Randomized Response Techniques (RRTs)

Let $Y \sim N(\mu, \sigma^2)$, and $S \sim N(\mu_s, \sigma_s^2)$ denote the sensitive variable and the scramble variable (μ_s and σ_s^2 known), respectively. We consider two cases. In the first case, to reduce response bias and enhance privacy, each respondent should add their response with a random value of the scramble variable, report only the result, and

then repeat the procedure m times. Let RR variable denote byZ, then for individual i, the j-th RR is given by: $z_{ij} = y_i + s_{ij} \quad i = 1, .., n\,, \quad j = 1, \dots, m,$

where y_i and s_{ij} denote the true value of the sensitive variable and a random value of the scramble variable for the i-th individual in j-th repeat of RR. If T denotes the predictor variable for the sensitive variable, the unbiased prediction for i-th individual in j-th RR is as follows: $t_{ij} = z_{ij} - \mu_s, \quad i = 1, .., n\,, \quad j = 1, \dots, m.$

Then, the estimation of the sensitive variable mean and its variance are calculated as follows:

$$\hat{\mu}_t = \bar{z} - \mu_s, \qquad V(\hat{\mu}) = \frac{\sigma^2 + \sigma_s^2}{n}$$

In the second case, considering the average of m RRs for each individual as observation, the predictor for i-th individual is $\bar{t}_i = \bar{z}_i - \mu_s, i = 1, .., n$, where $\overline{T} \sim N(\mu, \sigma^2 + \frac{\sigma_s^2}{m})$, and $\hat{\mu} = \bar{\bar{z}} - \mu_s\,, V(\hat{\mu}) = (\frac{\sigma^2}{n} + \frac{\sigma_s^2}{nm})$.

The matrix form of the model for the sensitive variable y is as follows:

$$y = X\beta + \varepsilon \qquad (1)$$

where X denotes the matrix of explanatory variables. Suppose the error term ε has zero mean and variance$v = \sigma_\varepsilon^2 I$, where I denote the identity matrix. Since its true value is not observable, the RR variable is used. Therefore, the model that uses the averaged RRs is as follows:

$$\bar{z} = g(X\beta) + \xi \qquad (2)$$

$g(X\beta) = X\beta + \mu_s\,, \quad \xi = \varepsilon + \delta_s\,,$

where $\delta_s = (\bar{s} - \mu_s) \sim N(0, \tau)$ is the error of selecting the scramble value, and $\tau = \frac{\sigma_s^2}{m} I$ is the covariance matrix of the vector $\bar{s}$. Assuming independence of S and Y, ξ is distributed as $N(0, \psi)$ where $\psi = \Sigma + \tau$.

The model parameters are estimated using the maximum likelihood method. The log-likelihood function is given by

$$l_Z(\beta, \sigma_\varepsilon^2) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\ln|\psi| - \frac{1}{2}(\bar{z} - X\beta - \mu_s)'\psi^{-1}(\bar{z} - X\beta - \mu_s),$$

and the maximum likelihood estimators (MLEs) of the unknown parameters are

$$\hat{\beta} = (X'X)^{-1}X'(\bar{z} - \mu_s),$$

$$\hat{\sigma}^2 = \frac{1}{n}(\bar{z} - X\hat{\beta} - \mu_s)'(\bar{z} - X\hat{\beta} - \mu_s) - {\sigma_s^2}/{m}.$$

The distribution of the regression coefficient estimators is as follows:

$$\hat{\beta} \sim N\left(\beta, \left(\sigma^2 + \frac{\sigma_S^2}{m}\right)(X'X)^{-1}\right).$$

Then, the use of RRTs can increase the variance of parameter estimates.

1.1. Additive-Scrambled RR Technique

Suppose respondents report RR variable $Z = aY + bS$ instead of the sensitive value Y where a and b are known constant values, and S denotes a random value from the independent scramble variable $S \sim N(\mu_S, \sigma_S^2)$. Then, $Z \sim N(a\mu + b\mu_S, a^2\sigma^2 + b^2\sigma_S^2)$ and, the j-th reported RR variable of individual i is $z_{ij} = ay_i + bs_{ij}\,, \quad i = 1, \dots, n\,, \quad j = 1, \dots, m.$

The predictor variable based on one observation, T, is given by $T = \frac{aY + bS - b\mu_S}{a} = \frac{Z - b\mu_S}{a} \sim N(\mu, (a^2\sigma^2 + b^2\sigma_S^2)/a^2)$, and the unbiased predictor of Y using m repetitions is $\overline{T} = \frac{\overline{Z} - b\mu_S}{a} \sim N(\mu, \sigma^2 + \frac{b^2}{ma^2}\sigma_S^2))$.

The log-likelihood function for the predictor variable is given by

$$l(\mu, \sigma^2) = -\frac{1}{2\left\{n\ln 2\pi + \ln\left(\sigma^2 + \frac{b^2}{ma^2}\sigma_S^2\right) + \frac{1}{\left(\sigma^2 + \frac{b^2}{ma^2}\sigma_S^2\right)}(\overline{T} - \mu)'(\overline{T} - \mu)\right\}}.$$

The MLEs of the model parameters are $\hat{\mu} = \bar{\bar{t}}$ and $\hat{\sigma}^2 = \frac{(\overline{T} - \hat{\mu})'(\overline{T} - \hat{\mu})}{n} - \frac{b^2}{a^2}\sigma_S^2$ and, the variance of the estimator of $\hat{\mu}$ is $V(\hat{\mu}) = V(\bar{\bar{t}}) = \frac{V(\overline{Z})}{na^2} = \frac{\sigma^2}{n} + \frac{b^2\sigma_S^2}{nma^2}$.

Let the model error for the sensitive variable distribute as $\varepsilon \sim N(0, \sigma^2 I)$. Due to the lack of the latent variable, its predictor variable, $\bar{t}$, is used. The model is given by

$$\bar{t} = X\beta + \varepsilon^*, \qquad (3)$$

in which $\varepsilon^* \sim N(0, \sigma_{\varepsilon^*}^2 I)$, where $\sigma_{\varepsilon^*}^2 = (a^2\,\sigma^2 + \frac{b^2\sigma_S^2}{m})/a^2$.

The log-likelihood function for the RR model is

$$l(\beta, \sigma^2) = -1/2\left\{n\ln 2\pi + \ln\left(\sigma^2 + \frac{b^2}{ma^2}\sigma_S^2\right) + \frac{1}{\left(\sigma^2 + \frac{b^2}{ma^2}\sigma_S^2\right)}(\overline{T} - X\beta)'(\overline{T} - X\beta)\right\}.$$

The MLEs of the proposed model parameters are as follows:

$$\hat{\beta} = (X'X)^{-1}X'\overline{T}, \text{ and } \hat{\sigma}^2 = \frac{1}{n}\left(\overline{T} - X\hat{\beta}\right)'\left(\overline{T} - X\hat{\beta}\right) - \frac{b^2\sigma_S^2}{ma^2}.$$

The distribution of the regression coefficients estimators is given by:

$$\hat{\beta} \sim N\left(\beta, \left(\sigma^2 + \frac{b^2\sigma_S^2}{ma^2}\right)(X'X)^{-1}\right).$$

1.2. Additive-Scrambled-Scrambled Technique

Suppose the respondents multiply their sensitive answer by a known constant value a, and randomly selecte two independent values, S_1 and S_2,from known scramble variables $N(\mu_{S_1}, \sigma_{S_1}^2)$ and $N(\mu_{S_2}, \sigma_{S_2}^2)$, respectively and it is reported the RR variable $Z = aY + bS_1 + cS_2$ to the researcher for two known constants b and c. Then, the reported variable Z is distributed as $Z \sim N(a\mu + b\mu_{S_1} + c\mu_{S_2}, a^2\sigma^2 + b^2\sigma_{S_1}^2 + c^2\sigma_{S_2}^2)$. The j-th reported value of Z for individual i is:

$$z_{ij} = ay_i + bs_{1ij} + cs_{2ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m.$$

The unbiased predictor variable for a single RR, T, is as follows:

$$T = \frac{aY + bS_1 + cS_2 - b\mu_{S_1} - c\mu_{S_2}}{a} = \frac{Z - b\mu_{S_1} - c\mu_{S_2}}{a} \sim N\left(\mu, \sigma^2 + \frac{b^2\sigma_{S_1}^2 + c^2\sigma_{S_2}^2}{a^2}\right),$$

where the averaged RR for each individual is:

$$\bar{T} = \frac{\bar{Z} - b\mu_{S_1} - c\mu_{S_2}}{a} \sim N\left(\mu, \sigma^2 + \frac{b^2\sigma_{S_1}^2 + c^2\sigma_{S_2}^2}{ma^2}\right).$$

The log-likelihood function is given by

$$l(\mu, \sigma^2) = -\frac{1}{2}\left\{ nln2\pi + nln\left(\sigma^2 + \frac{b^2\sigma_{S_1}^2 + c^2\sigma_{S_2}^2}{ma^2}\right) + \frac{1}{\left(\sigma^2 + \frac{b^2\sigma_{S_1}^2 + c^2\sigma_{S_2}^2}{ma^2}\right)}(\bar{T} - \mu)'(\bar{T} - \mu)\right\},$$

The MLE's of parameters are given by $\hat{\mu} = \bar{\bar{t}}$, and $\hat{\sigma}^2 = \frac{(\bar{T} - \hat{\mu})'(\bar{T} - \hat{\mu})}{n} - \frac{b^2\sigma_{S_1}^2 + c^2\sigma_{S_2}^2}{ma^2}$ where, The variance of $\hat{\mu}$ is $V(\hat{\mu}) = V(\bar{\bar{t}}) = \frac{V(\bar{Z})}{na^2} = \frac{\sigma^2}{n} + \frac{b^2\sigma_{S_1}^2 + c^2\sigma_{S_2}^2}{nma^2}$.

Consider the sensitive variable, which is defined using equation (1). Consequently, the variance of the model error is $\sigma_{\varepsilon^*}^2 = \sigma^2 + \frac{b^2\sigma_{S_1}^2 + c^2\sigma_{S_2}^2}{ma^2}$ for the predictor variable in equation (3), where $\varepsilon^* \sim N(0, \sigma_{\varepsilon^*}^2 I)$.

The log-likelihood function for estimated MLE's of model parameters is as follows:

$$l(\beta, \sigma^2) = -1/2\left\{ nln2\pi + nln\left(\sigma^2 + \frac{b^2\sigma_{S_1}^2 + c^2\sigma_{S_2}^2}{ma^2}\right) + \frac{1}{\left(\sigma^2 + \frac{b^2\sigma_{S_1}^2 + c^2\sigma_{S_2}^2}{ma^2}\right)}(\bar{T} - X\beta)'I(\bar{T} - X\beta)\right\},$$

Then, the parameters MLE's are $\hat{\beta} = (X'X)^{-1}X'\bar{T}$ and $\hat{\sigma}^2 = \frac{1}{n}(\bar{T} - X\hat{\beta})'(\bar{T} - X\hat{\beta}) - \frac{b^2\sigma_{S_1}^2 + c^2\sigma_{S_2}^2}{ma^2}$.

The distribution of the regression coefficients estimators is $\hat{\beta} \sim N\left(\beta, \left(\sigma^2 + \frac{b^2\sigma_{S_1}^2 + c^2\sigma_{S_2}^2}{ma^2}\right)(X'X)^{-1}\right)$.

1.3. Optional RR Technique

In the additive-optional RRT, respondents either report the sensitive value or add it with a random value of the scramble variable. Let $Y \sim N(\mu, \sigma^2)$, and G be the sensitive variable and a Bernoulli random variable with probability p, respectively, then the reported variable is $Z = YG + (Y + S)(1 - G)$. It shows that the sensitivity level of the variable Y is $(1 - p)$. The j-th reported value of Z for individual i is as follows:

$$z_{ij} = \begin{cases} y_i, & with\ probability\ p \\ y_i + s_{ij}, & with\ probability\ 1 - p \end{cases}, \quad i = 1, \dots, n, \quad j = 1, \dots, m$$

Then, the additive-optional RR variable has the following mixture density function

$$f_Z(Z) = p\varphi\left(\frac{y - \mu}{\sigma}\right) + (1 - p)\varphi\left(\frac{y + s - (\mu + \mu_s)}{\sqrt{\sigma^2 + \sigma_s^2}}\right). \quad (4)$$

From density (4), we have the following equation: $E_p(E_R([(Z|G)])) = E_p(pY + (1 - p)(Y + S))$.

The mean and variance of Z are $\mu_z = \mu + (1 - p)\mu_S$ and, $\sigma_z^2 = (p - p^2){\mu_s}^2 + (1 - p)\sigma_s^2 + \sigma^2$ respectively and, the MLE's of the parameters are $\hat{\mu} = \bar{z} - (1 - p)\mu_S$ and $\hat{\sigma}^2 = \left(\frac{(Z - \hat{\mu}_z)'(Z - \hat{\mu}_z)}{n} - (1 - p)\sigma_s^2 - (p - p^2){\mu_s}^2\right)$ where, the variance of $\hat{\mu}$ is:

$$V(\hat{\mu}) = \frac{(p - p^2)\mu_S^2 + (1 - p)\sigma_S^2 + \hat{\sigma}^2}{n}.V(\hat{\mu}).$$

The unbiased predictor variable, T, and the averaged RR, $\bar{T}$, have variances $\sigma_t^2 = \sigma^2 + (1 - p)\sigma_s^2$ and $\sigma_{\bar{z}}^2 = \sigma_{\bar{t}}^2 = \sigma^2 + (1 - p)\,\sigma_s^2/m + (p - p^2)\mu_S^2/m$, respectively where, $T = Z - (1 - G)\mu_S$ and, $\bar{T} = \bar{Z} -$

$(1-p)\mu_S$. Based on $\bar{T}$, the log-likelihood function for the predictor variable is:

$$l(\mu,\sigma^2) \propto \sum ln\left(\frac{1}{\sigma_{\bar{z}}} exp\left\{-\frac{1}{2}\left(\frac{\bar{z}_i-\mu_z}{\sigma_{\bar{z}}}\right)^2\right\}\right)$$
$$= \sum ln\left(\frac{1}{\sigma_{\bar{t}}} exp\left\{-\frac{1}{2}\left(\frac{\bar{t}_i-\mu}{\sigma_{\bar{t}}}\right)^2\right\}\right).$$

Furthermore, the MLE's of the parameters are:

$$\hat{\mu} = \bar{\bar{z}} - (1-p)\mu_S\,,$$
$$\hat{\sigma}^2 = \left(\frac{(\bar{t}-\hat{\mu})'(\bar{t}-\hat{\mu})}{n} - (1-p)\,\sigma_s^2/m - (p-p^2)\mu_S^2/m\right) \quad (5)$$

Consider regression equation (1) for the sensitive variable, the log-likelihood function is rewritten as

$$l(\mu,\sigma^2) = -1/2\left\{nln2\pi + nln(\sigma_{\bar{z}}^2) + \frac{1}{\sigma_{\bar{z}}^2}(\bar{Z}-(X\beta+(1-p)\mu_S))'(\bar{Z} - (X\beta+(1-p)\mu_S))\right\}$$
$$= -1/2\left\{nln2\pi + nln\left(\sigma_{\bar{t}}^2\right) + \frac{1}{\sigma_{\bar{t}}^2}(\bar{t}-X\beta)'(\bar{t}-X\beta)\right\}$$

Then, the MLE's of parameters are as follows:

$$\hat{\beta} = (X'X)^{-1}X'(\bar{Z}-(1-p)\mu_S\,),$$
$$\hat{\sigma}^2 = \frac{1}{n}\left(\bar{Z}-X\hat{\beta}-(1-p)\mu_S\right)'\left(\bar{Z}-X\hat{\beta}-(1-p)\mu_S\right) - (1-p)\,\sigma_s^2/m - \frac{(p-p^2){\mu_s}^2}{m}.$$

Therefore, the distribution of the regression coefficients estimators is given by:

$$\hat{\beta}\sim N\left(\beta,\left((1-p)\,\sigma_s^2/m + (p-p^2){\mu_s}^2/m + \sigma^2\right)(X'X)^{-1}\right).$$

1.4. Productive RR Technique

Assume that respondents multiply their sensitive value Y by a known value from the scramble variable S. Then, the RR variable is $Z = YS$, and the j-th answer for individual i is given by:

$$z_{ij} = y_i s_{ij}\,, \qquad i = 1,\dots,n\,, \qquad j = 1,\dots,m.$$

Then, the unbiased predictor variable $T = \frac{Z}{\mu_s}$ is defined for a single RR. The mean and variance estimators of the sensitive variable are $\hat{\mu} = \bar{t}$ and $\hat{\sigma}^2 = \frac{\left(\frac{(t-\hat{\mu})'(t-\hat{\mu})}{n} - \frac{\mu^2\sigma_s^2}{{\mu_s}^2}\right)}{\left(1+\frac{\sigma_s^2}{{\mu_s}^2}\right)}$ respectively when, the variance of $\hat{\mu}$ is $(\hat{\mu}) = \sigma^2 + (\frac{\mu^2+\sigma^2}{n})\frac{\sigma_s^2}{\mu_s^2}$.

The averaged RR for each individual is $\bar{T} = \frac{\bar{Z}}{\mu_s}$ which is an unbiased predictor variable with the mean and variance μ and, $\sigma^2 + (\frac{\mu^2+\sigma^2}{m})\frac{\sigma_s^2}{\mu_s^2}$, respectively. Then, the estimators of the mean and variance are $\hat{\mu} = \bar{\bar{t}}$, and $\hat{\sigma}^2 = \frac{\left(\frac{(t-\hat{\mu})'(t-\hat{\mu})}{n} - \frac{\mu^2\sigma_s^2}{m{\mu_s}^2}\right)}{\left(1+\frac{\sigma_s^2}{m{\mu_s}^2}\right)}$, respectively where, the variance of $\hat{\mu}$ is $V(\hat{\mu}) = \frac{\sigma^2}{n} + (\frac{\mu^2+\sigma^2}{mn})\frac{\sigma_s^2}{\mu_s^2}$ and the privacy level is calculated as $P_L = (\mu^2+\sigma^2)\left(\frac{\sigma_s^2}{m} + (\mu_s-1)^2\right)$.

The log-likelihood equation using the predictor variable $\bar{T}$ is given by

$$l(\mu,\sigma^2) = log\left(\int_{-\infty}^{\infty} \frac{1}{2\pi|\bar{s}|\sqrt{\frac{\sigma^2\sigma_s^2}{m}}} exp\left(-\frac{m}{2}\left(\frac{\bar{s}-\mu_s}{\sigma_s}\right)^2 - \frac{1}{2\sigma^2}\left(\frac{\bar{z}}{\mu_s}-\mu\right)^2\right) ds\right).$$

Numerical methods estimate these parameters since the likelihood equation does not lead to a closed-form solution

Considering the regression equation (1) for sensitive variable, the log-likelihood function using equation (3), is as follows:

$$l(\beta,\sigma^2) = log\left(\int_{-\infty}^{\infty} \frac{1}{2\pi|\bar{s}|\sqrt{\frac{\sigma^2\sigma_s^2}{m}}} exp\left(-\frac{m}{2}\left(\frac{\bar{s}-\mu_s}{\sigma_s}\right)^2 - \frac{1}{2\sigma^2}(\bar{T}-X\beta)^2\right) ds\right).$$

where $= \sigma_{\varepsilon^*}^2 + (\frac{\mu^2+\sigma^2}{m})\frac{\sigma_s^2}{\mu_s^2}$. The likelihood equations do not have closed-form solutions, so numerical methods are used.

2. Application

In this section, the real data is applied to investigate the proposed RR method. In the data collection, fifty random values from the normal scramble variable $S\sim N(3.6, 0.5^2)$ are selected and recorded in fifty cards. The income of the family head is one of the sensitive questions in social sciences studies. In a questionnaire, 512 bachelor's students at Shahid Chamran University were asked to report the number of family members, education level, occupation, and age of the head of the family. they also summed up the monthly income (of millions) of the family head with one of the given random

scramble values and repeated this process five times. The students randomly selected one card from the deck of 50 cards and without anyone noticing, added the income of their family head to the number on the card and returned the card to the deck. The cards were then shuffled to maintain privacy, and only the sum of two values was reported. We repeated the process five times and reported the results for each repetition. The j-th RR for an i-th individual was as follows:

$Z_{ij} = Y_i + S_{ij} \quad i = 1, \dots, 512, \quad j = 1, \dots, 5.$

Considering the RR Model (case one), the MLEs of the mean and variance of the family head income were obtained as $\hat{\mu}_Y = 3.50662$ and $\hat{\sigma}^2 = 1.97$, respectively. The explanatory variables include the number of family members, the age of the family head, the level of education (coded as a binary variable: 1 for university attendance and 0 for non-attendance), and the occupation of the family head. Occupation is treated as a nominal variable with five categories: "others" (used as the reference level), "self-employed," "doctor," "engineer," and "retired or deceased".

The results summarized in Table 1 indicate that the number of family members and the age of the family head were not statistically significant. Considering "others" as the reference level for the occupation, levels of "doctor" and "engineer" had a significant impact on income compared to employees. The results also showed that having university attendance compared to non-attendance led to a significant increase in income.

Table 1 also shows that the family head jobs "doctor" and "engineer" had a significantly increasing effect on family head income compared to "others". However, "self-employed" and "retired" did not significantly affect family head income compared with "others". Our findings indicated that the variance of the sensitive variable was estimated at 1.83.

For RR model 1 (case two), the estimated mean and variance of the family head income were $\hat{\mu}_Y = 3.579$ and, $\hat{\sigma}^2 = 1.932$, respectively. The estimated parameters and their significant levels are presented in Table 2 where the estimated variance of the sensitive variable is 2.09.

The results of Tables 1 and 2, are consistent with previous ones; however, the standard error of estimates decreased (Table 2).

3. Simulation Study

For the models presented in Section 2, simulation and comparison were conducted using privacy criteria. Let $\beta_0 = 5$ and $\beta_1 = 2$, and the covariate X and model error ε were generated from normal $N(1,4)$ and $N(0,1)$, respectively. Therefore, the sensitive variable had a normal distribution of $N(5 + 2x, 1)$. On the other hand, the distribution of the scramble variables must be such that their mean falls within the parameter space of the

Table 1. Estimated Parameters of the RR model (case one)

Parameter		Coefficient	SE	p-value
Intercept		2.46	0.65	< .001
Age		0.0091	0.01	0.414
Education	non-attendance	-	-	-
	university attendance	1.13	0.2	< .001
Family number		-0.04	0.06	0.447
Occupation of	others	--	--	--
the family head	self-employed	-0.16	0.21	0.449
	Doctor	2.6	0.48	< .001
	Engineer	2.37	0.37	< .001
	Retired	0.1	0.35	0.765

Table 2. Estimated Parameters for averaged RR (case two).

Parameter		Coefficient	SE	95% CI	t -value	p-value
Intercept		2.16	0.62	[1.18, 3.75]	3.51	< .000
Age		0.017	0.01	[-0.01, 0.03]	1.6	0.11
Education	Non-attendance	-	-	-	-	-
	University attendance	1.146	0.19	[0.73, 1.53]	5.96	< .001
Family number		-0.042	0.055	[-0.16, 0.07]	-0.76	0.45
Occupation of the	Others	--	--	--	--	--
family head	self-employed	-0.2	0.2	[-0.57, 0.25]	-1	0.317
	Doctor	2.77	0.46	[1.65, 3.55]	6.08	< .001
	Engineer	2.28	0.35	[1.64, 3.11	6.45	< .001
	Retired	0.076	0.33	[-0.57, 0.78]	0.23	0.815

Table 3. MSE and bias of parameter Estimations for additive-scrambled RR.

	n	σ_Y	β_1	β_0
Est.	100	0.944	2	4.995
Bias		-0.056	0?	-0.005
MSE		0.046	0.019	0.104
Est.	50	0.893	1.993	5.021
Bias		-0.107	-0.007	0.021
MSE		0.11	0.04	0.219
Est.	20	0.817	1.993	5.005
Bias		-0.183	-0.007	0.005
MSE		0.177	0.07	0.378

Table 4. MSE and bias of parameter Estimations for averaged-additive-scrambled RR.

	n	σ_Y	β_1	β_0
Est.	100	0.984	2.001	5
Bias		-0.016	0.0006	0.0004
MSE		0.0094	0.0095	0.049
Est.	50	0.96	2.001	4.998
Bias		-0.04	0.0008	-0.002
MSE		0.019	0.019	0.105
Est.	20	0.9	2.006	4.98
Bias		-0.101	0.006	-0.021
MSE		0.052	0.054	0.279

sensitive variable. The parameters were estimated using the maximum likelihood method. The simulation was repeated K times, and the results included the average parameter estimatesand the bias and mean squared error (MSE) of these estimates.

The simulations are as follows:

3.1 We considered the additive model with one scramble variable. This variable was sampled from normal $N(6,4)$. We consider $a = 3$ and $b = 2$, so the RR variable has a normal distribution of $Z \sim N(27 + 6x, 25)$. For $m = 5$ times repeat of RR for each individual, the averaged RR variable has a normal distribution of $\bar{Z} \sim N(27 + 6x, 12.2)$.

Tables 3 and 4 present the simulation results for $k = 2000$ repetitions for both RR and averaged RR models, respectively.

3.2. We considered the additive-scrambled-scrambled RR model with two scramble variables. The scramble data were generated from a normal distribution of $S_1 \sim N(6,4)$ and$S_2 \sim N(8,16)$. Setting $a = 3, b = 2$ and, $c = 2$, the RR variable $Z = aY + bS_1 + cS_2$ had a normal distribution of $N(43 + 6x, 89)$. The mean of m = 5 times the repeat of RR for each individual had a normal distribution of $\bar{Z} \sim N(27 + 6x, 25)$. Simulation results for both cases are provided in Tables 5 and 6, respectively.

3.3. Given a normal distribution $N(6,4)$, and a sensitivity level of 0.6, we used an optional RRT model. So, the probability of answering the sensitive variable was $p = 0.4$. The regression model is as follows:
$y_i = 5 + 2x_i + \varepsilon_i$, $\quad i = 1, \dots, n$, $\varepsilon \sim N(0,1)$.

Table 5. MSE and bias of parameter Estimations for additive-scrambled-scrambled RR.

	n	σ_Y	β_1	β_0
Est.	100	0.793	1.997	5.012
Bias		-0.207	-0.003	0.012
MSE		0.398	0.0675	0.364
Est.	50	0.731	2.003	4.986
Bias		-0.27	0.003	-0.014
MSE		0.55	0.138	0.752
Est.	20	0.707	2.016	4.938
Bias		-0.293	0.016	-0.061
MSE		0.662	0.254	1.32

Table 6. MSE and bias of parameter Estimations for averaged-additive-scrambled-scrambled RR.

	n	σ_Y	β_1	β_0
Est.	100	0.953	2	5.001
Bias		-0.047	0?	0.001
MSE		0.043	0.019	0.103
Est.	50	0.893	2.008	4.975
Bias		-0.106	0.0078	-0.025
MSE		0.102	0.04	0.211
Est.	20	0.743	1.991	5.037
Bias		-0.257	-0.009	-0.037
MSE		0.243	0.114	0.607

The optional RR variable had a mean of $8.6 + 2X$ and variance of 12.04. The mean of RRs for $m = 5$ observations per individual had the same mean and a variance of 10.12. Parameter estimatesand their MSE and biases are provided for k =2000 repetitions in Tables 7 and 8.

3.4. Finaly, simulation results were provided for the multiplicative RR. The scramble variable data were sampled from $N(6,4)$, so the mean and variance of the multiplicative RR variable were $\mu_z = 30 + 12x$ and$\sigma_z^2 = 16(5 + 2x)^2 + 52$, respectively.

The mean of the multiplicative RRs for $m = 5$ observations per individual had the same mean and variance $\sigma_z^2 = 3.2(5 + 2x)^2 + 39.2$. The simulation results are provided for k=2000 repetitions in Tables 9 and 10.

According to the simulation results, the maximum likelihood estimates were very close to the true values with high accuracy. Moreover, as the number of simulated data, n, increases, the accuracy of estimates improves, whereas the variance and bias of the estimates decrease.

3.5 Privacy criteria

The privacy criterion, P_L, (the privacy level), is the mean squared difference between the RR, Z, and the true response Y or $P_L = E(Z - Y)^2$. The measure $\delta = \frac{V(\hat{\mu})}{P_L}$ was proposed for comparing quantitative RR methods (42). The privacy evaluation criteria for single and repeated observations of techniques are presented in Table 11. For each n, the techniques in terms of the P_L criterion are sorted as follows:

The technique with two scramble variables is the best, and after that, the techniques are sorted as follows: The technique with one scramble variable, the multiplicative technique, and the optional technique. However, considering the δ criterion, the multiplicative technique came first, followed by the techniques with one and two

Table 7. MSE and bias of parameter Estimates for optional RR.

	n	σ_Y	β_1	β_0
Est.	100	0.791	2.005	5
Bias		-0.209	0.005	-0.001
MSE		0.318	0.079	0.444
Est.	50	0.686	2.001	5.006
Bias		-0.314	0.001	0.006
MSE		0.422	0.17	0.911
Est.	20	0.615	1.984	5.029
Bias		-0.385	-0.016	0.029
MSE		0.521	0.301	1.61

Table 8. MSE and bias of parameter Estimates for averaged-optional RR.

	n	σ_Y	β_1	β_0
Est.	100	0.924	2.001	4.998
Bias		-0.076	0.001	-0.002
MSE		0.066	0.022	0.116
Est.	50	0.849	2.005	4.991
Bias		-0.15	0.005	-0.008
MSE		0.138	0.048	0.246
Est.	20	0.717	2.007	4.985
Bias		-0.283	0.007	-0.015
MSE		0.276	0.129	0.69

Table 9. MSE and bias of parameter Estimates for multiplicative RR.

	n	σ_Y	β_1	β_0
Est.	100	3.246	2.002	4.998
Bias		2.246	0.002	-0.002
MSE		0.078	0.084	0.27
Est.	50	3.18	2.003	4.988
Bias		2.18	0.003	-0.012
MSE		0.145	0.166	0.57
Est.	20	3.039	1.993	4.983
Bias		2.038	-0.007	-0.017
MSE		0.352	0.5	1.73

Table 10. MSE and bias of parameter Estimates for averaged-multiplicative RR.

	n	σ_Y	β_1	β_0
Est.	100	1.61	1.999	5.008
Bias		0.613	-0.001	0.008
MSE		0.017	0.02	0.08
Est.	50	1.695	1.992	5.014
Bias		0.695	-0.008	0.014
MSE		0.034	0.048	0.18
Est.	20	1.67	1.997	5.01
Bias		0.67	-0.003	0.01
MSE		0.088	0.124	0.511

Table 11. Privacy criteria of the RR techniques.

	n	Privacy evaluation criteria	Results from one observation		Results for the mean of $m = 5$ observations	
			Mean	Var.	Mean	**Var.**
$Z = aY + bS,$	20	P_L	943.1	5209.87	928	**5719.02**
		δ	0.0016	3×10^{-8}	-0.0011	$\mathbf{55 \times 10^{-7}}$
	50	P_L	944.36	3193.87	932.19	**2280.68**
		δ	0.00098	54×10^{-9}	-0.00045	$\mathbf{18 \times 10^{-9}}$
	100	P_L	943.58	1597.88	930.127	**1155.3**
		δ	0.00049	69×10^{-10}	-0.00012	$\mathbf{2 \times 10^{-9}}$
$Z = aY + bS_1 + cS_2,$	20	P_L	2214.71	31578.69	2165.62	**17363.07**
		δ	0.0018	14×10^{-8}	0.0011	$\mathbf{15 \times 10^{-8}}$
	50	P_L	2219.29	17797.91	2159.43	**7626.24**
		δ	0.0011	29×10^{-9}	0.00045	$\mathbf{11 \times 10^{-9}}$
	100	P_L	2225.44	9425.88	2158.821	**3702.596**
		δ	0.00054	42×10^{-9}	0.00023	$\mathbf{13 \times 10^{-10}}$
$Z = YG + (Y + S)(1 - G),$	20	P_L	23.93	24.81	15.17	**6.13**
		δ	0.111	0.00016	0.04	**0.0001**
	50	P_L	24.05	14.77	15.14	**2.5**
		δ	0.066	37×10^{-6}	0.017	$\mathbf{10 \times 10^{-6}}$
	100	P_L	24.1	7.61	15.16	**1.21**
		δ	0.033	64×10^{-6}	0.009	$\mathbf{14 \times 10^{-7}}$
$Z = YS$	20	P_L	88.32	122.42	2280.34	**126752**
		δ	0.0004	39×10^{-10}	0.0001	$\mathbf{46 \times 10^{-11}}$
	50	P_L	87.75	47.41	2266.81	**47221.5**
		δ	0.00015	29×10^{-11}	0.00004	$\mathbf{29 \times 10^{-12}}$
	100	P_L	87.88	23.23	2271.15	**22501.78**
		δ	0.00008	37×10^{-12}	0.00002	$\mathbf{38 \times 10^{-13}}$

scramble variables, and the optional technique was the last one.

For techniques with averaged RRs, the best-performing technique in terms of the P_L criterion was the multiplicative technique.The technique with two scramble variables was the second one, followed by the technique with a single scramble variable.The last was the optional technique. The behavior of the δ criterion

for this case is consistent with the single-observation case.

When comparing privacy criteria between single-observation models and models with averaged RRs, the P_L criterion significantly increased for the multiplicative model with averaged observations. Models with one and two scramble variables showed a slight reduction in P_L, while the optional model had nearly a halving of P_L. The δ criterion favors single-observation responses across all techniques, emphasizing the preference for models with averaged RRs.

Results and Discussion

In social surveys, when studying a sensitive variable, respondents may refuse to answer questions or provide socially desirable responses. The RR techniques help mitigate this issue. The RRR technique is one approach that increases privacy levels while moderating the increase in estimates variance. When studying continuous RR data, collecting multiple observations from each increases the sample size and improves parameter estimates. Averaging the observations for each respondent helps achieve more precise estimations. Linear models are applied for the mean of observations. The findings of this study demonstrate that the averaged RRs for each individual in various RR techniques yield more accurate estimations and reduce their variance.

In the study of the family head income, modeling the RR techniques are evaluated with demographic variables, including the number of family members and age, education level, and occupation of the family head. The results of the averaged RR model indicate that the number of family members and the age of the family head are not statistically significant. Levels of “doctor” and “engineer” of occupation variable, have a significant impact on income compared to the reference category, “others”. The results also show that having a university education may lead to a significant increase in income. This finding provides a valuable avenue for further investigations in this field.

References

1. Warner SL. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. Journal of the American Statistical Association. 1965;60(309):63-9.
2. Abul-Ela A-LA. Randomized Response Models for Sample Surveys on Human Population: University of North Carolina, Chapel Hill; 1966.
3. G. Horvitz BVS, Walt R. Simmons. The unrelated question randomized response modeld. Research Triangle Institute and -National Center for Health Statistics;1967.
4. Greenberg BG, Abul-Ela A-LA, Simmons WR, Horvitz DG. The unrelated question randomized response model: Theoretical framework. Journal of the American Statistical Association. 1969;64(326):520-39.
5. Moors J. Optimization of the unrelated question randomized response model. Journal of the American Statistical Association. 1971;66(335):627-9.
6. Boruch RF. Assuring Confidentiality of Responses in Social Research: A Note on Strategies. The American Sociologist. 1971;6(4):308-11.
7. Alavi SMR, Tajodini M. Maximum likelihood estimation of sensitive proportion using repeated randomized response techniques. Journal of Applied Statistics. 2016;43(3):563-71.
8. Maddala GS. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge University Press; 1983.
9. Scheers NJ, Dayton CM. Improved estimation of academic cheating behavior using the randomized response technique. Research in Higher Education. 1987;26(1):61-9.
10. Cruyff MJLF, Böckenholt U, van der Heijden PGM, Frank LE. Chapter 18 - A Review of Regression Procedures for Randomized Response Data, Including Univariate and Multivariate Logistic Regression, the Proportional Odds Model and Item Response Model, and Self-Protective Responses. In: Chaudhuri A, Christofides TC, Rao CR, editors. Handbook of Statistics. 34: Elsevier; 2016. p. 287-315.
11. Blair G, Imai K, Zhou Y-Y. Design and Analysis of the Randomized Response Technique. Journal of the American Statistical Association. 2015;110(511):1304-19.
12. Chang CH, Cruyff M, Giam X. Examining conservation compliance with randomized response technique analyses. Conserv Biol. 2018;32(6):1448-56.
13. Chang P-C, Pho K-H, Lee S-M, Li C-S. Estimation of parameters of logistic regression for two-stage randomized response technique. Computational Statistics. 2021;36(3):2111-33.
14. Hsieh S-H, Perri PF. A Logistic Regression Extension for the Randomized Response Simple and Crossed Models: Theoretical Results and Empirical Evidence. Sociological Methods & Research. 2022;51(3):1244-81.
15. Halim A, Arshad IA, Alomair AM, Alomair MA. Estimation of hidden logits using several randomized response techniques. Symmetry. 2023;15(9):1636.
16. Sayed KH, Cruyff MJ, van der Heijden PG. The analysis of randomized response “ever” and “last year” questions: A non-saturated Multinomial model. Behavior Research Methods. 2024;56(3):1335-48.
17. Chaudhuri A, Christofides TC, Chaudhuri A, Christofides TC. Randomized response techniques to capture qualitative features. Indirect questioning in sample surveys. 2013:29-94.
18. Trappmann M, Krumpal I, Kirchner A, Jann B. Item sum: a new technique for asking quantitative sensitive questions. Journal of Survey Statistics and Methodology. 2014;2(1):58-77.
19. Qureshi MN, Balqees S, Hanif M. Mean estimation of scrambled responses using systematic sampling. 2022.
20. Khalil S, Zhang Q, Gupta S. Mean estimation of sensitive variables under measurement errors using optional RRT

models. Communications in Statistics-Simulation and Computation. 2021;50(5):1417-26.
21. Tiwari KK, Bhougal S, Kumar S, Rather KUI. Using randomized response to estimate the population mean of a sensitive variable under the influence of measurement error. Journal of Statistical Theory and Practice. 2022;16(2):28.
22. Gupta S, Qureshi MN, Khalil S. Variance estimation using randomized response technique. REVSTAT-Statistical Journal. 2020;18(2):165–76–-76.
23. Gupta S, Zhang J, Khalil S, Sapra P. Mitigating lack of trust in quantitative randomized response technique models. Communications in Statistics-Simulation and Computation. 2024;53(6):2624-32.
24. Azeem M, Ijaz M, Hussain S, Salahuddin N, Salam A. A novel randomized scrambling technique for mean estimation of a finite population. Heliyon. 2024;10(11).
25. Khan M. A randomized response technique for reducing the effect of initial non-response of the regression estimator in panel surveys. 2021.
26. Ahmed S, Shabbir J. On use of randomized response technique for estimating sensitive subpopulation total. Communications in Statistics-Theory and Methods. 2023;52(5):1417-30.
27. Kivaranovic D, Leeb H. A (tight) upper bound for the length of confidence intervals with conditional coverage. arXiv preprint arXiv:200712448. 2024.
28. Greenberg BG, Kuebler RR, Abernathy JR, Horvitz DG. Application of the Randomized Response Technique in Obtaining Quantitative Data. Journal of the American Statistical Association. 1971;66(334):243-50.
29. Warner SL. The Linear Randomized Response Model. Journal of the American Statistical Association. 1971;66(336):884-8.
30. Eichhorn BH, Hayre LS. Scrambled randomized response methods for obtaining sensitive quantitative data. Journal of Statistical Planning and Inference. 1983;7(4):307-16.
31. Gupta S, Gupta B, Singh S. Estimation of sensitivity level of personal interview survey questions. Journal of Statistical Planning and Inference. 2002;100(2):239-47.
32. Gupta S, Kalucha G, Shabbir J. A regression estimator for finite population mean of a sensitive variable using an optional randomized response model. Communications in Statistics - Simulation and Computation. 2017;46(3):2393-405.
33. Singh N, Vishwakarma G, Kumar N, Singh PH. Estimation of Mean of Sensitive Variable Using Multiplicative Scramble Variable Under Measurement Error. Journal of Statistical Theory and Practice. 2022;16:46.
34. Cao M, Breidt FJ, Solomon JN, Conteh A, Gavin MC. Understanding the drivers of sensitive behavior using Poisson regression from quantitative randomized response technique data. PloS one. 2018;13(9):e0204433.
35. Singh S, Tracy DS. Ridge regression using scrambled responses. Metron-International Journal of Statistics. 1999:147-57.
36. Strachan R, King M, Singh S. Theory and Methods: Likelihood-based Estimation of the Regression Model with Scrambled Responses. Australian & New Zealand Journal of Statistics. 1998;40(3):279-90.
37. Singh S, Joarder A, King M. Regression analysis using scrambled responses. Australian Journal of Statistics. 2008;38:201-11.
3. Rueda MD, Cobo B, Arcos A. Regression Models in Complex Survey Sampling for Sensitive Quantitative Variables. Mathematics [Internet]. 2021; 9(6).
39. Arcos A, Rueda MdM, Singh S. A generalized approach to randomised response for quantitative variables. Quality & Quantity. 2015;49(3):1239-56.
40. Fox J-P, Veen D, Klotzke K. Generalized Linear Mixed Models for Randomized
Responses. Methodology. 2018;15(1):1-18.
41. Yan Z, Wang J, Lai J. An Efficiency and Protection Degree-Based Comparison Among the Quantitative Randomized Response Strategies. Communications in Statistics - Theory and Methods. 2008;38(3):400-8.
42. Gupta S, Mehta S, Shabbir J, Khalil S. A unified measure of respondent privacy and model efficiency in quantitative RRT models. Journal of Statistical Theory and Practice. 2018;12:506-11.
43. Anthony YCK. Asking Sensitive Questions Indirectly. Biometrika. 1990;77(2):436-8.
44. Arellano-Valle R, Bolfarine H, Lachos V. Skew-normal Linear Mixed Models. 2004;3.
45. Arellano-Valle RB, Genton MG. On fundamental skew distributions. Journal of Multivariate Analysis. 2005;96(1):93-116.

Journal of Sciences, Islamic Republic of Iran 35(3): 233 - 241 (2024) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Bounds for the Varentropy of Basic Discrete Distributions and Characterization of Some Discrete Distributions

F. Goodarzi*

Department of Statistics, Faculty of Mathematical Sciences, University of Kashan, Kashan, Islamic Republic of Iran

Received: 9 July 2024 / Revised: 31 December 2024 / Accepted: 1 February 2025

Abstract

Given the importance of varentropy in information theory, and since a closed form cannot be derived for some discrete distributions, we aim to establish bounds for the varentropy of these distributions and introduce the past varentropy for discrete random variables. In this article, we first acquired lower and upper bounds for the varentropy of the Poisson, binomial, negative binomial, and hypergeometric distributions. Since the resulting upper bounds are expressed as squared logarithmic expectations, we provide an equivalent formulation using squared logarithmic difference coefficients. Similarly, we present lower bounds in terms of logarithmic difference coefficients. Furthermore, an upper bound is derived for the variance of a function of discrete reversed residual lifetime function. We also investigate inequalities involving moments of selected functions via the reversed hazard rate and characterize certain discrete distributions by the Cauchy-Schwarz inequality.

Keywords: Varentropy; Reversed hazard rate; Binomial transform; Cauchy-Schwarz inequality.

Introduction

If X is absolutely continuous with probability density function $f(x)$, then the entropy of X is given by

$$H(X) = -\int_{-\infty}^{+\infty} f(x)\log f(x)dx, \quad (1)$$

where $-\log f(X)$ is the information content of X. Notably, the existence of $H(X)$ is not guaranteed. When it exists, its values range belongs to $[-\infty, \infty]$, while the entropy of discrete random variables (RVs) does not take negative values.

It is noteworthy that the variance entropy (for short varentropy) of a RV X is given by

$$V(X) = \int_{-\infty}^{+\infty} f(x)[\log f(x)]^2 dx - [H(X)]^2. \quad (2)$$

The importance of this measure in the fields of mathematics and physics has been emphasized in various studies, such as those by (1), (2), and (3).

As an application of varentropy, we consider a system with complex network. A complex network, in reality, contains a large amount of information necessary to describe the system's behaviors. (1) stated that varentropy is utilized as a general measure of probabilistic uncertainty for a complex network in terms of the laws of thermodynamics. Next, we will mention the application of variance of entropy in computer science. One of the most significant threats internet users and cloud computing services face is denial-of-service (DDoS) attacks. The nonlinear time series model is employed to predict future network traffic states by (4) and used to predict the future values of entropy variance.

* Corresponding Author: Tel: 03145249136; Email: f-goodarzi@kashanu.ac.ir

Also, they determined prediction errors by comparing the actual variance of entropy and the predicted variance of entropy. (5) have derived an explicit formula of the varentropy measure for the invariant density of one-dimensional ergodic diffusion processes.

Furthermore, (3) found an optimal varentropy bound applicable to log-concave distributions. (2) obtained a sharp varentropy bound on Euclidean spaces for convex probability measures. Another method to compute a bound for varentropy is given in (6) and (7) via reliability theory. (8) proposed the concept of varentropy for doubly truncated RVs and extensively analyzed its theoretical properties. A method for computing varentropy measure for the order statistics is introduced by (9). (10) introduced the variance residual entropy measure. (11) and (12) obtained bounds for past varentropy for continuous RV. Also, (13) obtained a bound for residual varentropy of discrete RV. Moreover, (14) recently offered a few estimators for varentropy for a continuous RV. The lossless source coding research (15) stated that the source dispersion equals its varentropy for Markov sources.

Suppose X is a discrete RV supporting $S = \{0,1,\dots,b\}$, where b is an integer and $0 < b \leq \infty$. If we express the probability mass function (PMF) and distribution function of X by $p(x)$ and $F(x)$, respectively, then, in comparison with (1) and (2), the entropy and varentropy of a nonnegative discrete RV X are given as follows.

$$H(X) = -\sum_{x=0}^{\infty} p(x)\log p(x), \quad (3)$$

$$V(X) = \sum_{x=0}^{\infty} p(x)[\log p(x)]^2 - [H(X)]^2. \quad (4)$$

The entropy of a discrete RV is the average amount of information, measured in bits, gained from observing a single symbol.

Characterizations of distributions are essential to many researchers in applied fields. In particular, in reliability theory, given an RV that often denotes a unit's lifetime, aging functions are assigned to it and characterize this variable. Among the most used are reversed failure rate and reversed mean residual life.

One can define the reversed hazard rate of X as

$$\varphi(x) = P(X = x | X \leq x) = \frac{p(x)}{F(x)},$$

hence, $F(x)$ is specified as follows

$$F(x) = \prod_{t=x+1}^{b} (1 - \varphi(t)), \quad x = 0,\dots,b-1. \quad (5)$$

Also, the reversed mean residual lifetime is given by

$$r(x) = E(x - X | X < x) = \frac{1}{F(x-1)} \sum_{t=1}^{x} F(t-1), \quad (6)$$

with defining $r(0) = 0$. See (16) for more details.

Definition 1. (a) F is said to be decreasing reversed hazard rate (DRHR) if $\varphi(x)$ is decreasing in x.

(b) F is said to increase expected inactivity time (IEIT) if $r(x)$ increases in x.

To derive variance bounds for functions of RVs, we employ Chernoff's inequality. For a discrete RV X with PMF $p(x)$, $x = 0,1,2,\dots$, bounds for $Var[g(X)]$ can be obtained using the forward difference of $g(X)$. Notably, these bounds were derived utilizing the Cauchy-Schwarz (C-S) inequality. We utilize the following lemma to derive these bounds, as presented in (17).

Lemma 2. Let X be a nonnegative and integer-valued RV with probability function $p(x)$ *with support* $\{0,1,2,\dots\}$ and let its mean be μ. Additionally, let $g(X)$ be a real-valued function with $Var[g(X)] < \infty$. Then

$$\sigma^2 E^2[w(X)\Delta g(X)] \leq Var[g(X)] \leq \sigma^2 E[w(X)(\Delta g(X))^2], \quad (7)$$

where $\Delta g(x) = g(x+1) - g(x)$ and $w(x)$ satisfies

$$\sigma^2 p(x) w(x) = \sum_{k=0}^{x} (\mu - k) p(k). \quad (8)$$

The equality satisfies iff g is linear.

The layout of the article is as follows. In Section 1, we compare two sequences by the coefficient of variation for coding a discrete source of information with three symbols and also define past varentropy for discrete RVs and obtain it by past entropy of order ζ for the discrete case. In Section 2, we get an upper and lower bound for the varentropy of the binomial, Poisson, negative binomial, and hypergeometric distribution. An upper bound for the variance of a function of the discrete reversed residual life RV is obtained in Section 3. Furthermore, we characterize some distributions through functions that ensure reliability for discrete RVs.

Coefficient of Variation and Past Varentropy

The significance of entropy is widely recognized in information theory and various other fields. However, varentropy has received comparatively less attention. Notably, the discrete entropy (3) quantifies the average number of symbols needed to code an event generated by an information source governed by the PMF of X. Varentropy, on the other hand, quantifies the variability associated with this coding. If the entropy of two sources of information is identical, then, during coding, the number of digits needed for the codeword of a symbol is closer to the expected value for the source with the lower varentropy.

Example 1.1. *Suppose that X has the PMF* $p(0) = \frac{1}{2}$ *and* $p(1) = p(2) = \frac{1}{4}$. *Also, let Y have Poisson distribution with parameter λ; then, it is easily calculated that* $\lambda \approx 0.620675$, *we have* $H(Y) \approx 1.039721$ *and* $V(Y) = 0.515302$. *Moreover,* $H(X) \approx 1.039721$ *and*

$V(X) = 0.120112$*; hence, the coding process is better suited for sequences produced by X.*

In the process of coding a discrete source of information with three symbols with probabilities p, q, and $1 - p - q$, the quantifiers entropy and varentropy, respectively, are stated as:

$$H(p,q) = H(X) = -p\log p - q\log q - (1-p-q)\log(1-p-q), \quad (9)$$

$$V(p,q) = V(X) = (p-p^2)(\log p)^2 + (q-q^2)(\log q)^2 + (1-p-q-(1-p-q)^2)(\log(1-p-q))^2 - 2pq\log p\log q - 2p(1-p-q)\log p\log(1-p-q) - 2q(1-p-q)\log q\log(1-p-q). \quad (10)$$

We sketch $H(X)$ and $SD(X) = \sqrt{V(X)}$ defined over p and q in Figure 1. Regarding p and q, seven limit cases have no varentropy. These points are $(0,0)$, $(0,0.5)$, $(1/3,1/3)$, $(0.5,0)$, $(0.5,0.5)$, $(0,1)$, and $(1,0)$. Notice that varentropy would be zero in case $(1/3,1/3)$, with equiprobable sequences and maximum entropy.

Now, we want to check the maximum variability in the information content. For this purpose, we are looking into the behavior of $\frac{dSD(p,q)}{dp}$ and $\frac{dSD(p,q)}{dq}$. By setting these terms equal to zero, we have

$$(\log p)^2 + 2\log p - (\log(1-p-q))^2 - 2\log(1-p-q) + (2(p\log p + q\log q + (1-p-q)\log(1-p-q)))(\log(1-p-q) - \log p) = 0, \quad (11)$$

and

$$(\log q)^2 + 2\log q - (\log(1-p-q))^2 - 2\log(1-p-q) + (2(p\log p + q\log q + (1-p-q)\log(1-p-q)))(\log(1-p-q) - \log q) = 0. \quad (12)$$

Note that the seven points mentioned earlier have infinite derivative values (singular points). Thus, we apply the Newton-Raphson algorithm to obtain approximate roots of derivatives given in (11) and (12) (see (18)). The values $p = 0.0616518191$ and $q = 0.0616518191$ were obtained with an initial value $(0.06,0.06)$ to start the algorithm. It is clear that the points $(0.0616518191,0.8766963618)$ and $(0.8766963618,0.0616518191)$ also maximize $SD(p,q)$ and their values is 0.8728128309.

Additionally, we consider the intersection curves of the two surfaces of Figure 1, where $H(p,q) = SD(p,q)$. The intersection areas can be shown in Figure 2 using the implicit plot function in Maple. For example, if $p = 0.2$, then entropy and the standard deviation of the entropy are equal for values of q equal to 0.06929839562 and 0.7307016044. The range between the curves in Figure 2, $SD(p,q)$ is less than $H(p,q)$, whereas, in the points outside of this region, for example, $(p,q) = (0.1,0.1)$,

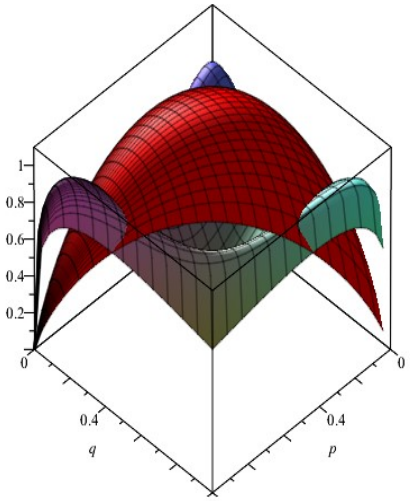


Figure 1. Plots of $H(X)$ and $SD(X)$ on p and q.

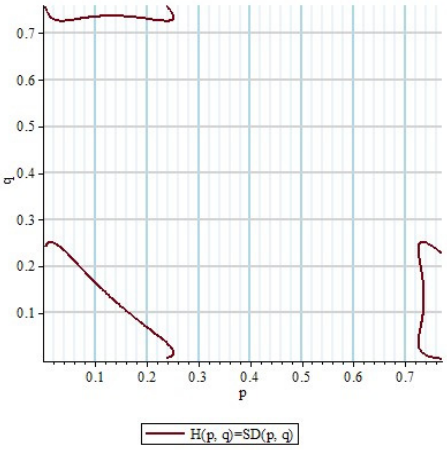


Figure 2. The curve of intersection of the two surfaces.

the entropy smaller than the standard deviation of the information content of RV. Now, considering the coefficient of variation of $-\log p(X)$, such that described as $CV(X) = \frac{SD(X)}{H(X)}$, if two sequences of symbols are generated by X and Y , a sequence with less coefficient of variation is more suitable for coding.

R $\acute{e}$ nyi entropy of order ζ for a discrete RV is expressed as $H_\zeta(x) = \frac{1}{1-\zeta}\log\sum_x p^\zeta(x)$ for $\zeta \neq 1$.

$H_\zeta(X)$ is additionally named the spectrum of R$\acute{e}$nyi information. R$\acute{e}$nyi information and the loglikelihood are related via the gradient, $\dot{H}_\zeta(X)$, of the spectrum at $\zeta = 1$. A straightforward computation demonstrates, assuming that the differentiation operations are legitimate, that

$$\dot{H}_1(X) =$$

$$\lim_{\zeta\to 1}\frac{(1-\zeta)(\sum_x p^\zeta(x))^{-1}\sum_x(p^\zeta(x)\log p(x))+\log\sum p^\zeta(x)}{(1-\zeta)^2} = -\frac{1}{2}\lim_{\zeta\to 1}\left\{\frac{\sum_x p^\zeta(x)\log^2 p(x)}{\sum_x p^\zeta(x)} - \left(\frac{\sum_x p^\zeta(x)\log p(x)}{\sum_x p^\zeta(x)}\right)^2\right\} = -\frac{1}{2}V(X). \quad (13)$$

Therefore, the varentropy is obtained as $V(X) = -2\dot{H}_1(X)$. In addition, the discrete past entropy is defined as

$$H(X;j) = -\sum_{x=0}^{j}\frac{p(x)}{F(j)}\log\left(\frac{p(x)}{F(j)}\right). \quad (14)$$

The past entropy of order ζ for a discrete case is expressed by

$$H_\zeta(X;j) = \frac{1}{1-\zeta}\log\left[\sum_{x=0}^{j}\left(\frac{p(x)}{F(j)}\right)^\zeta\right], \quad (15)$$

for $\zeta \neq 1$ and $\zeta > 0$. It is well known that when ζ tends to 1, $H_\zeta(X;j)$ tends to $H(X;j)$. Similarly, also, we

can show that $V(X;j) = -2\dot{H}_1(X;j)$, in which we call $V(X;j)$,

$$V(X;j) = \sum_{x=0}^{j}\frac{p(x)}{F(j)}\left(\log\frac{p(x)}{F(j)}\right)^2 - (H(X;j))^2, \quad (16)$$

as the past varentropy.

Example 1.2. *If X is distributed geometrically with parameter p, then*

$$H_\zeta(X;j) = \frac{1}{1-\zeta}\log\left[\sum_{x=0}^{j}\frac{p^\zeta q^{x\zeta}}{(1-q^{j+1})^\zeta}\right] = \frac{1}{1-\zeta}\left\{\log\frac{p^\zeta}{1-q^\zeta} + \log\frac{1-q^{(j+1)\zeta}}{(1-q^{j+1})^\zeta}\right\}, \quad (17)$$

where $q = 1 - p$ *and therefore,*

$$V(X;j) = -2\lim_{\zeta\to 1}\dot{H}_\zeta(X;j) = \frac{q(\ln q)^2}{(1-q)^2} - \frac{((j+1)\ln q)^2 q^{j+1}}{(1-q^{j+1})^2}. \quad (18)$$

It is observed that, for $j = 0$, the past varentropy is zero and increases for j, as shown in Figure 3.

To estimate $V(X;j)$, we generate a sample of size $n = 100$ from a geometric distribution with 1000 replicates. For this sample, we set $p_0 = 0.6$. Then, the Maximum Likelihood Estimator (MLE) for $\hat{p}$ is calculated to be 0.5978. For example by plugging $\hat{p}$ into (18) for j=1, the MLE of $V(X;j)$ is 0.1697.

Like the discrete case, (19) has previously obtained a relationship between varentropy and Rényi information for continuous RV. He expressed that varentropy can identify a distribution's shape, while the kurtosis measure is not applicable.

Bounds for Varentropy

Obtaining expressions for the entropy and varentropy of well-known distributions is significant in information and communication theory, physics, probability and statistics, and economics. An exact expression and closed form for the varentropy were obtained for most distributions. Among these distributions, we can mention the uniform, Bernoulli, geometric, exponential, Beta, Cauchy, Cramér, F, gamma, Gumbel, Laplace, Lévy, logistic, log-logistic, lognormal, normal, parabolic, Pareto, power exponential, t-distribution, triangular, von Mises and Weibull distribution. However, for many distributions, there is no closed form and an explicit expression for the varentropy using elementary functions. In such cases, we can obtain an upper and lower bound for varentropy via the expectation of a function of a logarithmic function.

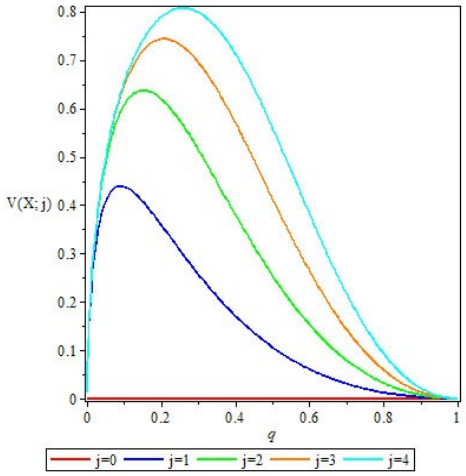


Figure 3. The curve of intersection of the two surfaces.

In this section, we find bounds for the varentropy of some nonnegative RVs. If X follows a discrete nonnegative RV, with variance σ^2, then by utilizing Lemma 2, we have

$$\sigma^2 E^2[w(X)\Delta\log p(X)] \le Var[-\log p(X)] \le \sigma^2 E[w(X)(\Delta\log p(X))^2]. \quad (19)$$

Example 2.1. *Suppose X has a binomial distribution distribution* $Bin(n,p)$ *then, since* $w(x) = \frac{n-x}{n(1-p)}$, *hence*

$$V(X) \le np(1-p)\sum_{x=0}^{n}\frac{n-x}{n(1-p)}\left(-\log\frac{(n-x)p}{(x+1)(1-p)}\right)^2\binom{n}{x}p^x(1-p)^{n-x} = np(1-p)E_{n-1}\left[\left(\log\frac{(n-X)p}{(X+1)(1-p)}\right)^2\right], \quad (20)$$

where E_{n-1} denotes expected value under the binomial distribution $Bin(n-1,p)$.

Likewise, we can derive a lower bound for

$Var[-\log p(X)]$ as follows

$$V(X) \geq np(1-p)E_{n-1}^2\left[\log\frac{(n-X)p}{(X+1)(1-p)}\right]. \quad (21)$$

If $n = 1$, then, since $-\log p(x)$ has a linear relation with x, hence the upper and lower bounds are equal to varentropy $p(1-p)(-\log(\frac{p}{1-p}))^2$ in the Bernoulli distribution.

Example 2.2. *Assume X is distributed according to a Poisson distribution with parameter $\lambda = 1$, then since $w(x) = 1$, the upper bound for varentropy is given as*

$$V(X) \leq \lambda E\left[\log\frac{X+1}{\lambda}\right]^2. \quad (22)$$

When $n \to \infty$ and $p \to 0$ so that $np = \lambda$, the upper bound (21) and (22) are approximately equal. Also,

since $\log x \leq x - 1$, we can obtain the upper bound $1 + \frac{1}{\lambda}$ for varentropy of Poisson distribution

Conversely, the lower bound for V (X) is computed as follows

$$V(X) \geq \lambda E^2\left[\log\frac{X+1}{\lambda}\right]. \quad (23)$$

In this section, we compute the equivalent expressions for the upper and lower bounds of varentropy according to series and integral expressions. To achieve a general expression for expectation of squared logarithmic function, that is, expressions like $E[\log^2(X+\omega)]$, we recall the ith forward difference of a function $g(\omega)$ is defined as

$$\Delta^i[g](\omega) := \sum_{k=0}^{i}\binom{i}{k}(-1)^{i-k}g(k+\omega), \quad (24)$$

where $\Delta^0[g](\omega) = g(\omega)$. Moreover, Newton series expansion of a function g around point ω is

$$g(k+\omega) = \sum_{i=0}^{\infty}\binom{x}{i}\Delta^i[g](\omega). \quad (25)$$

By considering $g(x) = (\log x)^2$ in equation (24), as (20) stated for $\log x$, we have

$$\Delta^i[\log]^2(\omega) = \sum_{k=0}^{i}\binom{i}{k}(-1)^{i-k}\log^2(k+\omega) = (-1)^{i+1}d_\omega(i+1), \quad (26)$$

where

$$d_\omega(i) = -\sum_{k=0}^{i-1}(-1)^k\binom{i-1}{k}\log^2(k+\omega),$$

so that, by using (26), we can obtain

$$(\log(x+\omega))^2 = \sum_{i=0}^{\infty}\binom{x}{i}(-1)^{i+1}d_\omega(i+1).$$

Now, to find the generating function for the coefficient d_ω, we use of Lerch transcendent such that it was recalled by (20), as follows:

$$\Phi(z,s,\omega) := \sum_{k=0}^{\infty}\frac{z^k}{(k+\omega)^s} = \frac{1}{\Gamma(s)}\int_0^{+\infty}\frac{t^{s-1}e^{-\omega t}}{1-ze^{-t}}dt. \quad (27)$$

By evaluating the second derivative of (27) for *s*, we have

$$\frac{d^2}{ds^2}\Phi(z,s,\omega) = \sum_{k=0}^{\infty}z^k(k+\omega)^{-s}(\log(k+\omega))^2 \quad (28)$$

and thus

$$\Phi''_\omega(z) := \frac{d^2}{ds^2}\Phi(z,s,\omega)|_{s=0} = \sum_{k=0}^{\infty}\log^2(k+\omega)z^k. \quad (29)$$

Next, putting $\omega = 1$ and using polylogarithm function $Li_s(z) := \sum_{k=1}^{\infty}z^k k^{-s}$, (29) can be written as

$$\Phi''(z) := \Phi_1''(z) = \frac{d^2}{ds^2}Li_s(z)/z|_{s=0}.$$

In fact, we have

$$\frac{d^2}{ds^2}Li_s(z)/z = \sum_{k=1}^{\infty}z^{k-1}(-\log k)^2k^{-s}. \quad (30)$$

At the same,

$$\Phi''(z) := \Phi_1''(z) = \sum_{k=0}^{\infty}z^k(\log(k+1))^2 = \sum_{k=1}^{\infty}z^{k-1}(\log(k))^2,$$

which is the equation of (30) for *s* = 0.

Using (27) and generating the function of the binomial transform, we get

$$D_\omega(z) = -\sum_{i=1}^{\infty}z^i\sum_{k=0}^{i-1}(-1)^k\binom{i-1}{k}\log^2(k+\omega) = \sum_{k=0}^{\infty}\sum_{i=k+1}^{\infty}(-1)^{k+1}z^i\binom{i-1}{k}\log^2(k+\omega) = \sum_{k=0}^{\infty}\log^2(k+\omega)(-1)^{k+1}\frac{z^{k+1}}{(1-z)^{k+1}} = \frac{-z}{1-z}\Phi''_\omega(\frac{-z}{1-z}).$$

Consider now the coefficient sequence $(-d_\omega(i+1))_{i=0}^{\infty}$, that is, the binomial transform of the sequence $(\log^2(k+\alpha))_{k=0}^{\infty}$. Let

$D_\omega(z) := \sum_{i=0}^{\infty}d_\omega(i)z^i$,

be the generating function for $(d_\omega(j))_{j=0}^{\infty}$, where $d_\omega(0)$ is defined as 0.

$$E\log^2(X+\omega) = \sum_{i=0}^{\infty}E[\frac{(X)_i}{i!}](-1)^{i+1}d_\omega(i+1) = \sum_{i=1}^{\infty}(-1)^iq(i-1)d_\omega(i).$$

The moment generating function $M(t)$ of the Poisson distribution is $\exp(\lambda(e^t-1))$, so as given in Theorem 1 in (20), we have

$$Q(z) = M(\log(z+1)) = e^{\lambda z} = \sum_{i=0}^{\infty}\frac{\lambda^i}{i!}z^i,$$

and hence $q(i) = \frac{\lambda^i}{i!}$. Furthermore, by using the equation

$Elog(X+\omega)=\sum_{i=1}^{\infty}(-1)^i q(i-1)c_\omega(i)=\int_0^{+\infty}\frac{e^{-t}-e^{-\omega t}M(-t)}{t}dt,$

in (20), where $c_\omega(i)=-\sum_{k=0}^{i-1}(-1)^k\binom{i-1}{k}\log(k+\omega)$, the lower bound

of varentropy of Poisson distribution is given as follows

$$\lambda E^2\left[\log\frac{X+1}{\lambda}\right]=\lambda\left\{\sum_{i=1}^{\infty}\frac{(-\lambda)^{i-1}}{(i-1)!}\left[\sum_{k=0}^{i-1}(-1)^k\binom{i-1}{k}\log\left(\frac{k+1}{\lambda}\right)\right]\right\}^2=\lambda\left\{\int_0^{\infty}\frac{e^{-t}}{t}(1-e^{\lambda(e^{-t}-1)})dt-\log\lambda\right\}.$$

Moreover, the upper bound for the varentropy of Poisson distribution is

$$\lambda E\left[\log\frac{X+1}{\lambda}\right]^2=\lambda\left\{\sum_{i=2}^{\infty}\frac{(-\lambda)^{i-1}}{(i-1)!}\left[\sum_{k=0}^{i-1}(\log^2(k+1)-2\log\lambda\log(k+1))\right]+\log^2(\lambda)\right\}=\lambda\left\{\sum_{i=1}^{\infty}\frac{(-\lambda)^{i-1}}{(i-1)!}\left[\sum_{k=0}^{i-1}(-1)^k\binom{i-1}{k}(\log(\tfrac{k+1}{\lambda}))^2\right]\right\}.$$

Example 2.3. *Let X follow a negative binomial distribution with a PMF* $p(x)=\binom{x+r-1}{r-1}p^r q^{x-r}$ *for* $x=0, 1, \ldots$ *. Then, since* $w(x)=p(1+\frac{x}{r})$*, the upper bound for varentropy is computed by*

$$V(X)\le\frac{r(1-p)}{p^2}\sum_{x=0}^{\infty}p(1+\frac{x}{r})(\log\frac{(r+x)(1-p)}{x+1})^2\binom{x+r-1}{r-1}p^r(1-p)^x=\frac{r(1-p)}{p^2}E_{r+1}\left[\log\frac{(1-p)(r+X)}{X+1}\right]^2, \quad (31)$$

where E_{r+1} *is the expected value of negative binomial distribution with parameters* parameters $r+1$ and p.

The lower bound for the distribution is determined as

$$V(X)\ge\frac{r(1-p)}{p^2}\left(E_{r+1}\left[\log\frac{(1-p)(r+X)}{X+1}\right]\right)^2. \quad (32)$$

It is trivial that if X has a geometric distribution with parameter p*, then varentropy is equal to the upper and lower bounds given in* (31) *and* (32) *for* $r=1$ *and hence* $\mathrm{Var}[-\log p(X)]=\frac{1-p}{p^2}(\log(1-p))^2$.

Now, by using equation (44) in (20), we can obtain an equivalent expression for the lower bound (32). We first have

$$E_{r+1}\left[\log\frac{(1-p)(r+X)}{X+1}\right]=\sum_{i=1}^{\infty}(-\frac{1-p}{p})^{j-1}\binom{i+r-1}{i-1}\left[\sum_{k=0}^{i-1}(-1)^k\binom{i-1}{k}\left\{\log(k+r)+\log\left(\frac{k+1}{1-p}\right)\right\}\right],$$

and therefore

$$Var[-\log p(X)]\ge\frac{r(1-p)}{p^2}\left(\sum_{i=1}^{\infty}(-\frac{1-p}{p})^{i-1}\binom{i+r-1}{i-1}\left[\sum_{k=0}^{i-1}(-1)^k\binom{i-1}{k}\left\{\log\frac{(1-p)(k+r)}{k+1}\right\}\right]\right)^2.$$

Also

$$Var[-\log p(X)]\le\frac{r(1-p)}{p^2}\left(\sum_{i=1}^{\infty}(-\frac{1-p}{p})^{i-1}\binom{i+r-1}{i-1}\left[\sum_{k=0}^{i-1}(-1)^k\binom{i-1}{k}\left\{\log\frac{(1-p)(k+r)}{k+1}\right\}^2\right]\right.$$

Example 2.4. *Let X have a hypergeometric distribution with PMF* $p(x)=\frac{\binom{m}{x}\binom{n-m}{r-x}}{\binom{n}{r}}$, $max(0,r-n+m)\le x\le min(r,m)$ *. Then, since* $w(x)=\frac{n(n-1)(m-x)(r-x)}{(n-m)(n-r)mr}$*, the upper bound for varentropy is computed by*

$$Var[-\log p(X)]\le\sigma^2\sum_{x=0}^{r}\frac{n(n-1)(m-x)(r-x)}{(n-m)(n-r)mr}\log(\frac{(m-x)(r-x)}{(x+1)(n-m-r+x+1)})\frac{\binom{m}{x}\binom{n-m}{r-x}}{\binom{n}{r}}=\sigma^2E_{m-1,n-2;r-1}\left[\log(\frac{(m-X)(r-X)}{(X+1)(n-m-r+X+1)})\right]^2,$$

where $E_{m-1,n-2;r-1}$ denotes expected value under the hypergeometric distribution with parameters $m-1$, $n-2$, and $\sigma^2=\frac{rm}{n}(1-\frac{m}{n})\frac{n-m}{n-1}$.

Characterization by Cauchy-Schwarz Inequality

(13) attained an upper bound for the variance of a function of the residual lifetime RV and characterized the type III and type I discrete Weibull distributions and the geometric distribution with the help of C-S inequality. Here, we derive a bound for the variance of a function of RV $X_x=(x-X|X<x)$ and characterize some distributions using inequalities involving the expectation of functions of reversed hazard rate.

The subsequent theorem gives an upper bound for $Var[g(X_x)]$ and characterizes the right-truncated geometric distribution.

Theorem 3.1. *Let X be a discrete and nonnegative RV with PMF* $p(x)$ *and distribution function* $F(x)$*. Suppose* g *is a function such that its forward difference is* $\Delta g(x)$ *then*

$$Var[g(X_x)]\le E\left[(\Delta g(X_{(x)}))^2(\frac{1}{\varphi(x-X_{(x)})}(r(x-\right.$$

$$X_{(x)}+1)-r(x)+X_{(x)}-1)\Big]. \quad (33)$$

Proof. We know

$$P(X_{(x)}=t)=\frac{P(X=x-t)}{P(X<x)}, \qquad t=1,\dots,x.$$

By applying Lemma 2 and noting that $E[X_{(x)}]=r(x)$, it follows that

$$\sum_{k=t}^{x}(k-r(x))\frac{P(X=x-k)}{P(X<x)}=\frac{1}{P(X<x)}[\sum_{k=t}^{x}kP(X=x-k)-r(x)\sum_{k=t}^{x}P\{X=x-k\}]$$
$$=\frac{P\{X<x-t+1\}}{P\{X<x\}}[r(x-t+1)+t-1-r(x)]$$
$$=\frac{1}{\varphi(x-t)}[r(x-t+1)+t-1-r(x)]\frac{P\{X=x-t\}}{P\{X<x\}}, \quad (34)$$

and again using Lemma 2 and replacing the right-hand side of (34) in inequality (7), we obtain

$$Var[g(X_x)]\le\sum_{t=1}^{x}[\Delta g(t)]^2\frac{1}{\varphi(x-t)}[r(x-t+1)+t-1-r(x)]P\{X_{(x)}=t\}$$
$$=E\Big\{[\Delta g(X_{(x)})]^2\frac{1}{\varphi(x-X_{(x)})}[r(x-X_{(x)}+1)+X_{(x)}-1-r(x)]\Big\},$$

Let $g(t)=-\log\frac{p(x-t)}{F(x-1)}$, then $\Delta g(t)=-\log\frac{p(x-t-1)}{p(x-t)}=\log(1-\eta_{x-t-1})$, hence

$$Var[-\log p(X_{(x)})]\le E\Big\{[\log(1-\eta_{x-X_{(x)}-1})]^2\frac{1}{\varphi(x-X_{(x)})}[r(x-X_{(x)}+1)+X_{(x)}-1-r(x)]\Big\}. \quad (35)$$

Under Lemma 2, the above equality holds iff $g(t)=-\log p(x-t)+\log F(x-1)$ is linear in t, which is equivalent to $\log p(x-t)$ being linear in t. Consequently, $\log p(x-t)=a_1t+b_1$ for some constants a_1 and b_1, and therefore $p(y)=e^{-a_1y}e^{a_1x+b_1}=de^{-a_1y}$ for $y=0,\dots,x-1$, where $d=e^{a_1x+b_1}$ is a constant.

We thus conclude that the equality holds in (35), iff $p(x)=\frac{c-1}{c^y-1}c^x, x=0,\dots,y-1,\ c>0$, which is the right truncated geometric distribution.

Remark 3.2. In Theorem 3.1, if X is a nonnegative RV and F is *DRHR,* then since the *DRHR* property implies the *IEIT* property (21), then

$$Var[g(X_{(x)})]\le E\Big[(\Delta g(X))^2(\frac{1}{\varphi(X)}-1)(X-1)\Big]. \quad (36)$$

Next, we aim to characterize certain distributions. Throughout the theorems presented below, we assume that Z is a discrete RV with a finite support $S=\{0,1,\dots,b\}$.

Given that $E(\frac{1}{\varphi(Z)})=b+1-E(Z)$, we can derive a useful lower bound for $E[\varphi(Z)]$, as presented in the next theorem.

Theorem 3.3. For any nonnegative discrete RV Z,

$$E[\frac{1}{\varphi(Z)}]\ge\frac{1}{E(\varphi(Z))}. \quad (37)$$

The equality satisfies iff for constant θ

$$F(z)=\begin{cases}(1-\theta)^{b-z}, & z=0,1,\dots,b,\quad 0<\theta<1,\quad b<\infty,\\ 1, & x\ge b.\end{cases} \quad (38)$$

Proof. To achieve (37), we make use of C-S inequality. The equality in (37) satisfies iff there's a positive constant A so that, for all $z\in\{0,1,\dots,b\}$,

$$\frac{\sqrt{P(Z=z)}}{\sqrt{\varphi(z)}}=A\sqrt{\varphi(z)P(Z=z)},$$

which is equivalent to $\varphi(z)=\theta=constant$. Now, using (5), we have a (38).

Theorem 3.4. Let Z be a nonnegative discrete RV. Then

$$E[\frac{\varphi(Z)}{Z}]\ge\frac{2}{b(b+1)-E(Z(Z-1))},$$

with equality iff Z distributed as,

$$F(z)=\begin{cases}\prod_{t=z+1}^{b}(1-\theta t), & z=0,1,\dots,b-1,\quad 0<\theta<\frac{1}{b},\\ 1, & z\ge b.\end{cases} \quad (39)$$

where θ is a constant.

Proof. By the C-S inequality, we have

$$1=(\sum_{z=0}^{b}P\{Z=z\}\sqrt{\frac{zF(z)}{zF(z)}})^2\le\sum_{z=0}^{b}\frac{P^2\{Z=z\}}{zF(z)}\sum_{z=0}^{b}zF(z)=\sum_{z=0}^{b}\frac{\varphi(z)}{z}P\{Z=z\}(\sum_{z=0}^{b}z-\sum_{z=0}^{b}zP\{Z>z\}). \quad (40)$$

Now, since

$$\sum_{z=0}^{b}zP\{Z>z\}=E\Big(\frac{Z(Z-1)}{2}\Big),$$

(40) reduces to

$$1\le E(\frac{\varphi(Z)}{Z})[\frac{b(b+1)}{2}-E\Big(\frac{Z(Z-1)}{2}\Big)],$$

and the desired result is obtained. The equality is gotten iff there's some positive constant θ so that

$$\frac{P(Z=z)}{\sqrt{zF(z)}}=\theta\sqrt{zF(z)}.$$

It follows that $\varphi(z)=\theta z$, which, using equation (5),

again satisfies iff Z has distribution given in equation (39).

The following two theorems derive lower bounds for $E(Z\varphi(Z))$.

Theorem 3.5. *Let Z be a discrete RV with $E(Z\varphi(Z)) < \infty$ and $E(\frac{1}{Z\varphi(Z)}) < \infty$. Then*

$$E[\frac{1}{Z\varphi(Z)}] \geq \frac{1}{E(Z\varphi(Z))}, \quad (41)$$

and equality holds iff Z is distributed as

$$F(z) = \begin{cases} \frac{(b-\theta)!z!}{b!(z-\theta)!}, & z = \theta, \dots, b-1, \quad \theta = 1,2, \dots, b-1 \\ 1, & z \geq b. \end{cases}$$

Proof. As in the proof of Theorem 3.4, the result is stablished.

Theorem 3.6. For any nonnegative discrete RV Z,

$$E(Z\varphi(Z)) \geq \frac{2E^2(Z)}{b(b+1)-E(Z(Z-1))}. \quad (42)$$

The equality satisfies iff Z has the distribution function (38).

Proof. By the C-S inequality, we find that

$$(E(Z))^2 \leq E(\frac{Z}{\varphi(Z)})E(Z\varphi(Z))$$
$$= [\sum_{z=0}^{b} zF(z)]E(Z\varphi(Z))$$
$$= [\frac{b(b+1)-E(Z(Z-1))}{2}]E[Z\varphi(Z)],$$

and thus (42) is obtained.

The equality satisfies iff there exists some nonnegative constant A so that, for all $z \in \{0, \dots, b\}$,

$$\sqrt{\frac{z}{\varphi(z)}} = A\sqrt{z\varphi(z)}.$$

This implies that $\varphi(z) = \theta =$ constant, and therefore the result is obtained.

Now, we proceed to compare the bounds utilized for $E(Z\varphi(Z))$ in inequalities (41) and (42).

Assume Z follows a discrete uniform distribution with support on $\{1, \dots, b\}$. In this case, $\varphi(z) = \frac{1}{z}$,and consequently, the lower bound in (41) becomes

$$\frac{1}{E(\frac{1}{Z\varphi(Z)})} = 1.$$

Besides that, the lower bound (42) is

$$\frac{2(E(Z))^2}{b(b+1) - E(Z(Z-1))} = \frac{3(b+1)}{2(2b+1)}.$$

Accordingly, for the distribution, we deduce that the bound (41) is superior to the bound (42) for $b > 1$.

Theorem 3.7. Let Z be a nonnegative discrete RV. Then

$$E[c^{-Z}\varphi(Z)] \geq \frac{(E[c^{-Z}])^2(c-1)}{cE(c^{-Z}) - c^{-b}},$$

for constant $c \neq 1$, where equality satisfies iff Z has the distribution function given in equation (38).

Proof. By utilizing the C-S inequality, we have

$$(E[c^{-Z}])^2 \leq E[c^{-Z}\varphi(Z)]E[\frac{c^{-Z}}{\varphi(Z)}].$$

Besides that, since

$$E\left[\frac{c^{-Z}}{\varphi(Z)}\right] = \sum_{z=0}^{b} c^{-z}F(z) = \sum_{y=0}^{b}\sum_{z=y}^{b} c^{-z}P\{Z = y\}$$
$$= \sum_{y=0}^{b} \frac{c^{-y} - c^{-b-1}}{1 - c^{-1}} P\{Z = y\}$$
$$= \frac{cE[c^{-Z}] - c^{-b}}{c-1}.$$

Thus, the result is obtained. The equality holds iff there exists some nonnegative constant A so that, for all $z \in \{0, \dots, b\}$,

$$\sqrt{\frac{c^{-z}}{\varphi(z)}} = A\sqrt{c^{-z}\varphi(z)}.$$

This concludes that $\varphi(z)$ is a constant, and the result is obtained.

Results

In this work, we first introduced the past varentropy for discrete RVs. Then, we obtained bounds for the varentropy of some discrete distributions. In the following, by considering the resulting upper

bounds, the squared logarithmic expectation, we obtained an expression for the bounds in terms of the squared logarithmic difference coefficients $d_\omega(j)$. In future work, we propose obtaining similar results for continuous distributions using logarithmic and log-gamma expectations. Moreover, we evaluated an upper bound for $Var[g(X_x)]$ and derived bounds for the expected values of specific functions in reliability theory.

Acknowledgments

I would like to express my sincere gratitude to Professor Mohsen Mohammadzadeh for his invaluable suggestion, which improved this article.

References

1. Jiang J, Wang R., Pezeril M. and Wang QA. Application of varentropy as a measure of
probabilistic uncertainty for complex networks, Science

Bulletin. 2011; 56: 3677–3682.
2. Li J, Fradelizi M. and Madiman M. Information concentration for convex measures, IEEE International Symposium on Information Theory, Barcelona. 2016; 1128-1132.
3. Fradelizi M, Madiman M. and Wang L. Optimal concentration of information content for logconcave densities. In C. Houdré, D. Mason, P. Reynaud-Bouret & J. Rosin´nski (eds.), High Dimensional Probability VII. Progress in Probability, vol. 71, Cham, Springer. 2016; 45-60.
4. Gupta BB and Badve OP. GARCH and ANN-based DDoS detection and filtering in cloud computing environment, International Journal of Embedded Systems. 2017; 9: 391-400.
5. De Gregorio A and Iacus SM. On Rényi information for ergodic diffusion processes, Information Sciences. 2009; 179: 279-291.
6. Goodarzi F, Amini M and Mohtashami Borzadaran GR. Characterizations of continuous distributions through inequalities involving the expected values of selected functions, Applications of Mathematics. 2017(a); 62: 493–507.
7. Goodarzi F, Amini M and Mohtashami Borzadaran GR. On lower bounds for the variance of functions of random variables, Applications of Mathematics. 2021; 66: 767–788.
8. Sharma A and Kundu C. Varentropy of doubly truncated random variable, Probability in the Engineering and Informational Sciences. 2022; 37(3): 852–871.
9. Maadani S, Mohtashami Borzadaran GR and Rezaei Roknabadi AH. Varentropy of order statistics and some stochastic, Communication in Statistics-Theory and Methods. 2022; 51: 6447-6460.
10. Goodarzi F, Amini M and Mohtashami Borzadaran GR. Some results on upper bounds for the variance of functions of the residual life random variables, Journal of Computational and Applied Mathematics. 2017(b); 320, 30-42.
11. Goodarzi F, Amini M and Mohtashami Borzadaran GR. On upper bounds for the variance of functions of the inactivity time, Statistics and Probability Letters. 2016; 117: 62–71.
12. Buono F, Longobardi M. Varentropy of past lifetimes, Mathematical Methods of Statistics. 2022; 31: 57-73.
13. Goodarzi F, Characterizations of some discrete distributions and upper bounds on discrete residual varentropy, Journal of the Iranian Statistical Society. 2022; 21(2): 233–250.
14. Alizadeh Noughabi H and Shafaei Noughabi M. Varentropy estimators with applications in testing uniformity, Journal of Statistical Computation and Simulation. 2023; 93: 2582-2599.
15. Kontoyiannis I and Verdú. S. Optimal lossless data compression: non-asymptotics and asymp-totics, IEEE Transactions on Information Theory. 2014; 60: 777-795.
16. Nair NU and Sankaran PG. Characterizations of discrete distributions using reliability concepts in reversed time, Statistics and Probability Letters. 2013; 83: 1939–1945.
17. Cacoullos T and Papathanasiou V. Characterizations of distributions by variance bounds, Statistics and Probabability Letters. 1989; 7: 351-356.
18. Kelley CT. Solving nonlinear equations with Newtonś method, SIAM, Philadelphia, 2003.
19. Song K.-S. Rényi information, loglikelihood and an intrinsic distribution measure, Journal of Statistical Planning and Inference. 2001; 93: 51-69.
20. Cheraghchi M. Expressions for the entropy of basic discrete distribution, IEEE Transactions on Information Theory. 2019; 65: 3999-4009.
21. Gupta L. Properties of reliability functions of discrete distributions, Communication in Statistics-Theory and Methods. 2015; 44: 4114-4131.

Journal of Sciences, Islamic Republic of Iran 35(3): 243 - 256 (2024) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Evaluating Feature Selection Methods for Macro-Economic Forecasting, Applied for Iran's Macro-Economic Variables

M. Goldani*

Department of Political Science and Economics, Faculty of Literature and Humanities, Hakim Sabzevari University, Sabzevar, Islamic Republic of Iran

Received: 15 October 2024 / Revised: 3 January 2025 / Accepted: 2 February 2025

Abstract

This research examines different feature selection methods to enhance the predictive accuracy of macroeconomic forecasting models, focusing on Iran's economic indicators derived from World Bank data. Fourteen feature selection techniques were thoroughly compared, classified into Filter, Wrapper, Embedded, and Similarity-based categories. The evaluation utilized Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) metrics under a 10-fold cross-validation scheme. The findings highlight that Stepwise Selection, Tree-based approaches, and Similarity-based methods, especially those employing Hausdorff and Euclidean distances, consistently outperformed others with average MAE values of 32.03 for Stepwise Selection and 62.69 for Hausdorff Distance. Conversely, Recursive Feature Elimination and Variance Thresholding exhibited weaker results, yielding significantly higher average MAE scores. Similarity-based approaches achieved an average rank of 9.125 across datasets, demonstrating their robustness in managing high-dimensional macroeconomic data. These outcomes underscore the value of integrating similarity measures with traditional feature selection techniques to improve the efficiency and reliability of predictive models, offering meaningful insights for researchers and policymakers in economic forecasting.

Keywords: Feature Selection; Predictive Accuracy; World Bank Indicators; Macroeconomic Analysis; Similarity Methods.

Introduction

The primary challenge of working with high-dimensional data lies in the exponential growth in complexity and sparsity that such data introduces. Additionally, the costs associated with storage and transmission increase, visualization becomes more challenging, and redundant or irrelevant features often complicate analysis (1). To address these challenges, dimensionality reduction techniques, feature selection, regularization methods, and meticulous data preprocessing are essential. These approaches help to extract valuable insights while mitigating the negative impacts of high dimensionality on data analysis and machine learning tasks. Feature selection is a key technique in dimensionality reduction, focusing on carefully identifying a relevant subset of features (variables or predictors) for model development. It plays a critical role in the data preprocessing workflow. Among various dimensionality reduction strategies, feature

* Corresponding Author: Tel: 05144410104; Email: m.goldani@hsu.ac.ir

Table 1. Summary of the specific objectives of the research

Objective	Details
Exploring Similarity-Based Feature Selection	Investigate using distance measures (e.g., Hausdorff, Euclidean, Dynamic Time Warping) as feature selection tools.
Benchmarking Against Conventional Methods	Compare similarity-based methods with Filter, Wrapper, and Embedded approaches using RMSE and MAE.
Assessing Practical Implications	Evaluate computational simplicity, robustness, and real-world applicability in economic forecasting.
Demonstrating Relevance with Case Studies	Use Iran's macroeconomic indicators (1990–2022) to validate findings and provide actionable insights.

selection stands out as a significant approach that retains only relevant features and eliminates redundant or irrelevant ones (2). Feature selection is vital in machine learning and data analysis, particularly when handling high-dimensional datasets. Identifying and selecting the most important features enhances model performance by improving predictive accuracy, reducing overfitting, and lowering computational costs. In the context of target variables, the feature selection process significantly contributes to achieving more accurate predictions.

Through systematically identifying and preserving relevant features while removing irrelevant or redundant ones, feature selection boosts a model's ability to capture underlying patterns and relationships within the data. This results in improved predictive accuracy, typically reflected in lower RMSE and MAE values. For instance, accurately forecasting GDP growth or inflation rates requires isolating key economic indicators such as broad money supply, government expenditure, and foreign direct investment. Similarly, understanding the drivers of unemployment or manufacturing growth necessitates focusing on the most impactful predictors. Effective feature selection not only enhances accuracy but also improves model interpretability and computational efficiency, aiding in better economic analysis and policymaking. Reducing model complexity also helps prevent overfitting, ensuring forecasts remain robust and reliable when applied to new data.

Feature selection (FS) is the process of identifying the most relevant and effective subsets of features to enhance the robustness of predictive models. This step is performed during the preprocessing phase of machine learning workflows. Before any training or testing, choosingthe most pertinent features based on the target variable is essential. While many FS techniques have been proposed in the literature, some methods, such as time series similarity methods, can also identify the most relevant features. A review of existing literature reveals that no studies have yet applied time series similarity methods specifically for feature selection. However, there are similarities between these two approaches that make time series similarity a promising alternative. Time series similarity measures the distance between two time series, which forms the foundation for clustering and classification tasks. A smaller distance between a feature and the target variable indicates that the feature is more relevant and should be included in the model. Thus, the goal of this research is to explore whether time series similarity methods can be as effective as traditional feature selection methods for identifying relevant feature subsets. The significance of this inquiry lies in the simplicity of the preprocessing step is just as important as the effectiveness of the methods employed, potentially saving both time and resources.

The overarching goal of this study is to evaluate the effectiveness of similarity-based methods as feature selection tools for high-dimensional macroeconomic forecasting. Table 1 summarizes the specific objectives of the research:

By addressing these objectives, the study contributes to advancing feature selection methodologies and provides practical recommendations for integrating similarity-based approaches in macroeconomic forecasting and other domains.

In the following sections, the methodology for integrating time series similarity measures into the feature selection framework is discussed (Section 2), the empirical results of the study are presented (Section 3), and the implications of the findings are analyzed in the discussion and conclusion (Section 4).

Literature review

Feature selection is essential for improving machine learning models accuracy, interpretability, and computational performance. By isolating the most significant features and eliminating those that are redundant or irrelevant, it addresses many of the challenges associated with high-dimensional datasets.

While feature selection has been extensively studied, the direct use of similarity measures as an independent method has not received much attention. Nevertheless, various studies have leveraged similarity measures indirectly to enhance feature selection techniques, as outlined below.

Similarity-based methods have shown potential, particularly in unsupervised feature selection. For example, Zhu et al. (3) proposed the Feature Selection-based Feature Clustering (FSFC) algorithm, which employs clustering driven by similarity measures to group and select features effectively. Similarly, Mitra (4) introduced an algorithm for unsupervised feature selection in large, high-dimensional datasets. This method evaluates features redundancy using similarity metrics, achieving greater efficiency and scalability.

Building on these ideas, Shi et al. (5) developed the Adaptive-Similarity-based Multi-modality Feature Selection (ASMFS) approach. This technique constructs a similarity matrix to capture inherent relationships across different modalities in high-dimensional data. The method demonstrated superior performance in tasks such as Alzheimer's disease classification, showcasing the value of similarity-based strategies in feature selection.

Recent research has refined similarity-based approaches to make them more robust and adaptable. Mehri et al. (6) employed similarity measures to identify and eliminate redundant features by examining their resemblance to others. Shen, Chen, and Garibaldi (7) proposed a meta-learning framework that integrates fuzzy similarity measures for recommending optimal feature selection techniques tailored to diverse datasets. Their approach automates feature selection, enhancing adaptability across dataset characteristics.

Goldani and Asadi (8) explored the application of similarity measures in financial forecasting, utilizing methods such as Haus Dorff distance and variance thresholds. These measures effectively selected predictive features, particularly in scenarios involving fluctuating data volumes. Similarly, Mathisen et al. (9) enhanced automated similarity measures for clustering, case-based reasoning, and one-shot learning, demonstrating their adaptability and utility in diverse applications.

Matrix factorization techniques have also leveraged similarity measures for feature selection. QI et al. (10) introduced the Regularized Matrix Factorization Feature Selection (RMFFS) method, which employs matrix factorization to capture feature correlations and applies a combination of l1 and l2 norms to ensure sparsity in the feature weight matrix. Du et al. (11) proposed the Robust Unsupervised Feature Selection via Matrix Factorization (RUFSM) method, which decomposes the data matrix into latent cluster centers and sparse representations. This approach achieves high-accuracy feature selection by identifying orthogonal cluster centers.

Hu et al. (12) extended this line of research with the Graph Self-Representation Sparse Feature Selection (GSR-SFS) method. Integrating a subspace-learning model into a sparse feature-level self-representation approach, improves both the interpretability and stability of the selected features.

Feature selection methods have found significant applications in medical and dynamic datasets. Remeseiro and Bolon-Canedo (2) reviewed feature selection techniques in medical imaging, biomedical signal processing, and DNA microarray data, highlighting their utility in solving domain-specific challenges. Venkatesh and Anuradha (13) addressed the limitations of traditional feature selection methods for dynamic, noisy datasets generated in IoT and web-based applications. Their work emphasized the need for scalable and robust methods to handle the evolving nature of such data.

The consensus among researchers, as highlighted by Guyon and Elisseeff (14), is that feature selection is crucial for improving the performance and interpretability of machine learning models. The choice of the feature selection method should be tailored to the specific problem and dataset, as there is no universal solution. Proper evaluation and validation are necessary to ensure the effectiveness of any feature selection technique. Jović et al. (15) investigated the calculation methods of standard filter, wrapper, and embedded methods. The result revealed that filters based on information theory and wrappers based on greedy stepwise approaches offer the best results.

The existing body of work highlights the potential of similarity-based methods to address challenges such as feature redundancy and relevance in high-dimensional data. While traditional feature selection methods such as Filter, Wrapper, and Embedded approaches have succeeded, integrating similarity measures directly into feature selection frameworks offers a promising alternative. However, their application remains underexplored in macroeconomic forecasting, which has motivated the current study to evaluate their feasibility and effectiveness in this context. This study bridges this gap by investigating the feasibility and effectiveness of using time series similarity methods as feature selection techniques. By systematically comparing these methods with established feature selection techniques, the research aims to evaluate their performance in identifying relevant subsets of features while ensuring computational simplicity and robustness. The findings have implications not only for improving the preprocessing of high-dimensional datasets but also for

advancing methodologies in domains such as economic forecasting, healthcare, and beyond.

Materials and Methods

This section outlines the methodology employed in this research, consisting of four key steps as depicted in Figure 1.

Dataset

This paper aims to compare the predictive performance of datasets selected using feature selection techniques and time series similarity methods. The data set employed for this purpose is derived from the World Bank Development Indicators. To validate and assess the effectiveness of the dataset chosen through these methods, various target variables were selected, as summarized in the Table 2. These variables represent "Macroeconomic Indicators" for Iran, with data sourced from the World Bank website for 1990–2022.

Preprocessing data

As an initial step in the data preprocessing process, variables with a high proportion of missing data—specifically, those with more than 80% of their values absent—are systematically removed from the dataset to ensure the reliability and integrity of subsequent analyses. This step helps eliminate variables that otherwise provide insufficient information for meaningful insights. For the remaining variables, which have a missing data rate of less than 80%, the gaps in the dataset are addressed through the application of the K-Nearest Neighbors (KNN) imputation method. This technique leverages the patterns and relationships between existing data points to estimate and fill in

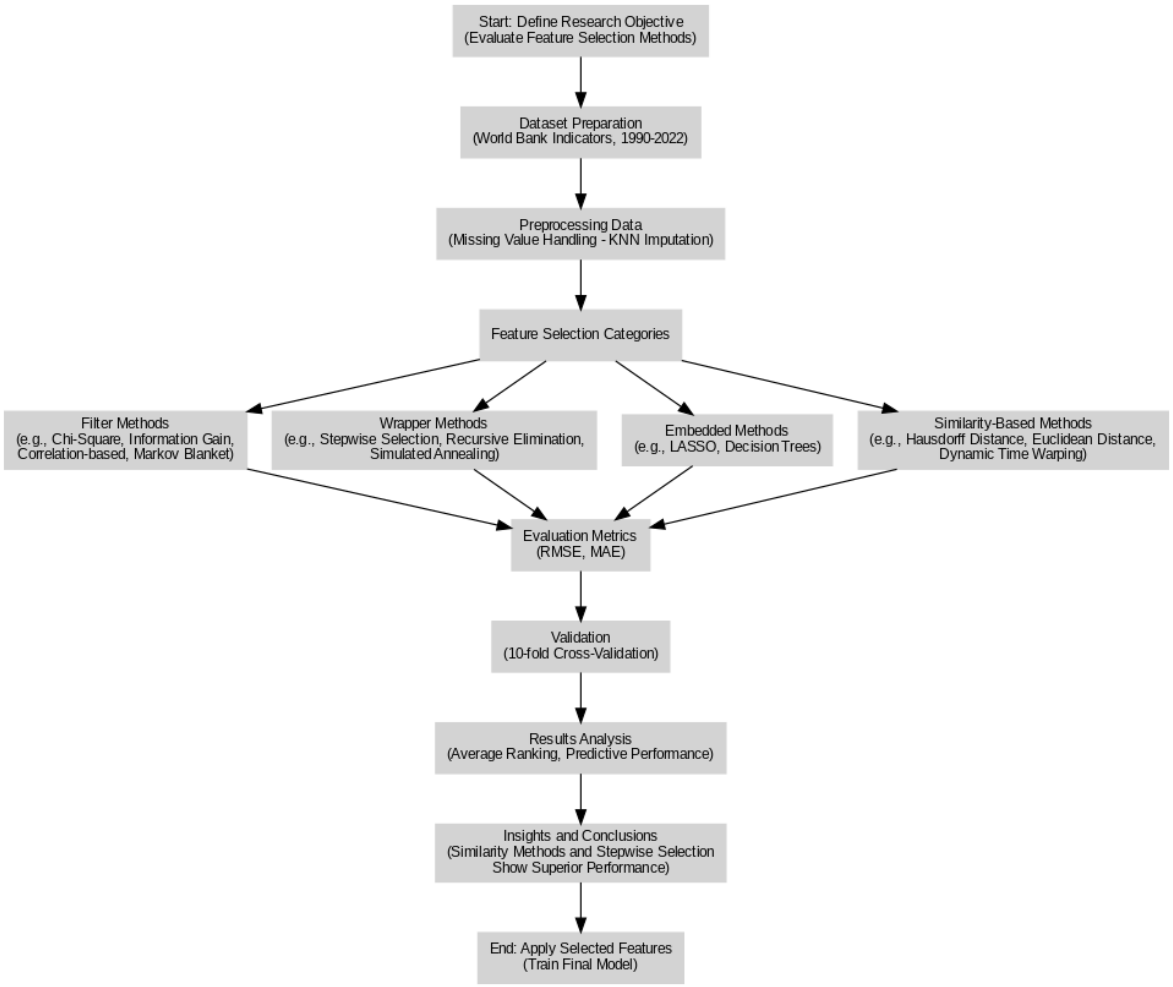


Figure 1. The complete methodology

Table 2. The list of target variables

Variable	Description
Adjusted Savings: Consumption of Fixed Capital	Annual adjusted savings considering fixed capital usage.
Broad Money	Total money supply in the economy.
Food Production Index (2014–2016 = 100)	Measure of food production, base year 2014–2016.
Foreign Direct Investment (Net Inflows as % of GDP)	Net inflows of FDI as a percentage of GDP.
GDP Growth	Annual growth rate of GDP.
General Government Final Consumption Expenditure (% of GDP)	Government consumption as a percentage of GDP.
GNI	Gross National Income.
Gross Domestic Income	Total income generated domestically.
Gross Domestic Saving	National saving as a percentage of GDP.
Gross National Expenditure (% of GDP)	Total expenditure as a percentage of GDP.
Gross Value Added at Basic Prices	Value addition by all sectors at basic prices.
Households and NPISHs Final Consumption Expenditure Per Capita (Constant 2015 US$)	Per capita household expenditure in constant dollars.
Imports of Goods and Services (Constant 2015 US$)	Value of imports adjusted to constant 2015 US$.
Manufacturing Value Added (Annual % Growth)	Annual growth in manufacturing output.
Official Exchange Rate (LCU per US$, Period Average)	Average local currency exchange rate per US dollar.
Stocks Traded (Total Value as % of GDP)	Value of traded stocks as a percentage of GDP.
Total Debt Service (% of Exports of Goods, Services, and Primary Income)	Debt repayment as a percentage of exports.
Unemployment (Total % of the Labor Force, Modeled ILO Estimate)	Total unemployment rate as estimated by ILO.
Wholesale Price Index (2010 = 100)	Index measuring wholesale price levels (base 2010).
Consumer Price Inflation	Annual inflation based on consumer prices.

missing values, thereby preserving the completeness of the dataset while maintaining its statistical validity (16). This approach ensures that the data set is robust and suitable for further analysis.

Conventional feature selection methods

Feature selection (FS) techniques are employed to determine and preserve the most significant and insightful features of the data, ensuring the construction of precise predictive models. The dataset includes many features, leading to the presence of noise, irrelevant details, and redundant information. Hence, this increases the computational time and error rate of the learning algorithm (17). Three main categories of feature selection methods exist: filter, wrapper, and embedded. A brief description of each selection method is given in Table 3.

They become particularly valuable in complex scenarios where neither filter, wrapper, nor can embedded methods alone achieve the desired outcomes.

The proposed approach

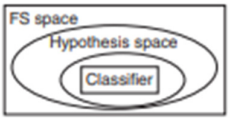
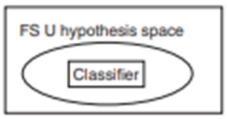
The proposed method falls under Filter techniques, which evaluate feature importance based on their correlation with the target variable. Figure 2 illustrates the framework of the suggested methodology, emphasizing its four main stages.

At the heart of this approach lies the application of similarity measures. This study examines feature selection (FS) by utilizing various distance metrics, including Euclidean Distance, Dynamic Time Warping (DTW), Edit Distance on Real Sequences (EDR), Longest Common Subsequence (LCSS), and Edit Distance with Real Penalty (ERP). These metrics are crucial for assessing the similarity between time series, a fundamental task in the clustering and classifying of temporal data. The primary goal is to determine the distance between two time series, which is vital for analyzing temporal patterns and trends.

In earlier applications, time series similarity was a direct statistical inference tool to uncover relationships between time series originating from different datasets (19). However, with the exponential growth of data collection in recent years, time series data has become increasingly prevalent, leading to a surge in analytical tasks such as regression, classification, clustering, and segmentation. These tasks often hinge on selecting a suitable distance metric to effectively quantify the degree of similarity between time series.

Given the importance of similarity measures, this study explores various methods to determine the distance between time series. These methods are broadly classified into three main categories: stepwise measures,

Table 3. Conventional feature selection methods

	**Univariate**		
	- Fast - Scalable	- Ignores feature dependencies - Ignores interaction with the classifier	χ^2 (Chi-square test) Euclidean distance
	- Independent of the classifier		i-test
Filter			Information gain Gain ratio
	Multivariate - Models feature dependencies	- Slower than univariate techniques	Correlation-based feature selection (CFS)
	- Independent of the classifier	- Less scalable than univariate techniques	Markov blanket filter (MBF)
	- Better computational complexity than wrapper methods	- Ignores interaction with the classifier	Fast correlation-based feature selection (FCBF)
	Deterministic - Simple	- Risk of overfitting	Sequential forward selection (SFS)
	- Interacts with the classifier	- More prone than randomized algorithms to	Sequential backward elimination (SBE)
Wrapper	- Models feature dependencies	getting stuck in a local optimum (greedy search)	Recursive Feature Elimination
	- Less computationally intensive than randomized methods	- Classifier dependent selection	
	Randomized - Less prone to local optima	- Computationally intensive	Simulated annealing
	- Interacts with the classifier	- Classifier dependent selection	Randomized hill climbing
	- Models feature dependencies	- Higher risk of overfitting than deterministic methods	Genetic algorithms
			Estimation of distribution algorithms
	- Interacts with the classifier	- Classifier dependent selection	Decision trees
	- Better computational complexity than wrapper methods		LASSO
Embedded	- Models feature dependencies		Feature selection using the weight vector of SVM

which align time series elements sequentially; distribution-based measures, which focus on statistical properties; and geometric methods, which emphasize spatial relationships and patterns. Understanding and leveraging these approaches is essential for advancing time series analysis and enhancing its applications across diverse fields.

Stepwise Metrics

These metrics compare time-series samples one by one based on their time indices (20). A significant limitation of these methods is the requirement for identical sample sizes in the time series. The most notable stepwise metrics are Euclidean Distance and Correlation Coefficient, which are detailed below.

- The Euclidean Distance is the simplest measure for comparing time series. It calculates the shortest distance between two points in Euclidean space using the Pythagorean theorem. The Euclidean Distance between two time series x and y of length n is defined as:

$$Deuc = (\sum_{i=1}^{n}(x_i - y_i)^{1/2} \quad (1)$$

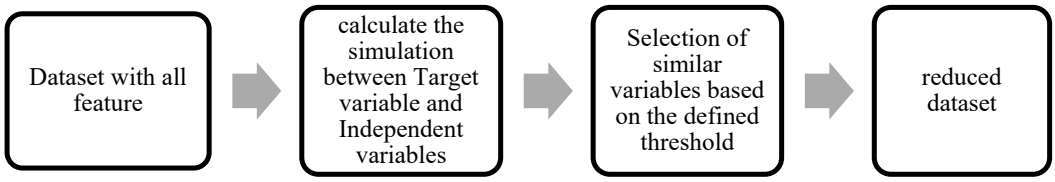


Figure 2. The framework of the proposed feature selection

This distance is widely used due to its simplicity and ease of understanding. However, a key limitation of Euclidean Distance is its sensitivity to time-axis transformations, such as scaling and shifting (21). Moreover, it cannot compare time series with different sample sizes. As it relies on point-to-point mapping, it is highly sensitive to noise and temporal misalignments, thus making it unsuitable for handling local shifts in time.

A straightforward extension of Euclidean Distance is to calculate the similarity using extracted features rather than raw time-series data.

- Pearson Correlation Coefficient is a widely used metric for assessing the linear relationship between two time series. It is defined as:

$$corr(x,y) = \frac{E(XY)-E(X)E(Y)}{std(X)std(Y)} \quad (2)$$

The Pearson Correlation Coefficient ranges between -1 and 1, where 1 indicates a perfect positive correlation, and -1 reveals a perfect negative correlation. However, it cannot distinguish between dependent and independent variables or capture non-linear relationships.

Elastic metrics

These metrics adjust the time axis by stretching or compressing it to minimize the effect of local variations. These methods are particularly effective for handling non-linear distortions on time. The most notable elastic methods include Dynamic Time Warping (DTW), Longest Common Subsequence (LCSS), and others.

- Dynamic Time Warping (DTW) is an algorithm for measuring similarity between time series that may vary in speed or timing. Unlike Euclidean Distance, DTW aligns sequences non-linearly by stretching or compressing the time axis to find the optimal alignment. The cumulative distance is calculated as:

$$DISTMATRIX = \begin{bmatrix} d(x_1,y_1) & d(x_1,y_2) & \dots & d(x_1,y_m) \\ d(x_2,y_1) & d(x_2,y_2) & \dots & d(x_2,y_m) \\ d(x_n,y_1) & d(x_n,y_2) & \dots & d(x_n,y_m) \end{bmatrix} \quad (3)$$

$$\begin{cases} r(i,j) = d(i,j) + \min\{r(i-1,j), r(i,j-1), r(i-1,j-1)\} \\ DTW(x,y) = \min\{r(n,m)\} \end{cases} \quad (4)$$

DTW allows comparisons between time series of different lengths and identifies similar shapes, even if they are out of phase. However, it is computationally intensive, making it less practical for large datasets.

- Longest Common Subsequence (LCSS) focuses on the longest matching subsequences between two time series while ignoring noise and distortions. For two sequences S_x and S_y of lengths n and m, the similarity is defined as:

$$M(i,j) = \begin{cases} 0 \quad ; \ i = 0 \ or \ j = 0 \\ 1 + M(i-1,j-1) \ ; \ x_i = y_j \ , \ i \geq 1 \ or \ j \geq 1 \\ Max\begin{cases} M(i-1,j) \\ M(i,j-1) \end{cases} ; x_i \neq y_j \ , \ i \geq 1 \ or \ j \geq 1 \end{cases} \quad (5)$$

Where M (n,m) is calculated recursively:

$$M(i,j) = \begin{cases} 0 \quad ; \ i = 0 \ or \ j = 0 \\ 1 + M(i-1,j-1) \ ; \ (x_i - y_j) \leq \varepsilon \ , \ i \geq 1 \ or \ j \geq 1 \\ Max\begin{cases} M(i-1,j) \\ M(i,j-1) \end{cases} ; \ (x_i - y_j) > \varepsilon \ , \ i \geq 1 \ or \ j \geq 1 \end{cases} \quad (6)$$

LCSS is robust to noise and suitable for comparing time series with different lengths. However, it heavily depends on the similarity threshold, which impacts its accuracy.

- The edit distance algorithm counts the number of insertion, deletion, and substitution operations required to transform one string into another. It can be applied to time series, where points X and Y match if their absolute distance is less than ε (22). Given two sequences Y, and X, of lengths n and mmm, respectively, the Edit Distance on Real sequence (EDR) between X and Y refers to the number of insertions, deletions, or substitutions required to transform X into Y. It is defined as follows:

$$EDR(X,Y) = \begin{cases} n \quad if \ m = 0 \\ m \quad if \ n = 0 \\ min\begin{cases} EDR(Rest(X), Rest(Y)) + subcost, \\ EDR(Rest(X), Y) + 1 \ , EDR(X, Rest(y)) + 1 \end{cases} \end{cases} \quad (7)$$

o ERP, as with the EDR method, is based on Edit Distance (ED) for measuring the similarity of time-series data (23). ERP, accompanied by the L1-norm and Edit Distance, supports local time shifts and is a metric, meaning it satisfies the triangular inequality. Non-metric distance functions complicate problems as violating the triangular inequality renders most indexing structures infeasible. The primary reason why EDR does not satisfy the triangular inequality is that when a gap needs to be added, it repeats the previous element. In contrast, ERP does not face this issue since it uses the L1-norm between two non-gap elements and is designed in such a way that it applies an actual penalty between two non-gap elements. However, it employs a fixed value for calculating the distance for gaps (23). When calculating ERP for two time series S_x and S_y with lengths n and mmm, they are aligned to the same length by adding certain symbols (referred to as gaps). Then, each element in one time series is matched with a gap or an element in another. Finally, the ERP distance between the two-time series S_x and S_y is defined recursively.

$$d_{erp} = \begin{cases} \sum_{i=1}^{m}|x_i - g| & if\ n = 0 \\ \sum_{j=1}^{n}|y_j - g| & if\ m = 0 \\ min\begin{Bmatrix} d_{erp}(Rest(x), Rest(y) + |x_1 - y_1|), \\ d_{erp}(Rest(x), y + |x_1 - g|) \\ d_{erp}(x, Rest(y) + |y_1 - g|) \end{Bmatrix} \end{cases} \quad (8)$$

o Time Warped Edit Distance (TWED) combines the strengths of DTW and edit distance algorithms by allowing elastic matching with additional constraints. The similarity is measured as the minimum sequence of edit operations required to align two time series.

Geometric distances

Geometric distances focus on the spatial characteristics of trajectories, particularly their shapes. Examples include Hausdorff Distance, Discrete Frechet Distance, and SSPD (Symmetric Segment Path Distance).

o The Hausdorff Distance measures the maximum mismatch between two trajectories, defined as:

$$Haus(X, Y) = Max\{\sup\inf\|xy\|_2, \sup\inf\|xy\|_2\} \quad (9)$$
$$x \in X\ y \in Y \quad x \in X\ y \in Y$$

o Frechet Distance measures the similarity between curves by calculating the minimal "leash length" required to connect a dog and its owner walking along two separate paths. It is mathematically defined as:

$$D_{Frechet}(T^1, T^2) = \min\{\max\|w_k\|_2\} \quad (10)$$
$$w\ k \in (0 \ldots |w|)$$

o SSPD shape-based distances such as Hausdorff and Frechet can align with corresponding paths but can not be compared as a unified entity. SSPD is a shape-based distance metric that does not consider the time index of the path. This metric calculates the point-to-segment distance for all samples of the reference path and all segments of the other path then report the average of the obtained distances for the path sample as the SSPD distance (24).

SSPD is defined as follows:

$$\begin{aligned} D_{SPD}(T^1, T^2) &= \frac{1}{n_1}\sum_{i_1=1}^{n_1} D_{pt}(p_i^1, T^2) \\ (p_i^1, T^2) &= min_{i_2 \in (0 \ldots . n_2 - 1)} D_{ps}(p_{i_1}^1, s_{i_2}^2) \\ D_{PT}(P_1^2, T^1) &= \min D_{PS}(P_1^2, S_{i_1}^1) \\ & \quad i_1 \in (0 \ldots . n_1 - 1) \end{aligned} \quad (11)$$

This distance is not symmetric. By considering the average of these distances, SSPD is defined as follows:

$$D_{SSPD}(T^1, T^2) = \frac{D_{SPD}(T^1, T^2) + D_{SPD}(T^2, T^1)}{2} \quad (12)$$

Validation methods

Stationarity is a key principle in time series analysis, defined as the condition where the statistical attributes of a time series, such as its mean, variance, and autocorrelation, remain unchanged over time (25). A stationary time series is essential for reliable analysis and modeling. In the subsequent phase of our methodology, statistical tests were conducted to evaluate significant variations among the reduced datasets.

To perform predictive analysis, a Linear Regression model was selected due to its straightforward nature and ease of interpretation. Nevertheless, alternative regression models may be applied based on the specific requirements of the study. To enhance the reliability of the model evaluation and mitigate the risk of overfitting, a 10-fold cross-validation technique was employed. This method involves splitting the dataset into ten roughly equal parts, with each subset alternately used for training and testing during the evaluation.

Model performance was measured using two key metrics: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). RMSE captures the deviation between predicted and observed values, whereas MAE quantifies the average error magnitude in predictions. The evaluation was carried out across 10 iterations, generating unique RMSE and MAE scores for each run. This iterative approach ensured the robustness and

consistency of the results, providing a comprehensive validation of the methodology.

Results and Discussion

In this study, we present the results of predicting performance across 14 datasets, each selected using a different feature selection (FS) technique. These techniques include seven filter methods, five wrapper methods, three embedded methods, and four similarity-based methods. The similarity methods as FS techniques are also evaluated within this framework. The chosen methods were selected for their widespread recognition in literature, allowing for a clear comparison. To assess predictive accuracy, we use two evaluation metrics: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), applied to the performance of a Linear Regression model. To evaluate the efficiency of each dataset selected by the FS methods, we implemented the techniques on the World Bank dataset, which includes

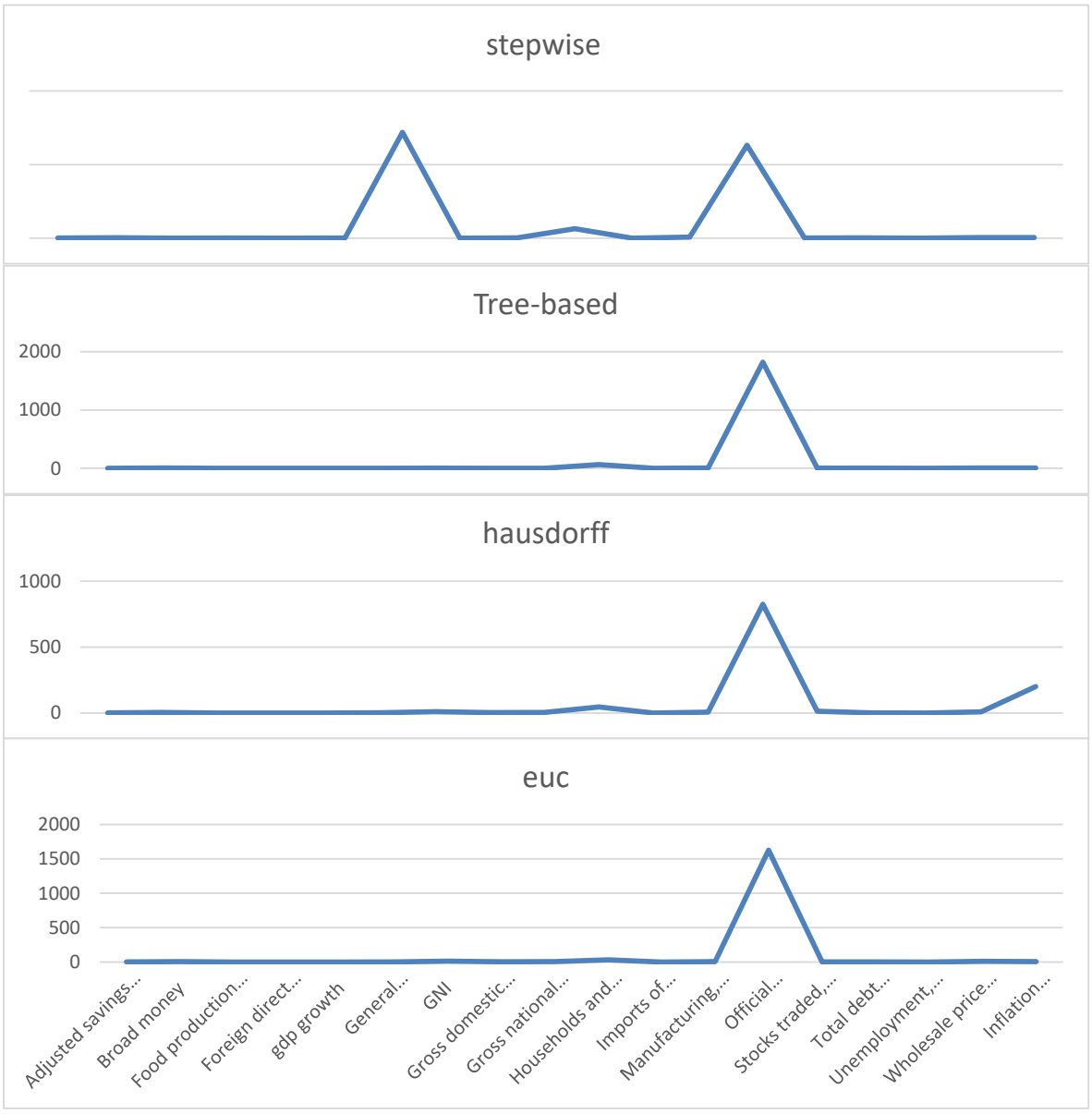


Figure 3. The top four feature selection models based on 14 datasets chosen by Feature selection techniques

Table 4. Average Mean Absolute Error (MAE) of datasets

Category	Methods	Average
Wrappers	stepwise	32/0299
similarity	frechet	51/6163
similarity	hausdorff	62/68829
similarity	sspd	91/70364
similarity	epr	91/88632
similarity	dtw	91/93176
similarity	euc	95/02939
Embedded	Tree-based	106/3909
Wrappers	recursive	270/5572
similarity	lcsso	292/8808
similarity	edr	298/4402
Filters	MI_Score	963/5397
Filters	inf	1683/06
similarity	Sparse	3/98E+08
Wrappers	forward	6/4E+08
Wrappers	simulated_annealing	8/13E+08
Filters	fisher	1/83E+09
Embedded	lasso	3/06E+12
Filters	chi	4/83E+13
Filters	corrolation	4/83E+13
Filters	data_dispersion	8/16E+13
Filters	var	6/41E+14
Wrappers	backward	6/47E+14

various target variables. In total, 20 different datasets were used, and FS methods were employed to identify the best feature subsets from each.

Figure 3 illustrates the results of a 10-fold cross-validation evaluation for each FS method. The datasets selected by these four methods consistently exhibited the lowest RMSE and MAE, indicating superior predictive accuracy.

Table 4 presents the average MAE values for datasets processed using various feature selection (FS) methods. The Mean Absolute Error (MAE) averaged across 20 datasets for each target variable.Those derived using the stepwise feature selection approach demonstrated superior predictive accuracy among the subsets generated. These subsets consistently exhibited the smallest MAE values compared to others. Following closely were the subsets identified through similarity-based techniques, which also achieved notably low average MAE scores, underscoring their effectiveness in prediction tasks.

Figure 4 indicates the average ranking of MAE selected based on FS methods. The ranking of each feature selection (FS) method was determined based on its ability to select the best subset of datasets with the lowest Mean Absolute Error (MAE). To provide a comprehensive analysis, the rank of each of the 20 datasets across all FS methods was averaged. According to the results, the best predictive accuracy methods were Stepwise Selection, Tree-based methods, Hausdorff, Euclidean (Euc), and MI_Score. In contrast, Recursive Feature Elimination with Cross-Validation (RFECV) and Variance Thresholding exhibited the poorest performance.

The average ranking across the feature selection categories (Figure 5) indicates that, on average, similarity-based methods outperformed the other approaches. Specifically, similarity methods achieved an average rank of 9.125, highlighting their superior performance in selecting the most relevant feature subsets compared to other methods.

The results underscore the potential of similarity-based methods as viable alternatives to traditional feature selection techniques, with implications for a wide range of applications, particularly macroeconomic forecasting.

Effectiveness of Similarity-Based Approaches

The strong performance of similarity-based methods, particularly Frechet and Hausdorff distances, demonstrates their ability to identify features that exhibit high relevance to target variables. These methods leverage the inherent structure of time series data, effectively capturing relationships that might be overlooked by traditional approaches. For instance, the Frechet Distance, which accounts for the shape and continuity of data trajectories, excels in handling time series with local distortions, while the Hausdorff Distance, which measures the greatest distance between points of two datasets, is robust against outliers and

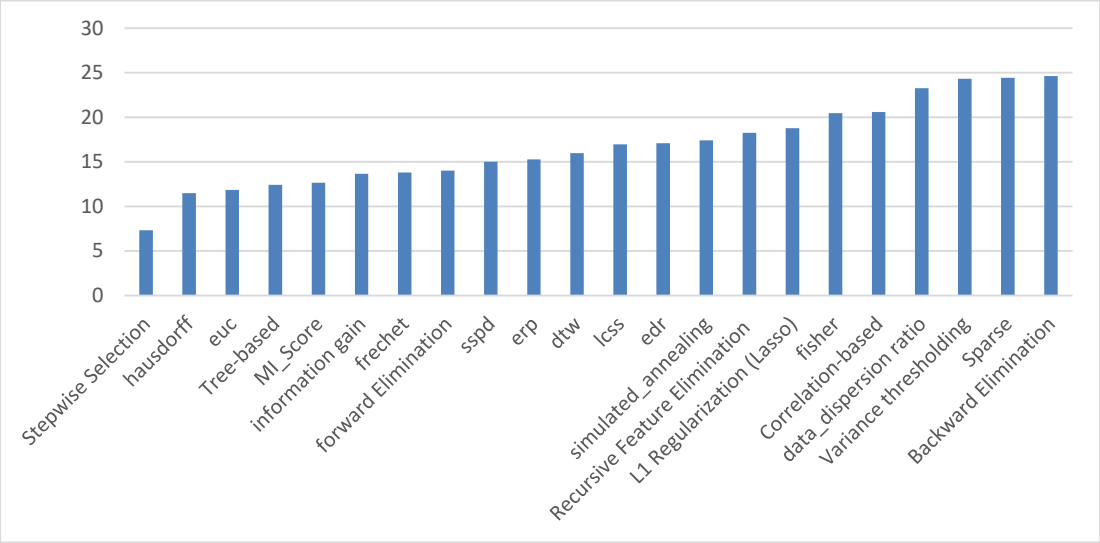


Figure 4. The average ranking of MAE selected based on FS methods

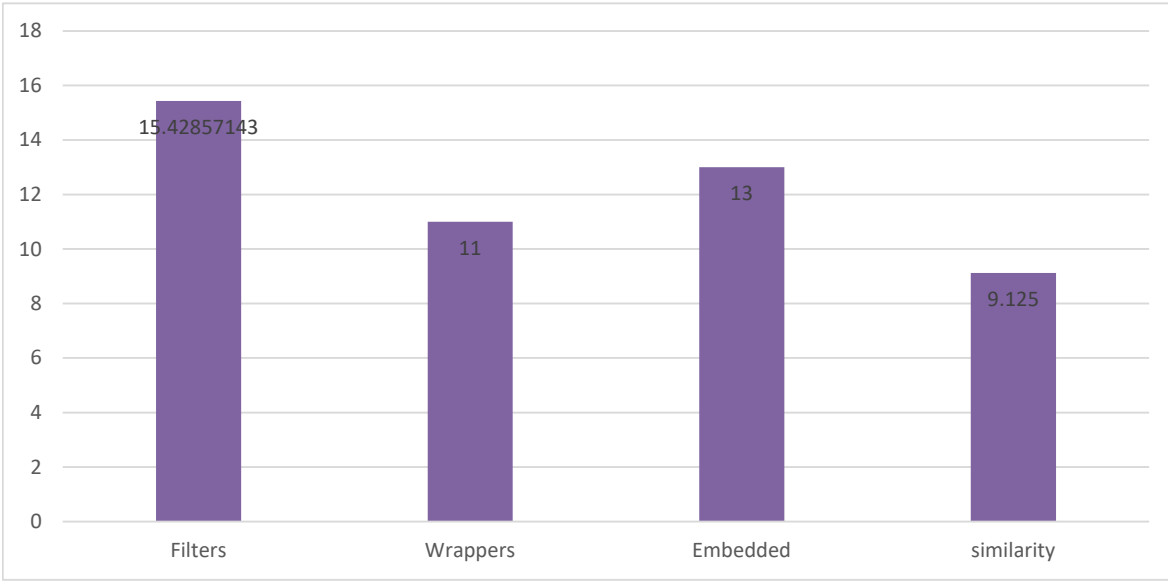


Figure 5. The ranking of the category of feature selection methods

noise.

This capability aligns with clustering and classification literature findings, where similarity measures are frequently employed to quantify relationships between data points. By applying these measures to feature selection, this study extends their utility into a new domain, validating their effectiveness in identifying subsets of features that enhance model performance. Further, their simplicity and computational efficiency make similarity-based methods suitable for real-world scenarios where quick and accurate analysis is critical.

Comparison with Traditional Methods

Traditional feature selection methods, such as Stepwise Selection and Tree-based approaches, remain benchmarks in the field due to their consistent performance and well-established methodologies.

Stepwise Selection, in particular, excels in identifying key features through iterative inclusion or exclusion, making it a preferred choice for many predictive modeling tasks. Similarly, Tree-based methods, such as Random Forests, offer an embedded mechanism for ranking features by their importance, balancing accuracy and interpretability.

However, similarity-based methods emerge as strong contenders, offering a computationally efficient alternative especially advantageous in high-dimensional scenarios. Unlike traditional methods that often rely on iterative testing or classifier-specific criteria, similarity-based approaches operate independently of classifiers, enabling faster preprocessing and reducing the risk of overfitting. This makes them particularly appealing for datasets with numerous variables, where computational resources and time constraints are significant considerations.

Implications for Macroeconomic Forecasting

Macroeconomic forecasting heavily relies on accurate predictions of key indicators, such as GDP growth, inflation rates, and unemployment levels. The use of similarity-based methods in this context provides several advantages:

✓ Simplification of Preprocessing: By directly measuring the relationship between features and target variables, similarity-based methods eliminate redundant preprocessing steps. This simplifies the pipeline and lowers the risk of introducing errors during data preparation.

✓ Enhanced Interpretability: The straightforward nature of similarity measures, such as distances or correlations, allows for easier interpretation of results. Policymakers and economists can gain clearer insights into which features drive predictions, thus facilitating more informed decision-making.

✓ Robust Forecasting Tools: By focusing on the most relevant features and minimizing noise, these methods contribute to developing robust and reliable forecasting models. This is particularly critical for policymaking, where accurate predictions can guide interventions and resource allocation.

Conclusion

In this study, we investigated which feature selection (FS) and similarity methods most effectively enhance the predictive performance of models for various macroeconomic variables. The analyzed indicators included a diverse range of metrics, such as adjusted savings (consumption of fixed capital), broad money, the food production index, imports of goods and services (constant 2015 US$), manufacturing value-added (annual % growth), official exchange rate (LCU per US$), stocks traded (total value as a % of GDP), total debt service (% of exports), unemployment (% of the total labor force, ILO estimate), the wholesale price index (2010 = 100), and consumer price inflation. To achieve this, we evaluated 23 different FS and similarity methods to identify the most effective techniques for selecting features that provide accurate predictions of these macroeconomic indicators.

Time series similarity algorithms, though rarely utilized as standalone feature selection methods, were a key focus of this research. By comparing these algorithms against traditional FS approaches, we aimed to assess their potential in identifying relevant features. Each FS and similarity method was applied to the datasets, and their performance was evaluated using both MAE and RMSE metrics. The current findings are hence in agreement with the studies of Zhu et al. and Mitra, who applied the methods of similarity measures for feature grouping and selection to increase clustering performance. Additionally, robustness from similarity metrics obtained herein further supports conclusions from Mehri et al. and Goldani and Asadi, demonstrating their viability in high-dimensional and financial forecasting setups. This work extends these approaches toward macroeconomic forecasting, hence addressing an important lacuna in the related literature. Besides, the traditional feature selection methods, such as stepwise selection and tree-based methods, were confirmed to be reliable benchmarks, which agrees with the results obtained by Jović et al. However, the similarity-based methods were their strong competitors, providing equal or higher predictive accuracy with computational simplicity. Unlike other methods, such as Recursive Feature Elimination and Variance Thresholding, which did not perform well in our analysis, results consistent with the critiques of Guyon similarity-based approaches provided a more robust alternative for high-dimensional datasets. Findings revealed that methods such as Stepwise Selection paired with Tree-based techniques, Hausdorff distance, Euclidean distance, and Mutual Information Score consistently outperformed other approaches, demonstrating higher predictive accuracy. Conversely, methods like Recursive Feature Elimination with Cross-Validation and Variance Thresholding showed comparatively weaker results, suggesting limited utility in this context. These results highlight the potential of similarity-based algorithms as effective tools for feature selection in macroeconomic forecasting.

By systematically comparing these methods with established feature selection techniques across 20 datasets of macroeconomic indicators, the key findings were obtained as follows:

Performance of Similarity-Based Methods:

- Similarity-based methods, particularly Frechet and Hausdorff distances, demonstrated strong performance in identifying relevant features, with competitive Mean Absolute Error (MAE) values compared to traditional techniques.
- The computational efficiency and robustness of similarity-based methods make them suitable for high-dimensional datasets, offering an alternative to Filter, Wrapper, and Embedded methods.

Advancing Feature Selection:

- Traditional approaches such as Stepwise Selection and Tree-based methods remain benchmarks thanks to their high accuracy and established methodologies. However, similarity-based methods provide a complementary approach, particularly in applications requiring computational simplicity and adaptability.

Macroeconomic Implications:

- The adoption of similarity-based feature selection can improve forecasting accuracy for critical economic indicators such as GDP growth, inflation, and unemployment. These tools enhance interpretability and simplify preprocessing, making them valuable for policymakers and economic analysts.

Studies could explore hybrid models that integrate similarity-based techniques with traditional feature selection frameworks to leverage the strengths of both approaches. For example, combining similarity measures with Wrapper methods could further boost accuracy while maintaining computational efficiency. Since the performance of similarity-based methods depends on the choice of distance metrics, research should focus on developing adaptive or data-driven methods for selecting optimal metrics based on dataset characteristics.

Adopting similarity-based feature selection methods significantly advances macroeconomic forecasting and policy analysis. These methods would improve the accuracy and efficiency of models while maintaining transparency and interpretability. By prioritizing adopting and developing these techniques, policymakers can make more informed decisions, better allocate resources, and enhance their ability to respond to economic challenges. Future efforts should focus on refining these methods, scaling their use across various domains, and integrating them into comprehensive, real-time forecasting systems to support dynamic and effective policymaking.

References

1. Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, et al. Feature selection: A data perspective. ACM Comput Surv. 2017; 50(6):1–45.
2. Remeseiro B, Bolon-Canedo V. A review of feature selection methods in medical applications. Comput Biol Med. 2019; 112:103375.
3. Zhu Z, Ong YS, Dash M. Wrapper-filter feature selection algorithm using a memetic framework. IEEE Trans Syst Man Cybern. 2007; 37(1):70–6.
4. Mitra P, Murthy CA, Pal SK. Unsupervised feature selection using feature similarity. IEEE Trans Pattern Anal Mach Intell. 2002; 24(3):301–12. doi:10.1109/34.990133
5. Shi J, Wang B, Shi Q, et al. Adaptive-similarity-based multi-modality feature selection for multimodal classification in Alzheimer's disease. Med Image Anal. 2020; 60:101618.
6. Mehri M, Chaieb R, Kalti K, Héroux P, Mullot R, Essoukri Ben Amara N. A comparative study of two state-of-the-art feature selection algorithms for texture-based pixel-labeling task of ancient documents. J Imaging. 2018; 4(8):97.
7. Shen Z, Chen X, Garibaldi JM. A novel meta-learning framework for feature selection using data synthesis and fuzzy similarity. In: 2020 IEEE Int Conf Fuzzy Syst (FUZZ-IEEE). 2020. p. 1–8.
8. Goldani M, Tirvan SA. Sensitivity assessing to data volume for forecasting: introducing similarity methods as a suitable one in feature selection methods. arXiv preprint arXiv:2406.04390. 2024.
9. Mathisen BM, Aamodt A, Bach K, Langseth H. Learning similarity measures from data. Prog Artif Intell. 2020; 9(2):129–43.
10. Qi M, Wang T, Liu F, Zhang B, Wang J, Yi Y. Unsupervised feature selection by regularized matrix factorization. Neurocomputing. 2017; 273:593–610.
11. Du S, Ma Y, Li S, Ma Y. Robust unsupervised feature selection via matrix factorization. Neurocomputing. 2017; 241:115–27.
12. Hu R, et al. Graph self-representation method for unsupervised feature selection. Neurocomputing. 2015; 220:130–7.
13. Venkatesh B, Anuradha J. A review of feature selection and its methods. Cybern Inf Technol. 2019; 19(1):3–26.
14. Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res. 2003; 3:1157–82.
15. Jović A, Brkić K, Bogunović N. A review of feature selection methods with applications. In: 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). 2015. p. 1200–5.
16. Goldani M. Comparative analysis of missing values imputation methods: a case study in financial series (S&P500 and Bitcoin value data sets). Iran J Finance. 2024; 8(1):47–70.
17. Ali M, Mazhar T, Shahzad T, Ghadi YY, Mohsin SM, Akber SMA, et al. Analysis of feature selection methods in software defect prediction models. IEEE Access. 2023.
18. Saeys Y, Inza I, Larranaga P. A review of feature selection techniques in bioinformatics. Bioinformatics. 2007; 23(19):2507–17.
19. Chen H, GAO X. A new time series similarity measurement method based on fluctuation features. Tehnički Vjesnik.

2020; 27:1134–41.
20. Salarpour A, Khatunloo H. A segmental distance-based similarity criterion using time deviation. J Electr Eng Univ Tabriz. 2019; (2):645–56.
21. Keogh E, Pazzani M. Derivative dynamic time warping. In: Proceedings of the 2001 SIAM International Conference on Data Mining. 2001. p. 1–11. doi:10.1137/1.9781611972719.1
22. Besse PC, Guillouet B, Loubes JM, Royer F. Review and perspective for distance-based clustering of vehicle trajectories. IEEE Trans Intell Transp Syst. 2016; 17(11):3306–17.
23. Chen L, Ng RT. On the marriage of Lp-norms and edit distance. In: Proceedings of the Thirtieth International Conference on Very Large Data Bases (VLDB). 2004. p. 792–803.
24. Besse PC, Guillouet B, Loubes JM, Royer F. Review and perspective for distance-based clustering of vehicle trajectories. IEEE Trans Intell Transp Syst. 2016; 17(11):3306–17.
25. Bergmeir C, Benítez JM. On the use of cross-validation for time series predictor evaluation. Inf Sci. 2012; 191:192–213.

Journal of Sciences, Islamic Republic of Iran 35(3): 257 - 265 (2024) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Comparison of Adaptive Neural-Based Fuzzy Inference System and Support Vector Machine Methods for the Jakarta Composite Index Forecasting

Ayu Mutmainnah*, Sri Astuti Thamrin, Georgina Maria Tinungki

Department of Statistics, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Makassar, 90245 Indonesia Makassar, Indonesia

Received: 2 September 2024 / Revised: 30 November 2024 / Accepted: 2 February 2025

Abstract

The Jakarta Composite Index (JCI) is a pivotal benchmark for evaluating the performance of all stocks listed on the Indonesia Stock Exchange (IDX). Given the inherent complexity, nonlinearity, and non-stationarity of stock market data, selecting robust forecasting methods is essential. This study compares the performance of the Adaptive Neuro-Fuzzy Inference System (ANFIS) and Support Vector Machine (SVM) in forecasting JCI movements. The researcher assessed prediction accuracy using Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The training phase revealed that the optimal ANFIS model employed the generalized bell membership function, outperforming trapezoidal and Gaussian alternatives. Concurrently, the best SVM configuration utilized a linear kernel (cost = 10), demonstrating superior performance compared to radial basis function (RBF) and sigmoid kernels. In the testing phase, ANFIS achieved an RMSE of 39.894 and MAPE of 0.4647, while SVM recorded an RMSE of 38.728 and MAPE of 0.4516. These results underscore the superior predictive capabilities of SVM, positioning it as a reliable tool for stock market forecasting. The study's findings provide valuable insights for investors and policymakers in navigating market uncertainties and optimizing investment strategies.

Keywords: Forecasting; Support Vector Machine; Jakarta Composite Index; Adaptive Neural-based Fuzzy Inference System.

Introduction

The Jakarta Composite Index (JCI) serves as a crucial benchmark, reflecting the overall performance of all stocks listed on the Main Board and Development Board of the Indonesia Stock Exchange (IDX) (1). Stock price movements within the JCI exhibit diverse patterns throughout the trading day, with some stocks experiencing gains, others losses, and a subset remaining unchanged (2). Figure 1 illustrates the general structure of a fuzzy inference system, while Figure 3 depicts the specific ANFIS architecture used in this study. A rising JCI trend signals an overall increase in stock prices, whereas a declining trend indicates a general downturn. For participants in the capital market, are closely monitoring stock price movements is essential to inform strategic investment decisions. However, forecasting stock market behavior poses significant challenges due to

* Corresponding Author: Tel: +62 821 3766 4330; Email: amutmainnah8@gmail.com

its inherently complex, nonlinear, and non-stationary nature.

In addressing these complexities, there have been various advanced forecasting approaches leveraging artificial intelligence developed, including the Adaptive Neuro-Fuzzy Inference System (ANFIS), Support Vector Machine (SVM), Genetic Programming (GP), and Artificial Neural Networks (ANN). In a study by When-Chuan Wang (3), the forecasting performance of these methods was compared to daily river flow data. The results demonstrated that ANFIS and SVM surpassed traditional statistical approaches such as ARMA, as well as other AI-based methods like GP and ANN, in terms of Coefficient of Determination (R^2), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE). Notably, the predicted values of ANFIS and SVM closely aligned with observed data trends, highlighting their efficacy in handling nonlinear datasets.

The ANFIS model excels at identifying intricate nonlinear patterns in data, combining the strengths of fuzzy inference systems and neural network architectures. While fuzzy inference systems can translate expert knowledge into rule-based models, determining optimal membership functions can be computationally intensive. ANFIS addresses this limitation by integrating neural network learning mechanisms, automating the search for optimal membership functions, and thus expediting the modeling process. This dual capability makes ANFIS a versatile tool for applications across various domains. For example, ANFIS has been successfully utilized to forecast and analyze air quality in Wuhan City, particularly in studying the effects of COVID-19 on environmental parameters (4).

Similarly, the Support Vector Machine (SVM) method offers a robust alternative for time series forecasting, including stock price prediction. SVM is particularly well-suited for complex, nonlinear datasets and has demonstrated high predictive accuracy when its hyperparameters are optimally tuned. A comparative study of SVM and Backpropagation-based ANN for forecasting foreign tourist arrivals in Bali Province revealed that SVM, using a radial basis function kernel, outperformed ANN by achieving the lowest forecasting errors (5).

The high potential returns offered by the Indonesian stock market have attracted significant interest from domestic and international investors, particularly in comparison to other regional markets. The potential underscores the importance of accurate stock price forecasting to maximize investment returns. Previous studies have consistently shown that ANFIS and SVM outperform other forecasting methods in terms of predictive accuracy. Therefore, this study seeks to apply ANFIS and SVM methodologies to forecast the Jakarta Composite Index (JCI) to contribute to more informed investment strategies.

Materials and Methods

Forecasting

Forecasting involves estimating future values based on historical data, typically employing statistical and computational methods. It is an essential tool in decision-making processes, allowing for predicting future trends using past observations. Time series analysis is widely used among the various approaches, relying on historical values and error patterns to predict future outcomes over time (6).

Adaptive Neural-Based Fuzzy Inference System (ANFIS)

ANFIS integrates fuzzy inference systems with neural network architecture, leveraging the strengths of both approaches. While fuzzy inference systems excel in translating expert knowledge into rule-based models, they often require significant effort to determine optimal membership functions. Neural networks streamline this process by automating the search for membership functions, enhancing the applicability of ANFIS across diverse fields (7). Assuming a Fuzzy inference system with two inputs, x_1 , x_2, and single output Y., the first-order Sugeno fuzzy model can be represented as follows:

if $x_1 = A_1$ and $x_2 = B_1$, than $f_1 = p_{1x} + q_{1y} + r_1$

if $x_1 = A_2$ and $x_2 = B_2$, than $f_1 = p_{2x} + q_{2y} + r_2$

Here $\boldsymbol{A_i}$ and $\boldsymbol{B_i}$ are linguistic labels (e.g., low, medium, high) represented by membership functions, and pi, qi, and ri are consequent parameters.

Member Functions of ANFIS

Fuzzy set theory extends classical set theory by allowing degrees of membership for elements. The degree of membership, denoted by μA(x), quantifies how much an element x belongs to a fuzzy set A (8). Membership values are defined using functions such as:

1. Trapezoidal Membership Function:

$$f(x,a,b,c,d) = \begin{cases} 0; & x < a \\ \dfrac{x-a}{b-a}; & a \le x \le b \\ 1; & b \le x \le c \\ \dfrac{d-x}{d-c}; & c \le x \le d \\ 0; & x > d \end{cases}$$

2. Generalized Bell Membership Function:

$$B(x,a,b,c,d) = \frac{1}{1+\left[\left(\frac{x-c}{a}\right)^2\right]^b}$$

3. Gaussian Membership Function:

$$G(x,\mu,\sigma) = exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

A Fuzzy Inference System (FIS) is a computational framework grounded in Fuzzy set theory, utilizing Fuzzy rules (in the form of IF-THEN statements) and Fuzzy reasoning. The system receives a crisp input, which is then processed by a knowledge base containing Fuzzy rules in the IF-THEN format. The system evaluates the "fire strength" for each rule. When multiple rules are present, the system aggregates the outcomes of all the rules. Finally, the aggregated results are defuzzified to produce a crisp output value (9).

Architecture of ANFIS

The network architecture of the ANFIS method consists of five layers, as illustrated in Figure 2 (10). The ANFIS network comprises five layers, each with distinct roles:

1. Fuzzification Layer: Calculates the degree of membership for each input using membership functions. The premise parameters are adapted in this layer. Suppose $x_1 = X_1$ and $x_2 = X_2$, the node function is described by the following equation:

$$O_{1,1} = \mu_{A_1}(X_1)$$

$$O_{1,2} = \mu_{A_2}(X_1)$$
$$O_{1,3} = \mu_{B_1}(X_2)$$
$$O_{1,4} = \mu_{B_2}(X_2)$$

2. Fuzzy Logic Operation Layer: Computes the firing strength of rules using the product of input memberships. The node function for this layer can be described by the following equation:

$$O_{2,1} = w_i = \mu_{A_i}(X_1)\mu_{B_i}(X_2)$$

3. Normalization Layer: Normalizes the firing strengths to ensure proportionality.

$$O_{3,i} = \overline{w}_i = \frac{w_i}{\sum_i w_i}$$

4. Defuzzification Layer: Calculates the weighted output of each rule using consequent parameters.

$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i(c_{i1}x_1 + c_{i2}x_2 + c_{i0})$$

5. Output Layer: Aggregates the results from all rules to produce the final model output.

$$O_5 = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}$$

Parameter estimation is performed using hybrid learning, combining the Recursive Least Squares Estimation (RLSE) for linear parameters and Backpropagation for nonlinear parameters (17).

Support Vector Machine (SVM)

SVM is a machine learning algorithm grounded in statistical learning theory, suitable for classification and regression tasks (11, 12). SVM maps input data into a

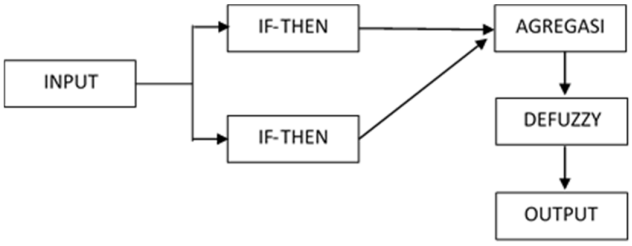


Figure 1. Fuzzy Inference System

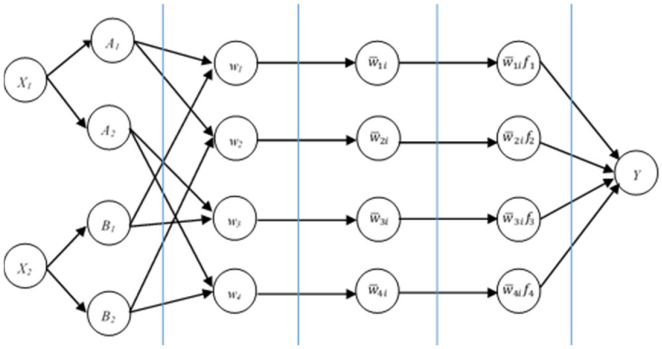


Figure 2. Architecture of ANFIS

high-dimensional feature space using kernel functions, enabling the separation of non-linearly separable data. The kernel trick transforms data into a higher-dimensional space, facilitating linear separation (13).

Commonly used kernels include:

1. Linear Kernel:
$$\mathrm{K(x,z)} = x^T z$$
2. Sigmoid Kernel:
$$\mathrm{K(x,z)} = \tanh\left(\gamma \,.\, x^T z + r\right)$$
3. Radial Basis Function (RBF) Kernel:
$$\mathrm{K(x,z)} = exp\left\{-\frac{\|x-z\|^2}{2\sigma^2}\right\}$$

The optimal SVM parameters are typically identified using a grid search algorithm, which systematically evaluates combinations of parameters.

Performance Metrics for Model Evaluation

The accuracy of forecasting models is assessed using the following metrics:

Root Mean Square Error (RMSE)

$$RMSE = \left(\frac{\sum(y_i - \hat{y}_i)}{n}\right)^{\frac{1}{2}}$$

Lower RMSE values indicate better predictive accuracy (14).

Mean Absolute Percentage error (MAPE)

$$MAPE = \frac{\sum_{t=1}^{n} |(\frac{y_i - \hat{y}_i}{y_i})| \, 100\%}{n}$$

Lower MAPE values denote higher forecasting precision (15).

Results

This study utilizes a dataset comprising daily closing prices of Jakarta Composite Index (JCI) shares over two years, specifically from January 2, 2020 to December 29, 2023. The movement pattern of stock prices is highly volatile and exhibits non-linear characteristics. To address this, the research applies artificial intelligence methods, including the Adaptive Neuro-Fuzzy Inference System (ANFIS) and Support Vector Machine (SVM). Before modeling, the dataset is divided into training and testing subsets, as detailed in Tables 1 and 2.

Table 3 shows that the dataset consists of 974 observations of JCI stock prices collected during the study period. The data is split so that 80% is allocated for training, and 20% is reserved for testing. The training data is utilized to build prediction models, which are subsequently validated using the testing data to evaluate their predictive performance.

ANFIS Training Process

The ANFIS model is implemented using MATLAB software, with an error tolerance of zero and a maximum of 20 epochs. Before the training process, the data is normalized to 0 to 1 to enhance computational efficiency and meet system requirements. The ANFIS model uses the JCI stock closing price as the target variable and includes six input variables derived from the prior six days. Mathematically, the ANFIS model is expressed as:

Table 1. Training Data and Target Data

Data	Training Data	Target Data
1	Data from the 1st day to the 6th day	Data from the 7th day
2	Data from the 2nd day to the 7th day	Data from the 8th day
3	Data from the 3rd day to the 8th day	Data from the 9th day
…	…	…
774	Data from the 768th day to the 773th day	Data from the 774th day

Table 2. Dataset division

No	Dataset division	Period	Number of Data
1.	Training Data	January 2, 2014 – March 6, 2023	780
2.	Testing Data	March 7, 2023 – December 29, 2023	194

Table 3. Nonlinear Parameters of Trapezoidal Function

Input	a	b	c	d
Input 1 mf1 (A1)	-0.7000	-0.3000	0.3032	0.6919
Input 1 mf2 (A2)	0.2536	0.6983	1.3000	1.7000
Input 2 mf1 (B1)	-0.7000	-0.3000	0.2986	0.6892
…	…	…	…	…
Input 6 mf2 (F2)	0.2879	0.6988	1.3000	1.7000

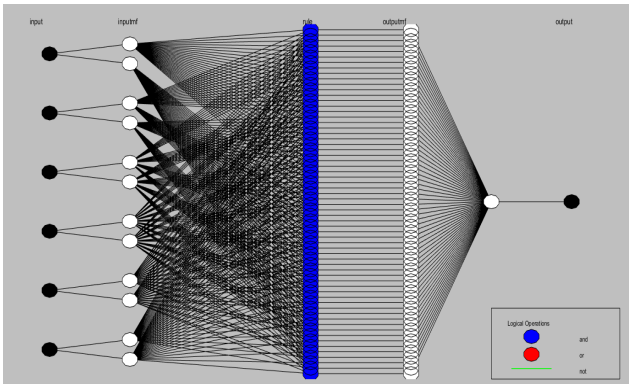


Figure 3. Anfis Structure Of Jci Stock Price Data

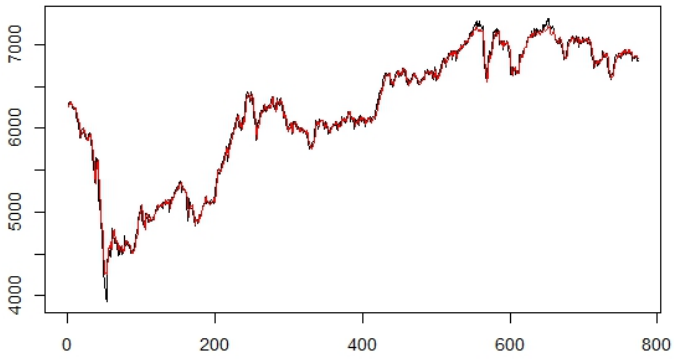


Figure 4. Svm Prediction Before Optimization

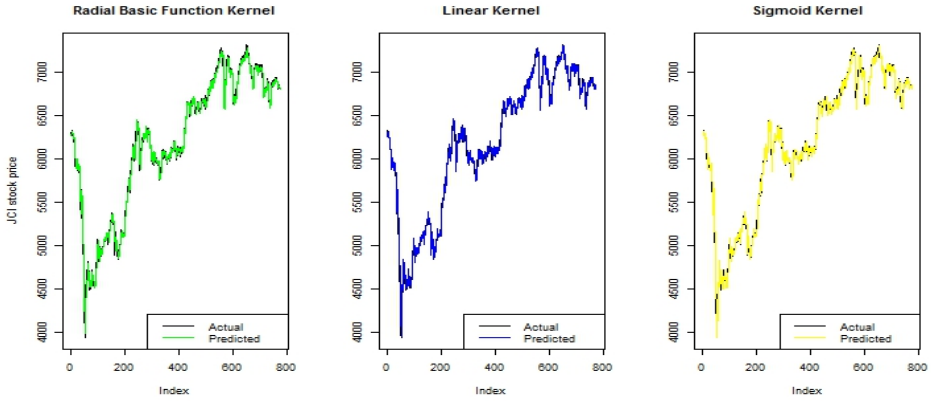


Figure 5. Svm Prediction Using Kernels

$$\begin{aligned}\bar{w}_{1t}.f_1 = \bar{w}_{1t}\big(&C_{1,1}X_1 + C_{1,2}X_2 + C_{1,3}X_3 + C_{1,4}X_4 \\ &+ C_{1,5}X_5 + C_{1,6}X_6 + C_{1,0}\big) + \cdots \\ &+ \bar{w}_{64t}\big(C_{64,1}X_1 + C_{64,2}X_2 + C_{64,3}X_3 \\ &+ C_{64,4}X_4 + C_{64,5}X_5 + C_{64,6}X_6 \\ &+ C_{64,0}\big)\end{aligned}$$

Here $X_1, X_2, X_3, X_4, X_5$and X_6 represent the six input variables.wwithh denotes the normalized. firing strength for rule,Cij,and i represent the linear parameters for rule i.

In the ANFIS architecture, the first layer performs fuzzification, transforming crisp values into Fuzzy numbers based on membership functions, such as Trapezoidal, Generalized Bell, and Gaussian. The nonlinear parameters in the membership functions are optimized using a backpropagation error method, as

Table 4. Nonlinear Parameters of Generalized Bell Function

Input	a	b	c
Input 1 mf1 (A1)	0.4823	2.0050	-0.0036
Input 1 mf2 (A2)	0.4414	2.0005	1.0465
Input 2 mf1 (B1)	0.4246	2.0113	-0.0250
…	...	...	...
Input 6 mf2 (F2)	0.4260	2.0132	1.0192

Table 5. Nonlinear Parameters of Gaussian Function

Input	μ	σ
Input 1 mf1 (A1)	0.3614	-0.0213
Input 1 mf2 (A2)	0.3477	1.0324
Input 2 mf1 (B1)	0.3641	-0.0187
…	...	...
Input 6 mf2 (F2)	0.3493	1.0255

illustrated in Tables 3–5. The mathematical representations of the membership functions are as follows:

a. Trapezoidal Membership Function:

$$\mu_{A1}(X_1) = \begin{cases} 0, & X_1 < -0.7 \\ \dfrac{X_1-(-0.7)}{-0.3-(-0.7)}, & 0.7 \le X_1 \le -0.3 \\ 1, & -0.3 \le X_1 \le 0.3032 \\ \dfrac{(0.5505-X_1)}{0.6919-0.3032}, & 0.3032 \le X_1 \le 0.6919 \\ 0, & X_1 > 0.6919 \end{cases}$$

b. Generalized Bell Membership Function:

$$\mu_{A1}(X_1, 0.4823, 2.0050, -0.0036) = \frac{1}{1+\left[\left(\frac{x-0.0036}{0.4823}\right)^2\right]^{2.0050}}$$

c. Gaussian Membership Function:

$$\mu_{A1}(X_1, 0.3614, -0.0213) = exp\left(\frac{-(x-0.3614)^2}{2(0.0213)^2}\right)$$

The second layer computes the firing strength (αα-predicate) using Zadeh's AND operator, combining the membership degrees generated in the first layer. The step produces 64 rules derived from 2^6, representing all possible combinations of inputs and membership functions.

In the third layer, the firing strengths are normalized by dividing each by the total sum of all firing strengths. The normalized values are then defuzzified in the fourth layer, where fuzzy outputs are converted to crisp values using linear parameters optimized through the Least Squares Estimation (LSE) method, Tables 6–8 present detailed model parameters. The fifth and final layer aggregates the outputs to generate the final predictions in Table 9.

A comparison of membership functions is carried out to find the best model with the following ANFIS best model criteria comparison. Based on the goodness of output model criteria in Table 10, the Generalized Bell function in ANFIS performs best with the lowest MAPE value of 0.4284. This function shows the highest relative accuracy in forecasting stock prices compared to the Gaussian and Trapezoidal functions. Therefore, the ANFIS model with the Generalized Bell membership function will be used for testing the Test Data.

SVM Training Process

SVM models are implemented using the e1071 package in R Studio. This method transforms input data into a high-dimensional feature space using kernel functions, constructing an optimal hyperplane for classification or regression. Initially, the SVM model uses default parameters, resulting in predictions that deviate significantly from the actual data, as shown in Figure 4.

Parameter optimization is conducted using a grid search method, testing combinations of parameters: $\varepsilon = \{0,0.1,0.2,\ldots,1\}$, $cost = \{2^{-2}, 2^{-1}, \ldots, 2^9\}$ and $gamma = \{2^9, 2^8, \ldots, 2^2\}$. Cross-validation is employed to evaluate model performance for each parameter combination. The tested kernel functions are the Radial Basis Function (RBF), linear, and sigmoid. The linear kernel produces predictions closest to actual data, as illustrated in Figure 5.

The linear kernel achieves the lowest Root Mean Square Error (RMSE) compared to other kernels, with optimal parameter values of cost=10cost=10. This kernel is selected to forecast test data and predict future JCI stock prices.

Discussion

Following the training data analysis using both

Table 6. Linear Parameters of Trapezoidal Function

Input	C_1	C_2	C_3	...	C_6	C_0
Output mf1	-0.1262	0.1866	-0.2717	...	0.8620	0.0656
Output mf2	-3.0476	-4.4460	1.1517	...	-0.7315	1.6118
Output mf3	0.5082	-10.7592	0.5622	...	-10.9661	2.7333
...	...	...	...	...	...	...
Output mf64	0.0305	-0.0868	-0.0596	...	0.9014	0.0273

Table 7. Linear Parameters of Generalized Bell Function

Input	C_1	C_2	C_3	...	C_6	C_0
Output mf1	0.4425	-0.0151	-0.9043	...	-0.4460	0.4707
Output mf2	-0.5735	-12.2232	3.5131	...	3.0817	1.8571
Output mf3	-7.0566	4.5672	6.7993	...	0.9636	-5.0750
...	...	...	...	...	...	...
Output mf64	-1.8858	0.7730	0.4699	...	0.6007	0.3415

Table 8. Linear Parameters of Gaussian Function

Input	C_1	C_2	C_3	...	C_6	C_0
Output mf1	0.1104	-0.2049	-0.5585	...	-0.4439	0.4432
Output mf2	-2.7432	-12.6954	5.3103	...	-3.1770	4.4031
Output mf3	-2.5234	4.1647	2.6423	...	0.3907	-2.8141
...	...	...	...	...	...	...
Output mf64	-1.9919	1.2109	0.6571	...	0.5844	-0.0962

Table 9. Forecasting Result of ANFIS

Date	Actual Data	Gaussian Training Output	Trapezoidal Training Output	Generalized Bell Training Output
1/10/2020	6274.941	6285.760	6282.905	6295.116
1/13/2020	6296.567	6253.760	6280.237	6251.545
1/14/2020	6325.406	6291.051	6299.900	6286.146
1/15/2020	6283.365	6310.446	6292.180	6304.323
…	…	…	…	…
12/29/2023	6807.001	6822.114	6819.001	6816.086

Table 10. Comparison of ANFIS Model Goodness Criteria

Member Functions	MAPE	RMSE
Trapezodial Function	0.6206	50.438
Generalized Bell Function	0.4284	37.637
Gaussian Funcstion	0.6520	52.153

Adaptive Neuro-Fuzzy Inference System (ANFIS) and Support Vector Machine (SVM) models, forecasting was performed on the testing data to identify the most effective approach for predicting JCI stock prices. In the case of ANFIS, the model employed a Generalized Bell membership function, which demonstrated strong performance in the training phase. Conversely, the SVM model utilized a linear kernel with a cost parameter set to 10, which was optimized during the grid search process.

To evaluate and compare the forecasting performance of both models on the testing data, accuracy metrics, including Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE), were calculated and are presented in Table 11. ANFIS achieved a MAPE of 0.4647 and an RMSE of 39.894, whereas SVM outperformed ANFIS with a MAPE of 0.4516 and an RMSE of 36.728. These results clearly indicate that SVM offers superior accuracy and predictive power for forecasting JCI stock prices compared to ANFIS.

As shown in Figure 6 and Table 12, the SVM model's predictions are noticeably closer to the actual stock prices than the ANFIS model's. The graphical representation further underscores the conclusion that the SVM method provides more accurate and reliable predictions for the Jakarta Composite Index (JCI) stock prices than the ANFIS model.

Table 11. Comparison of SVM model goodness criteria

Kernel	MAPE	RMSE
Radial Basic Function	0.7851	64.483
Sigmoid	0.7789	63.893
Linear	0.7787	63.564

Table 12. Comparison of ANFIS and SVM Forecasting Results

Time	Actual Data	ANFIS	SVM
27-Mar-2023	6628.137	6657.343	6667.709
28-Mar-2023	6565.728	6628.467	6658.872
29-Mar-2023	6678.237	6564.476	6574.031
30-Mar-2023	6612.490	6663.021	6675.338
31-Mar-2023	6691.611	6628.246	6607.709
3-Apr-2023	6762.254	6698.378	6696.484
4-Apr-2023	6708.933	6766.400	6727.627
5-Apr-2023	6760.328	6724.016	6719.136
6-Apr-2023	6839.436	6772.216	6759.006
15-Mar-2023	6808.951	6833.755	6811.873
16-Mar-2023	6805.277	6818.610	6812.346
17-Mar-2023	6827.175	6819.812	6814.796
20-Mar-2023	6833.178	6826.516	6820.83
…	…	…	…
29-Dec-2023	7272.797	7288.872	7302.981

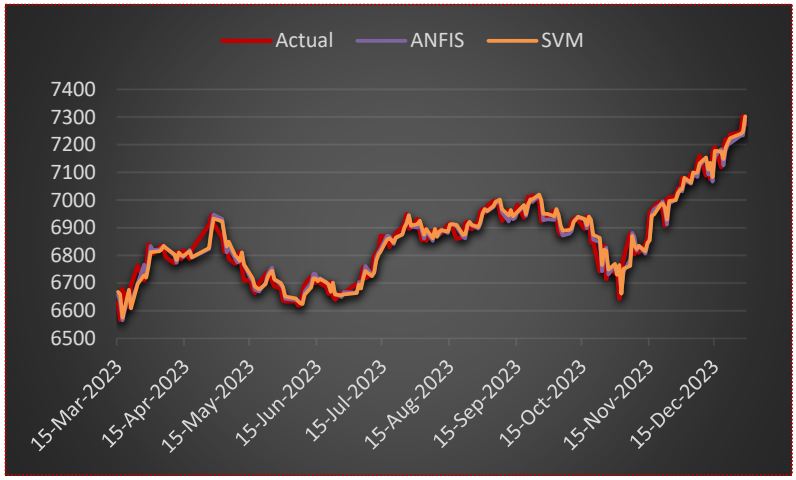


Figure 6. Anfis And Svm Forecasting Result Graphs

This outcome suggests that SVM can effectively optimize hyperparameters using kernel methods and is better suited for capturing the non-linear patterns inherent in stock price data. Meanwhile, although ANFIS showed reasonable performance, SVM outperformed it in terms of both MAPE and RMSE, highlighting the advantage of SVM in this particular forecasting context.

References

1. Bursa Efek Indonesia. Bursa Efek Indonesia (Internet). 2023 Mar 4 (cited 2024 Nov 25). Available from: https://www.idx.co.id/id/produk/saham
2. Arviana, Nerissa, Geofanni. Sebelum Investasi, Pahami Dulu Apa Itu IHSG dan Berbagai Fungsinya [Internet]. Glints.com. 2023 Jan 19 [cited 2024 Nov 30]. Available from: https://glints.com/id/lowongan/apa-itu-ihsg/#.ZAMkgh_P3IU
3. Wang WC, Chau KW, Cheng CT, Qiu L. A comparison of

performance of several artificial intelligence methods for forecasting daily discharge time series. J Hydrol. 2009;368(1–4):294–306.
4. Al-qaness MA, Fan H, Ewees AA, V DY. Improved ANFIS model for forecasting Wuhan City Air Quality and analysis COVID-19 lockdown impacts on air quality. Environ Res. 2021;194:110607.
5. Tarigan IA, Bayupati IP, Putri GA. Komparasi model Support Vector Machine dan Backpropagation dalam peramalan jumlah wisatawan mancanegara di Provinsi Bali. J Teknol Syst Komput. 2021;9(2):90–5.
6. Makridakis S, Wright S, McGee VE. Metode dan aplikasi peramalan. 2nd ed. Jakarta: Binaputra Aksara; 1999.
7. Widyapratiwi LK, Mertasana IP, Arjana IG. Peramalan beban listrik jangka pendek di Bali menggunakan pendekatan Adaptive Neuro-Fuzzy Inference System (ANFIS). Teknol Elektro. 2012;11(1):7–13.
8. Widodo PP. Penerapan soft computing dengan Matlab. Bandung: Rekayasa Sains; 2012.
9. Wulandari A, Rosita, Gernowo R. Metode Autoregressive Integrated Moving Average (ARIMA) dan metode Adaptive Neuro Fuzzy Inference System (ANFIS) dalam analisis curah hujan. Berkala Fisika. 2019;22(1):41–8.
10. Jang JS, Sun CT, Mizutani E. Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence. New Jersey: Prentice Hall; 1997.
11. Vapnik VC. Support-vector networks. Mach Learn. 1995;20(3):273–97.
12. Santoso B. Data mining teknik pemanfaatan data untuk keperluan bisnis. Yogyakarta: Graha Ilmu; 2007.
13. Campbell C, Ying Y. Learning with Support Vector Machine. Morgan & Claypool Publishers; 2011.
14. Albert S, Shantika M, Syahru R. Sistem informasi peramalan tren pelanggan dengan menggunakan metode double exponential smoothing di MESS GM. J Komput Apl. 2020;10(4):237–46.
15. Sezer OB, Gudelek MU, Ozbayoglu AM. Financial time series forecasting with deep learning: a systematic literature review (2005–2019). Appl Soft Comput. 2020;90:106181.
16. Novitasari HR. Weather parameters forecasting as variables for rainfall prediction using Adaptive Neuro Fuzzy Inference System (ANFIS) and Support Vector Regression (SVR). In: International Conference on Science & Technology (ICoST 2019). 2020. p. 1–6

Journal of Sciences, Islamic Republic of Iran 35(3): 267 - 277 (2024) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Second-ordered Characterization of Generalized Convex Functions and Their Applications in Optimization Problems

M. T. Nadi[1*], J. Zafarani[2]

[1] *School of Mathematics, Institute for Research in Fundamental Sciences (IPM), P. O. Box 19395-5746, Tehran, Islamic Republic of Iran*
[2] *Department of Mathematics, Sheikhbahaee University and University of Isfahan, Isfahan, Islamic Republic of Iran*

Received: 26 January2025 / Revised: 16 April 2025 / Accepted: 20 May 2025

Abstract

This survey investigates some developments in the second-order characterization of generalized convex functions using the coderivative of subdifferential mapping. More precisely, it presents the second-order characterization for quasiconvex, pseudoconvex and invex functions. Furthermore, it gives some applications of the second-order subdifferentials in optimization problems such as constrained and unconstrained nonlinear programming.

Keywords: Second-order subdifferential; Positive semidefinite property; Regular second-order subdifferential; Second-order optimality conditions.
Mathematics Subject Classification (2010): 26B25, 49J40, 49J52, 49J53, 90C33

Introduction

Second-order subdifferentials and their application in the optimization and characterization of various kinds of convexity have attracted the attention of the literature. It is well known that the second-order differential of a twice continuously differentiable function $g: \mathbb{R}^{\mathbf{n}} \to \mathbb{R}$ is convex if and only if $\nabla^2 g$ (its Hessian matrix) is positive semidefinite and g is strictly convex when $\nabla^2 g$ is positive definite everywhere.

This result is true even in normed spaces:

Theorem 1.1 (Flett, 1980) Let X be a real normed space and let $g: X \to \mathbb{R}$ be a twice Fréchet differentiable function, then g is convex if and only if $d^2g(x)(y)^2 \geq 0$ for all $x, y \in X$.

Convex functions and their generalizations have many applications in optimization, economy, control theory and several other sciences; thus the characterization of convex functions is fundamental and useful. We know that when a C^2 function $g: \mathbb{R}^{\mathbf{n}} \to \mathbb{R}$ attains its minimum at x, its Hessian is positive semidefinite and conversely, the positive definiteness of its Hessian is sufficient for g to reach its minimum at x when $(\nabla g)(x) = 0$. Indeed the strict local convexity of g guaranteed by positive definiteness of $\nabla^2 g(x)$. Some authors have studied the characterization of convex functions and their generalizations by their subdifferentials. Also, the second-order optimality conditions have received much attention in optimization theory, in recent years; see (1,2,3) for example.

Theorem 1.2 (4, Rockafellar 1970) The maximal monotonicity of Fréchet subdifferential of a lower semicontinuous function is a necessary and sufficient condition for its convexity.

Characterization of generalized convex functions by second-order subdifferentials can be more useful, especially in optimization.

* Corresponding Author: Tel: +989394520191; Email: mt_nadi@yahoo.com

Second-order characterization of convex functions by generalized second-order directional derivatives have studied by some authors.

The upper Dini-directional derivative of g at $x \in X$ in direction $v \in X$ is defined as an element of $\mathbb{R}$ by

$$g'_+(x;v) = \limsup_{t\downarrow 0} t^{-1}(g(x+tv) - g(x)). \quad (1)$$

The second-order upper Dini-directional derivative of g at $x \in X$ in direction $v \in X$ for which $g'_+(x;v)$ is defined by

$$g''_+(x;v) = \limsup_{t\downarrow 0} 2t^{-2}(g(x+tv) - g(x) - tg'_+(x;v)). \quad (2)$$

In the case of an infinite $g'_+(x;v)$, the derivative $g''_+(x;v)$ will not be considered.

Theorem 1.3 (5, Ginchev and Ivanov 2003) Let $g: X \to \mathbb{R}$ be u.s.c. Then g is convex on X if and only if the following Conditions (C_1) and (C_2) hold for each $x, u \in X$:

(C_1) $g'_+(x;v) + g'_+(x;-v) \geq 0$,

if the expression on the left-hand side has the sense

(C_2) $g'_+(x;v) + g'_+(x;-v) = 0$, implies that $g''_+(x;u) \geq 0$.

Example 1.1 The function $g(x) = -|x|, x \in \mathbb{R}$, satisfies the equality $g''_+(x;v) = 0$ for all $x, v \in \mathbf{R}$. It is continuous, but not convex. Obviously, $g'_+(x;v) + g'_+(x;-v) = -2$.

Example 1.2 The function $g: \mathbb{R} \to \mathbb{R}$ defined as

$$g(x) = \begin{cases} x^2, & \text{if x is rational ;} \\ 0, & \text{otherwise} \end{cases}$$

satisfies conditions (C_1) and (C_2), but g is not convex. This function is not u.s.c.

Some other authors used the second-order Fréchet (Second-order regular subdifferentials) and Mordukhovich (limiting) subdifferentials defined by the coderivative of the subdifferential mappings. See (6,7) for the following definitions and more details.

Let X be a Banach space endowed with a norm $\|.\|$, X^* its dual space, X^{**} its second dual space and $\langle .,. \rangle$ be the dual pairing between X and X^*. For a set-valued mapping $T: X \rightrightarrows Y$ between Banach spaces, we define the effective domain and the graph of T by

$$domT = \{x \in X: T(x) \neq \emptyset\}, \quad gphT = \{(x,y) \in X \times Y: y \in T(x)\}.$$

The sequential Painlevé-Kuratowski upper limit of T at x in the topology of Y is defined by

$$\limsup_{x\to\bar{x}} T(x) = \{y \in Y: \exists \text{ sequences } x_k \to \bar{x}, y_k \to y \text{ with } y_k \in T(x_k),\ \forall k = 1,2,\dots\}.$$

Given $\varepsilon \geq 0$ and $\Omega \subseteq X$, the $\varepsilon - normals$ to Ω at $\bar{x} \in cl(\Omega)$ is defined by

$$\widehat{N}_\varepsilon(\bar{x};\Omega) := \{x^* \in X^*: \limsup_{x\xrightarrow{\Omega}\bar{x}} \frac{\langle x^*, x-\bar{x}\rangle}{||x-\bar{x}||} \leq \varepsilon\},$$

where the symbol $x \xrightarrow{\Omega} \bar{x}$ means that $x \to \bar{x}$ with $x \in \Omega$. When $\varepsilon = 0$, the set $\widehat{N}_0(\bar{x},\Omega) = \widehat{N}(\bar{x},\Omega)$ is named the prenormal cone or Fréchet normal to Ω at $\bar{x}$.

The limiting or Mordukhovich normal cone to Ω at $\bar{x}$ is

$$N(\bar{x};\Omega) := \limsup_{x\to\bar{x},\varepsilon\downarrow 0} \widehat{N}_\varepsilon(x;\Omega),$$

where the sequential Painlevé-Kuratowski upper limit is taking in the $weak^*$ topology of X^*. When X is an Asplund Banach space and Ω is closed, we can put $\varepsilon = 0$.

Definition 1.1 (6) The Fréchet or regular coderivative of T at $(\bar{x},\bar{y})$ is

$$\widehat{D}^*T(\bar{x},\bar{y})(y^*) = \{x^* \in X^*: (x^*,-y^*) \in \widehat{N}((\bar{x},\bar{y}), gph\ T)\}\ \forall y^* \in Y^*.$$

The limiting or Mordukhovich coderivative of T at $(\bar{x},\bar{y})$ is

$$D^*T(\bar{x},\bar{y})(y^*) = \{x^* \in X^*: (x^*,-y^*) \in N((\bar{x},\bar{y}), gph\ T)\}\ \forall y^* \in Y^*.$$

i.e.,

$$D^*T(\bar{x},\bar{y})(y^*) = \{x^* \in X^*: \exists \varepsilon_k \downarrow 0, (x_k,y_k) \text{ with } (x_k^*,-y_k^*) \in \widehat{N}_{\varepsilon_k}((x_k,y_k), gph\ T) \text{ as } k \to \infty\}.$$

Definition 1.2 (6) The mixed coderivative of T at $(\bar{x},\bar{y})$ is

$$D_M^*T(\bar{x},\bar{y})(y^*) = \{x^* \in X^*: \exists \varepsilon_k \downarrow 0, (x_k,y_k,y_k^*) \to (\bar{x},\bar{y},y^*), x_k^* \xrightarrow{w^*} x^* \text{ with } (x_k^*,-y_k^*) \in \widehat{N}_{\varepsilon_k}((x_k,y_k), gph\ T) \text{ as } k \to \infty\}.$$

Definition 1.3 (6) A single-valued mapping $g: X \to Y$ is said to be strictly differentiable at $\bar{x}$ if there is a linear continuous operator $\nabla g(\bar{x}): X \to Y$ such that

$$\lim_{u,x\to\bar{x}} \frac{g(x) - g(u) - \langle \nabla g(\bar{x}), u-x\rangle}{\|x-u\|} = 0.$$

When g is single-valued and strictly differentiable at $\bar{x}$ or continuously differentiable around $\bar{x}$, with the adjoint operator $\nabla g(\bar{x})^*: Y^* \to X^*$, we have

$D^*g(\bar{x})(y^*) = \widehat{D}^*g(\bar{x})(y^*) = \{\nabla g(\bar{x})^* y^*\}$ for all $y^* \in Y^*$.

Let $g: X \to \overline{\mathbb{R}} = [-\infty,+\infty]$ be an extended real-valued function. We define

$$domg = \{x \in X: |g(x)| < \infty\} \text{ and } epi(g) = \{(x,\mu) \in (X \times \mathbb{R}): \mu \geq g(x)\}.$$

The Fréchet subdifferential or presubdifferential of g at $\bar{x} \in$ dom g is defined by

$\hat{\partial}g(\bar{x}) = \{x^* \in X^*: (x^*, -1) \in \hat{N}((\bar{x}, g(\bar{x})), \text{epi } g)\}$

and the basic or Mordukhovich limiting subdifferential is defined by

$\partial g(\bar{x}) = \{x^* \in X^*: (x^*, -1) \in N((\bar{x}, f(\bar{x})), \text{epi } g)\}$.

For $\bar{x} \notin \text{dom } f$, we put $\hat{\partial} g(\bar{x}) = \partial g(x) = \emptyset$. Also, g is said to be lower regular at $\bar{x}$ if $\hat{\partial} g(\bar{x}) = \partial g(\bar{x})$.

Definition 1.4 (6) Let $g: X \longrightarrow \overline{\mathbb{R}}$ be a function and its value at $\bar{x}$ is finite,

(i) For any $\bar{y} \in \partial g(\bar{x})$, the mapping $\partial^2 g(\bar{x}, \bar{y}): X^{**} \rightrightarrows X^*$ with the values

$$\partial^2 g(\bar{x}, \bar{y})(v) = (D^* \partial g)(\bar{x}, \bar{y})(v), (v \in X^{**}),$$

is called the limiting or Mordukhovich second-order subdifferential of g at $\bar{x}$ relative to $\bar{y}$.

(ii) For any $\bar{y} \in \hat{\partial} g(\bar{x})$, the mapping $\hat{\partial}^2 g(\bar{x}, \bar{y}): X^{**} \rightrightarrows X^*$ with the values

$$\hat{\partial}^2 g(\bar{x}, \bar{y})(v) = (\hat{D}^* \hat{\partial} g)(\bar{x}, \bar{y})(v), (v \in X^{**}),$$

is called the Fréchet second-order subdifferential of g at $\bar{x}$ relative to $\bar{y}$.

(iii) For any $\bar{y} \in \partial g(\bar{x})$, the mapping $\breve{\partial}^2 g(\bar{x}, \bar{y}): X^{**} \rightrightarrows X^*$ with the values

$$\breve{\partial}^2 g(\bar{x}, \bar{y})(v) = (\hat{D}^* \partial g)(\bar{x}, \bar{y})(v), (v \in X^{**}),$$

is called the Combined second-order subdifferential of g at $\bar{x}$ relative to $\bar{y}$.

(iv) For any $\bar{y} \in \partial g(\bar{x})$, the mapping $\partial_M^2 g(\bar{x}, \bar{y}): X^{**} \rightrightarrows X^*$ with the values

$$\partial_M^2 g(\bar{x}, \bar{y})(v) = (D_M^* \partial g)(\bar{x}, \bar{y})(v), (v \in X^{**}),$$

is called the mixed second-order subdifferential of g at $\bar{x}$ relative to $\bar{y}$. When the function g is C^2 around $\bar{x}$ and $v \in X^{**}$, we have

$$\hat{\partial}^2 g(\bar{x})(v) = \partial^2 g(\bar{x})(v) = \partial_M^2 g(\bar{x})(v) = \breve{\partial}^2 g(\bar{x})(v) = \{(\nabla^2 g(\bar{x}))^* v\},$$

where $(\nabla^2 g(\bar{x}))^*$ is the adjoint operator of the Hessian $\nabla^2 g(\bar{x})$.

Definition 1.5 (PSD) holds for $g: X \longrightarrow \overline{\mathbb{R}}$, in the Fréchet sense, when $\langle z, v\rangle \geq 0$ for every $v \in X^{**}$ and $z \in \hat{\partial}^2 g(x, y)(v)$ with $(x, y) \in \text{gph } \hat{\partial} g$.

When $\langle z, v\rangle > 0$ whenever $v \neq 0$, (PD) holds in the Fréchet sense for g.

Also, (PSD) holds in the limiting sense, when $\langle z, v\rangle \geq 0$ for every $v \in X^{**}$ and $z \in \partial^2 g(x, y)(v)$ with $(x, y) \in \text{gph } \partial g$.

When $\langle z, v\rangle > 0$ whenever $v \neq 0$, (PD) holds in the limiting sense for g.

Chieu and Huy considered these cases and extended those results for the class of C^1 functions $g: X \to \mathbb{R}$, where X is a Hilbert space or an Asplund space.

Theorem 1.4 (8, Chieu, Huy 2011) Let $g: X \to R$ be a C^1 function and X be an Asplund space. Then g is convex if the following condition holds:

$\langle z, v\rangle \geq 0$ for all $v \in X^{**}, z \in \hat{\partial}^2 g(x, y)(v)$ with $(x, y) \in \text{gph } \hat{\partial} g$.

Results

Definition 2.1 A proper subdifferentials and their application in the optimization and characterization of various kinds of convexity have attracted the attention of the literature. It is well known that the second-order differential of a twice continuously differentiable function $g: \mathbb{R}^n \to \mathbb{R}$ is convex if and only if $\nabla^2 g$ (its Hessian matrix) is positive semidefinite and g is strictly convex when $\nabla^2 g$ is positive definite everywhere.

This result is true even in normed spaces:

Theorem 1.1 (Flett, 1980) Let X be a real normed space and let $g: X \to \mathbb{R}$ be a twice Fréchet differentiable function, then g is convex if and only if $d^2 g(x)(y)^2 \geq 0$ for all $x, y \in X$.

Convex functions and their generalizations have many applications in optimization, economy, control theory and several other sciences; thus the characterization of convex functions is fundamental and useful. We know that when a C^2 function $g: \mathbb{R}^n \to \mathbb{R}$ attains its minimum at x, its Hessian is positive semidefinite and conversely, the positive definiteness of its Hessian is sufficient for g to reach its minimum at x when $(\nabla g)(x) = 0$. Indeed the strict local convexity of g guaranteed by positive definiteness of $\nabla^2 g(x)$. Some authors have studied the characterization of convex functions and their generalizations by their subdifferentials. Also, the second-order optimality conditions have received much attention in optimization theory, in recent years; see (1,2,3) for example.

Theorem 1.2 (4, Rockafellar 1970) The maximal monotonicity of Fréchet subdifferential of a lower semicontinuous function is a necessary and sufficient condition for its convexity.

Characterization of generalized convex functions by second-order subdifferentials can be more useful, especially in optimization.

Second-order characterization of convex functions by generalized second-order directional derivatives have studied by some authors.

The upper Dini-directional derivative of g at $x \in X$ in direction $v \in X$ is defined as an element of $\mathbb{R}$ by

$$g'_+(x; v) = \limsup_{t \downarrow 0} t^{-1}(g(x + tv) - g(x)). \quad (1)$$

The second-order upper Dini-directional derivative of g at $x \in X$ in direction $v \in X$ for which $g'_+(x; v)$ is defined by

$$g''_+(x; v) = \limsup_{t \downarrow 0} 2t^{-2}(g(x + tv) - g(x) - tg'_+(x; v)). \quad (2)$$

In the case of an infinite $g'_+(x; v)$, the derivative

$g''_+(x;v)$ will not be considered.

Theorem 1.3 (5, Ginchev and Ivanov 2003) Let $g: X \to \mathbb{R}$ be u.s.c. Then g is convex on X if and only if the following Conditions (C_1) and (C_2) hold for each $x, u \in X$:

(C_1) $g'_+(x;v) + g'_+(x;-v) \geq 0$,

if the expression on the left-hand side has the sense

(C_2) $g'_+(x;v) + g'_+(x;-v) = 0$, implies that $g''_+(x;u) \geq 0$.

Example 1.1 The function $g(x) = -|x|, x \in \mathbb{R}$, satisfies the equality $g''_+(x;v) = 0$ for all $x, v \in \mathbf{R}$. It is continuous, but not convex. Obviously, $g'_+(x;v) + g'_+(x;-v) = -2$.

Example 1.2 The function $g: \mathbb{R} \to \mathbb{R}$ defined as

$$g(x) = \begin{cases} x^2, & \text{if x is rational ;} \\ 0, & \text{otherwise} \end{cases}$$

satisfies conditions (C_1) and (C_2), but g is not convex. This function is not u.s.c.

Some other authors used the second-order Fréchet (Second-order regular subdifferentials) and Mordukhovich (limiting) subdifferentials defined by the coderivative of the subdifferential mappings. See (6,7) for the following definitions and more details.

Let X be a Banach space endowed with a norm $\|.\|$, X^* its dual space, X^{**} its second dual space and $\langle .,. \rangle$ be the dual pairing between X and X^*. For a set-valued mapping $T: X \rightrightarrows Y$ between Banach spaces, we define the effective domain and the graph of T by

$$\mathrm{dom} T = \{x \in X: T(x) \neq \emptyset\}, \quad \mathrm{gph} T = \{(x,y) \in X \times Y: y \in T(x)\}.$$

The sequential Painlevé-Kuratowski upper limit of T at x in the topology of Y is defined by

$\limsup_{x \to \bar{x}} T(x) = \{y \in Y: \exists$ sequences $x_k \to \bar{x}, y_k \to y$ with $y_k \in T(x_k), \ \forall k = 1,2,\dots\}$.

Given $\varepsilon \geq 0$ and $\Omega \subseteq X$, the $\varepsilon - normals$ to Ω at $\bar{x} \in cl(\Omega)$ is defined by

$$\widehat{N}_\varepsilon(\bar{x};\Omega) := \{x^* \in X^*: \limsup_{\substack{\Omega \\ x \to \bar{x}}} \frac{\langle x^*, x - \bar{x} \rangle}{||x - \bar{x}||} \leq \varepsilon\},$$

where the symbol $x \overset{\Omega}{\to} \bar{x}$ means that $x \to \bar{x}$ with $x \in \Omega$. When $\varepsilon = 0$, the set $\widehat{N}_0(\bar{x},\Omega) = \widehat{N}(\bar{x},\Omega)$ is named the prenormal cone or Fréchet normal to Ω at $\bar{x}$.

The limiting or Mordukhovich normal cone to Ω at $\bar{x}$ is

$$N(\bar{x};\Omega) := \limsup_{x \to \bar{x}, \varepsilon \downarrow 0} \widehat{N}_\varepsilon(x;\Omega),$$

where the sequential Painlevé-Kuratowski upper limit is taking in the $weak^*$ topology of X^*. When X is an Asplund Banach space and Ω is closed, we can put $\varepsilon = 0$.

Definition 1.1 (6) *T*he Fréchet or regular coderivative of T at $(\bar{x},\bar{y})$ is

$$\widehat{D}^*T(\bar{x},\bar{y})(y^*) = \{x^* \in X^*: (x^*, -y^*) \in \widehat{N}((\bar{x},\bar{y}), \mathrm{gph}\ T)\} \ \forall y^* \in Y^*.$$

The limiting or Mordukhovich coderivative of T at $(\bar{x},\bar{y})$ is

$$D^*T(\bar{x},\bar{y})(y^*) = \{x^* \in X^*: (x^*, -y^*) \in N((\bar{x},\bar{y}), \mathrm{gph}\ T)\} \ \forall y^* \in Y^*.$$

i.e.,

$$D^*T(\bar{x},\bar{y})(y^*) = \{x^* \in X^*: \exists \varepsilon_k \downarrow 0, (x_k, y_k) \to (\bar{x},\bar{y}), (x_k^*, y_k^*) \overset{w^*}{\to} (x^*, y^*)$$
$$\text{with } (x_k^*, -y_k^*) \in \widehat{N}_{\varepsilon_k}((x_k, y_k), \mathrm{gph}\ T) \text{ as } k \to \infty\}.$$

Definition 1.2 (6) The mixed coderivative of T at $(\bar{x},\bar{y})$ is

$$D_M^*T(\bar{x},\bar{y})(y^*) = \{x^* \in X^*: \exists \varepsilon_k \downarrow 0, (x_k, y_k, y_k^*) \to (\bar{x},\bar{y},y^*), x_k^* \overset{w^*}{\to} x^*$$
$$\text{with } (x_k^*, -y_k^*) \in \widehat{N}_{\varepsilon_k}((x_k, y_k), \mathrm{gph}\ T) \text{ as } k \to \infty\}.$$

Definition 1.3 (6) A single-valued mapping $g: X \to Y$ is said to be strictly differentiable at $\bar{x}$ if there is a linear continuous operator $\nabla g(\bar{x}): X \to Y$ such that

$$\lim_{u,x \to \bar{x}} \frac{g(x) - g(u) - \langle \nabla g(\bar{x}), u - x \rangle}{\| x - u \|} = 0.$$

When g is single-valued and strictly differentiable at $\bar{x}$ or continuously differentiable around $\bar{x}$, with the adjoint operator $\nabla g(\bar{x})^*: Y^* \to X^*$, we have

$$D^*g(\bar{x})(y^*) = \widehat{D}^*g(\bar{x})(y^*) = \{\nabla g(\bar{x})^* y^*\} \text{ for all } y^* \in Y^*.$$

Let $g: X \to \overline{\mathbb{R}} = [-\infty, +\infty]$ be an extended real-valued function. We define

$$\mathrm{dom} g = \{x \in X: |g(x)| < \infty\} \text{ and } \mathrm{epi}(g) = \{(x,\mu) \in (X \times \mathbb{R}): \mu \geq g(x)\}.$$

The Fréchet subdifferential or presubdifferential of g at $\bar{x} \in \mathrm{dom}\ g$ is defined by

$$\hat{\partial} g(\bar{x}) = \{x^* \in X^*: (x^*, -1) \in \widehat{N}((\bar{x}, g(\bar{x})), \mathrm{epi}\ g)\}$$

and the basic or Mordukhovich limiting subdifferential is defined by

$$\partial g(\bar{x}) = \{x^* \in X^*: (x^*, -1) \in N((\bar{x}, f(\bar{x})), \mathrm{epi}\ g)\}.$$

For $\bar{x} \notin \mathrm{dom}\ f$, we put $\hat{\partial} g(\bar{x}) = \partial g(x) = \emptyset$. Also, g is said to be lower regular at $\bar{x}$ if $\hat{\partial} g(\bar{x}) = \partial g(\bar{x})$.

Definition 1.4 (6) Let $g: X \longrightarrow \overline{\mathbb{R}}$ be a function and its value at $\bar{x}$ is finite,

(i) For any $\bar{y} \in \partial g(\bar{x})$, the mapping $\partial^2 g(\bar{x},\bar{y}): X^{**} \rightrightarrows X^*$ with the values

$$\partial^2 g(\bar{x},\bar{y})(v) = (D^*\partial g)(\bar{x},\bar{y})(v), (v \in X^{**}),$$

is called the limiting or Mordukhovich second-order subdifferential of g at $\bar{x}$ relative to $\bar{y}$.

(ii) For any $\bar{y} \in \hat{\partial} g(\bar{x})$, the mapping $\hat{\partial}^2 g(\bar{x},\bar{y}): X^{**} \rightrightarrows X^*$ with the values

$$\hat{\partial}^2 g(\bar{x},\bar{y})(v) = (\widehat{D}^*\hat{\partial} g)(\bar{x},\bar{y})(v), (v \in X^{**}),$$

is called the Fréchet second-order subdifferential of g at $\bar{x}$ relative to $\bar{y}$.

(iii) For any $\bar{y} \in \partial g(\bar{x})$, the mapping $\check{\partial}^2 g(\bar{x},\bar{y}): X^{**} \rightrightarrows X^*$ with the values

$$\check{\partial}^2 g(\bar{x},\bar{y})(v) = (\widehat{D}^*\partial g)(\bar{x},\bar{y})(v), (v \in X^{**}),$$

is called the Combined second-order subdifferential of g at $\bar{x}$ relative to $\bar{y}$.

(iv) For any $\bar{y} \in \partial g(\bar{x})$, the mapping $\partial_M^2 g(\bar{x},\bar{y}): X^{**} \rightrightarrows X^*$ with the values

$$\partial_M^2 g(\bar{x},\bar{y})(v) = (D_M^*\partial g)(\bar{x},\bar{y})(v), (v \in X^{**}),$$

is called the mixed second-order subdifferential of g at $\bar{x}$ relative to $\bar{y}$. When the function g is C^2 around $\bar{x}$ and $v \in X^{**}$, we have

$$\hat{\partial}^2 g(\bar{x})(v) = \partial^2 g(\bar{x})(v) = \partial_M^2 g(\bar{x})(v) = \check{\partial}^2 g(\bar{x})(v) = \{(\nabla^2 g(\bar{x}))^* v\},$$

where $(\nabla^2 g(\bar{x}))^*$ is the adjoint operator of the Hessian $\nabla^2 g(\bar{x})$.

Definition 1.5 (PSD) holds for $g: X \longrightarrow \overline{\mathbb{R}}$, in the Fréchet sense, when $\langle z, v\rangle \geq 0$ for every $v \in X^{**}$ and $z \in \hat{\partial}^2 g(x,y)(v)$ with $(x,y) \in \text{gph}\ \hat{\partial} g$.

When $\langle z, v\rangle > 0$ whenever $v \neq 0$, (PD) holds in the Fréchet sense for g.

Also, (PSD) holds in the limiting sense, when $\langle z, v\rangle \geq 0$ for every $v \in X^{**}$ and $z \in \partial^2 g(x,y)(v)$ with $(x,y) \in \text{gph}\ \partial g$.

When $\langle z, v\rangle > 0$ whenever $v \neq 0$, (PD) holds in the limiting sense for g.

Chieu and Huy considered these cases and extended those results for the class of C^1 functions $g: X \to \mathbb{R}$, where X is a Hilbert space or an Asplund space.

Theorem 1.4 (8, Chieu, Huy 2011) Let $g: X \to R$ be a C^1 function and X be an Asplund space. Then g is convex if the following condition holds:

$$\langle z, v\rangle \geq 0 \text{ for all } v \in X^{**}, z \in \hat{\partial}^2 g(x,y)(v) \text{with } (x,y) \in \text{gph}\ \hat{\partial} g.$$

Convex case

The following questions were raised (8, Chieu, Huy 2011):

1. Is it true that, for any Fréchet differentiable function $g: X \longrightarrow \overline{\mathbb{R}}$, PSD implies convexity?

2. Which class of locally Lipschitz functions does PSD, imply the convexity of the corresponding function?

3. How to extend the characterizations to a general Banach setting?

We proved in (11), that (PSD) holds for any function $g: X \to \overline{\mathbb{R}}$, defined on an arbitrary Banach space, where g is a lower semicontinuous strongly convex.

Theorem 2 .1 (11, Nadi, Yao, Zafarani) Let X be a Banach space and $g: X \longrightarrow \overline{\mathbb{R}}$ be a lower semicontinuous strongly convex function. Then (PSD) holds.

The foregoing result also holds when we replace second-order Fréchet coderivative with mixed second-order coderivative:

Corollary 2.1 (11, Nadi, Yao, Zafarani) Let X be a Banach space and $g: X \to \overline{\mathbb{R}}$ be a lower semicontinuous strongly convex function. Then (PSD) holds in the mixed second-order sense, that is

$\langle z, v\rangle \geq 0$ for any $v \in X^{**}$ and $z \in D_M^* \partial g(\bar{x},\bar{y})(v) = \partial_M^2 g(\bar{x},\bar{y})(v)$.

Also, (PSD) guarantees the convexity of $g: X \to \mathbb{R}$ for some classes of functions. For example, (PSD) guarantees convexity for the class of continuously differentiable functions (C^1 functions) defined on Asplund spaces. Theorem 2.1 of (8, Chieu, Huy, 2011) and (PSD) imply convexity of lower-C^2 functions on $\mathbb{R}^n$ (12, Theorem 4.1). In the following, we illustrate that (PSD) is not a sufficient condition for convexity, when the function is differentiable at a point.

Example 2.1 (11, Nadi, Yao, Zafarani) Consider the function $g: \mathbb{R} \longrightarrow \mathbb{R}$ as follows:

$$g(x) = \begin{cases} \frac{1}{n^2}, & x \in]0,1], \frac{1}{n+1} < x \leq \frac{1}{n}, n \in \mathbb{N} \\ 0, & x \leq 0 \\ 2, & x > 1, \end{cases}$$

It is clear that g is differentiable at zero, but is not convex. Also, by an easy calculation, we can show that (PSD) holds for g.

In the following theorem, we showed that (PD) guarantees the convexity of $g: X \to \mathbb{R}$ when g is differentiable on X and $\hat{\partial} g$ is non-empty on X.

We proved it for $X = \mathbb{R}$ and afterwards for Banach spaces.

Theorem 2.2 (11, Nadi, Yao, Zafarani) Let $g: \mathbb{R} \to \mathbb{R}$ be a differentiable function and (PSD) holds in the Fréchet sense and $\hat{\partial} g'$ be nonempty on $\mathbb{R}$. Then g is convex.

We concluded the following corollary for g on Banach spaces by using the above argument. For arbitrary $a, v \in X$, $g: X \to \mathbb{R}$ and $s \in \mathbb{R}$, define $g_{a,v}(s) = g(a + sv)$. We know that g is convex on X if and only if $g_{a,v}$ is convex on $\mathbb{R}$ for any $a, v \in X$; See,

(13) for more details.

Corollary 2.2 Let $g: X \to \mathbb{R}$ be a differentiable function on X, $(\widehat{D}^*\nabla g)(x)(v)$ be non empty for any $x, v \in X$ and $\langle z, v\rangle \geq 0$ for every $x, v \in X$ and $z \in (\widehat{D}^*\nabla g)(x)(v)$. Then g is convex.

Corollary 2.3 Let $g: X \to \mathbb{R}$ be a differentiable function on X, $(\widehat{D}^*\nabla g)(x)(v)$ be non empty for any $x, v \in X$ and $\langle z, v\rangle \geq 0$ for every $x, v \in X$ and $z \in (D^*\nabla g)(x)(v)$ or $z \in (D_M^*\nabla g)(x)(v)$. Then g is convex.

We concluded that (PSD) and differentiability, imply the continuity of differential mapping.

Corollary 2.4 Let $g: X \to \mathbb{R}$ be differentiable on X and $\widehat{D}^*(\nabla g)(x)(v)$ be non-empty for any $x, v \in X$. If (PSD) holds in the Fréchet sense, then g is of class C^1.

Theorem 2.3 (14, Nadi, Zafarani) Let $g: X \to \mathbb{R}$ be a locally Lipschitz approximately convex function and X be an Asplund Banach space. Then g is convex, if (PSD) holds in the regular sense:

$$\langle z, v\rangle \geq 0, \forall v \in X \text{ and } z \in \hat{\partial}^2 g(x, y)(v) \text{ with } (x, y) \in \operatorname{gph} \hat{\partial} g$$

Theorem 2.4 (15, Nadi, Zafarani) Let $g: X \to \mathbb{R}$ be a lower semicontinuous approximately convex function, X be an Asplund space and (PSD) holds. Then g is convex.

For $X = \mathbb{R}^n$, two classes of lower-C^1 functions and lower semicontinuous approximately convex functions are the same (16, Daniilidis, Georgiev, 2004). The class of lower- C^1 functions was initially introduced by Spingarn (1981) and afterwards, the smaller class of lower- C^k functions was introduced in 1982 by Rockafellar. The function $g: \mathbb{R}^n \to \mathbb{R}$ is said to be lower-C^k for ($k \in \mathbb{N}$) if, for each $\bar{x} \in \mathbb{R}^n$, there exists a neighbourhood of $\bar{x}$ as V such that g has the representation

$$g(x) = \max_{s \in S} g_s(x),$$

where the index set S is compact, the functions g_s are of class C^k on V, and $g_s(x)$ and all of the partial derivatives of the functions g_s of order k are jointly continuous on (s, x).

Definition 2.4 We say that a locally Lipschitz function $g: X \to \mathbb{R}$ is directionally Clarke regular (d-regular) at z if, for every $v \in X$, the Clarke directional derivative of g at z in the direction v coincides with $d^-g(z, v)$, where

$$d^-g(z, v) := \liminf_{t \to 0^+} \frac{g(z + tv) - g(z)}{t}.$$

Remark 2.1 *The above Theorem is the lower-C^1 version of Theorem 4.1 (12, Chieu, Lee, Mordukhovich, Nghia, 2016). We know that in finite dimensional spaces, a lower-C^1 function g is approximately convex and locally Lipschitz* (16, Daniilidis, Georgiev, 2004). *Also, we answer question 2 posed in* (8, Chieu, Huy, 2011) *by this result. By a similar proof, we concluded that* (PSD) *holds for d-regular and semismooth functions defined on $X = \mathbb{R}^n$.*

We show by the following example that in the foregoing theorem, approximate convexity is essential. It means that, the class which was asked in question 2 of (8, Chieu, Huy, 2011) is approximately convex functions (the class of lower-C^1 functions when the space is finite-dimensional). We show that the following function which is Lipschitz and was given in (8, Chieu, Huy, 2011), theorem 4.2, is not approximately convex (lower-C^1).

Example 2.2 (15, Nadi, Zafarani) For all $x \in \mathbb{R}$; define

$g(x) = \int_0^x \chi_E(t)dt$, where E is a subset of $\mathbb{R}$ which is measurable and the intersection of both E and its complement with each nonempty open interval of $\mathbb{R}$ has positive Lebesgue measure. The function g is Lipschitz, and (PSD) holds but it is not convex.

Corollary 2.5 (15, Nadi, Zafarani) Let $g: X \to \mathbb{R}$ be a lower semicontinuous approximately convex function and X be a Hilbert space. Then the function g is strongly convex (with modulus $\kappa > 0$) if and only if

$$\langle z, v\rangle \geq \kappa \| v \|^2, \forall v \in X \text{ and } z \in \hat{\partial}^2 g(x, y)(v) \text{ with } (x, y) \in \operatorname{gph} \hat{\partial} g. \quad (3)$$

Convex mappings

We assume that the spaces X and Y are Banach spaces and X is reflexive, $K \subseteq Y$ is a closed convex and pointed cone ($K \cap -K = 0$) and K^* is the positive dual cone of K; that is $K^* = \{y^* \in Y^*: y^*(k) \geq 0, \text{for all } k \in K\}$.

Definition 2.5 Let $g: X \longrightarrow Y$ be a vector valued function. g is K-convex on X if for any $x_1, x_2 \in X$ and $\lambda \in [0,1]$,

$$g(\lambda x_1 + (1 - \lambda)x_2) \leq_K \lambda g(x_1) + (1 - \lambda)g(x_2).$$

Theorem 2.5 (15, Nadi, Zafarani) Let $g: X \to Y$ be a C^1 mapping. If (PSD) holds in the limiting sense, then g is K-convex.

Also, the converse holds for twice continuously differentiable case:

Theorem 2.6 (15, Nadi, Zafarani) Let Y and X be Banach spaces and $g: X \to Y$ be a C^2 mapping. Then *(PSD)* holds if and only if g is K-convex.

The following example illustrates the foregoing theorem.

Example 2.3 (15, Nadi, Zafarani) Consider $g: \mathbb{R}^2 \to \mathbb{R}^2$ with $g(z) = g(z_1, z_2) = (z_1^2 + z_2^2, z_1^2 + z_1)$ and $C = \{(z_1, z_2) \in \mathbb{R}^2 : z_1, z_2 \geq 0 \text{ and } z_2 \leq z_1\}$. Then g is a C-convex mapping, twice continuously differentiable and *(PSD)* holds because for every $z = (z_1, z_2)$ and $v = (v_1, v_2) \in \mathbb{R}^2$ we have:

$$\nabla^2 g(z)(v) = 2v_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + 2v_2 \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + 2v_1 \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} v_1 + v_2 \\ v_1 \end{pmatrix}.$$

But this means that

$$\nabla^2 g(z)(v) = \begin{pmatrix} 2v_1^2 + 2v_2^2 \\ 2v_1^2 \end{pmatrix} \in C.$$

Quasi convex functions

Characterization of pseudoconvexity and quasiconvexity by their second-order subdifferentials and their applications are studied in the literature. For twice differentiable pseudoconvex and quasiconvex functions $g: C \subseteq \mathbb{R}^n \to \mathbb{R}$, where ∇g is locally Lipschitz, the second-order characterization has been extended by (13, Crouziex and Ferland, 1996).

Given a normed linear space X and a convex subset K of X, a function $g: K \to \mathbb{R}$ is called

(i) quasiconvex on K, where for every $x, y \in K$ and $t \in]0,1[$,

$$g(x + t(y - x)) \leq \max\{g(x), g(y)\},$$

or equivalently where its level sets $(Lev_\alpha g)$ are convex, i.e.,

for every $\alpha \in \mathbb{R}$, $\text{Lev}_\alpha g =: \{x \in K : g(x) \leq \alpha\}$ is convex,

(ii) pseudoconvex on K if for every $x, y \in K$, $x \neq y$ and $x^* \in \hat{\partial} g(x)$,

$$\langle x^*, y - x \rangle \geq 0 \Longrightarrow g(y) \geq g(x).$$

Definition 2.6 [(14), Nadi, Zafarani] Let X be a Banach space and $F: X \rightrightarrows X^*$ be a set-valued mapping and, for every $\bar{x} \in X$ and $v \in X^{**}$, define:

$$\widehat{D}_+ F(\bar{x}, v) := \sup\{\langle z, v \rangle : z \in \widehat{D}^* F(x, y)(v), x \to \bar{x}, y \to \bar{y}, y \in F(x)\}.$$

$$\widehat{D}_- F(\bar{x}, v) := \inf\{\langle z, v \rangle : z \in \widehat{D}^* F(x, y)(v), x \to \bar{x}, y \to \bar{y}, y \in F(x)\}.$$

Here we mention a result for the quasiconvex case:

Theorem 2.7 (14, Nadi, Zafarani) Let $g: X \to R$ be a locally Lipschitz function. If the following assertions hold for every $\bar{x}, u \in X$:

(i) $\varphi_u(\bar{x}) = \inf\{\langle y, v \rangle : y \in \partial_c g(\bar{x})\} = 0$ implies that $\widehat{D}_+ \partial_c g(\bar{x}, v) \geqslant 0$;

(ii) $\varphi_u(\bar{x}) = 0$, $\widehat{D}_+ \partial_c g(\bar{x}, v) \geq 0, \widehat{D}_- \partial_c g(\bar{x}, v) \leq 0$ and $\langle y_{\bar{t}}, v \rangle > 0$ (for some $\bar{t} < 0$ and $y_{\bar{t}} \in \partial_c g(\bar{x} + \bar{t}v)$), implies that there exists $\hat{t} > 0$ such that $\langle y_t, v \rangle \geq 0$ for every $t \in [0, \hat{t}]$ and $y_t \in \partial_c g(\bar{x} + tv)$.

(iii) g is approximately convex.

Then g is quasiconvex.

Example 2.4 (14, Nadi, Zafarani) Consider the function $g: S = \{z : \| z \| < \frac{1}{2}\} \subseteq \mathbb{R}^2 \to \mathbb{R}$ defined as

$$g(z_1, z_2) = f(z) = -\| z \|^2 + \| z \|.$$

It is easy to see that g is continuously differentiable on $S \backslash \{(0,0)\}$. Also, the Clarke subdifferential at $(0,0)$ is

For every $0 \neq v \in \mathbb{R}^2$, we have $\inf\{\langle y, v \rangle : y \in \partial_c g((0,0))\} < 0$, because the closed unit ball is a balanced subset of $\mathbb{R}^2$. Therefore, clearly (i) holds.

For (ii), assume that $v \neq (0,0)$ is arbitrary. Now, an easy calculation shows that

$$\langle \nabla g(tv), v \rangle = (v_1^2 + v_2^2)(-2t + \frac{1}{\sqrt{v_1^2 + v_2^2}}) \geq 0,$$

for every $t \in [0, \hat{t}]$ with $\hat{t} := 2(v_1^2 + v_2^2)^{-\frac{1}{2}}$, which means that (ii) holds.

Pseudo convex functions

A similar result holds for the pseudoconvex case:

Theorem 2.8 (14, Nadi, Zafarani) Let $g: X \to \mathbb{R}$ be a locally Lipschitz function. Suppose that the following conditions hold for every $\bar{x}, v \in X$:

(i) $\varphi_v(\bar{x}) = \inf\{\langle y, v \rangle : y \in \partial_c g(\bar{x})\} = 0$ implies that $\widehat{D}_+ \partial_c g(\bar{x}, v) \geqslant 0$;

(ii) $\varphi_v(\bar{x}) = 0$, $\widehat{D}_+ \partial_c g(\bar{x}, v) \geq 0$ and $\widehat{D}_- \partial_c g(\bar{x}, v) \leq 0$, implies that: there exists $\hat{t} > 0$ such that $\langle y_t, u \rangle \geq 0$ for every $t \in [0, \hat{t}]$ and $y_t \in \partial_c f(\bar{x} + tu)$.

(iii) g is approximately convex.

Then g is pseudoconvex.

For the case of strictly pseudoconvex functions, the following result is interesting:

Theorem 2.9 (17, Khanh Phat 2020) Let $g: \mathbb{R}^n \to \mathbb{R}$ be a $C^{1,1}$-smooth function satisfying

$$x \in \mathbb{R}^n, v \in \mathbb{R}^n \backslash \{0\}, \langle \nabla g(x), v \rangle = 0, \Rightarrow \langle z, v \rangle > 0, \text{for all } z \in \partial^2 g(x)(v).$$

Then g is a strictly pseudoconvex function.

Also, for the case of strictly quasiconvex functions, the following result is interesting:

Theorem 2.10 (17, Khanh Phat 2020) Let $g: \mathbb{R}^n \to \mathbb{R}$ be a $C^{1,1}$-smooth function satisfying

$$x \in \mathbb{R}^n, v \in \mathbb{R}^n \backslash \{0\}, \langle \nabla g(x), v \rangle = 0, \Rightarrow \langle z, v \rangle > 0, \text{for all } z \in \hat{\partial}^2 g(x)(v) \cup -\hat{\partial}^2 g(x)(-v).$$

Then g is a strictly quasicoconvex function.

Invex function

In recent years, the mathematical landscape has witnessed numerous extensions and generalizations of

classical convexity, particularly through the invex functions by Hanson in 1981(18). This pivotal advancement sparked a wave that has substantially enriched the applications of invexity within nonlinear optimization and related fields. Notably, Hanson demonstrated that the Kuhn-Tucker conditions, which are fundamental in optimization theory, serve as sufficient criteria for optimality when dealing with invex functions. This revelation has prompted further exploration into the properties and applications of generalized convexity.

Definition 2.7 A set C is said to be invex with respect to $\eta: X \times X \to X$, when for any $x, y \in C$ and $0 \le \lambda \le 1$,

$$y + \lambda\eta(x, y) \in C.$$

Definition 2.8 A vector valued function $\eta: X \times X \to X$ is said to be skew, if

$$\eta(x, y) + \eta(y, x) = 0, \text{ for any } x, y \in X.$$

The following assumptions are frequently used in the literature:

ASSUMPTION **A**: Let C be an invex set with respect to η, and $g: C \to \mathbb{R}$. Then

$$g\big(y + \eta(x, y)\big) \le g(x) \text{ for any } x, y \in C.$$

ASSUMPTION **C**: Let $\eta: X \times X \to X$. Then, for any $x, y \in X$ and for any $\delta \in [0,1]$,

$$\eta(y, y + \delta\eta(x, y)) = -\delta\eta(x, y),$$
$$\eta(x, y + \delta\eta(x, y)) = (1 - \delta)\eta(x, y)$$

Definition 2.9 A differentiable function $g: X \to \mathbb{R}$ is said to be invex with respect to η, if for any $x, y \in C$, one has

$$\langle \nabla g(y), \eta(x, y) \rangle \le g(x) - g(y).$$

Definition 2.10 A locally Lipschitz function $g: C \subseteq X \to \mathbb{R}$ is called invex with respect to η, if for any $x, y \in C$ and any $\xi \in \partial g(x)$, one has

$$\langle \xi, \eta(x, y) \rangle \le g(x) - g(y).$$

Remark 2.2 Note that, in the above definitions by letting $\eta(x, y) = x - y$, we reduce to the convex case. Indeed, invex functions reduce to convex functions, and invex sets, to convex sets.

Proposition 2.1 (19, Nadi, Zafarani) Let $g: \mathbb{R}^n \to \mathbb{R}$ be an invex function with respect to a skew $\eta: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$, be twice differentiable at $x \in \mathbb{R}^n$ and $\eta(., x)$ be differentiable at x. Then $\langle \eta_x(x, x)v, D^2 g(x)v \rangle \ge 0$ for any $v \in \mathbb{R}^n$.

Theorem 2.11 (19, Nadi, Zafarani) Suppose that $g: \mathbb{R}^n \to \mathbb{R}$ is $C^{1,1}$, invex function with respect to a skew $\eta: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$, where η is differentiable in the first argument at x and continuous. Then $\langle \eta_x(x, x)v, x^* v \rangle \ge 0$, for any $v \in \mathbb{R}^n$ and $x^* \in \partial g'(x)$.

Remark 2.3 The above results are the natural extensions of the convex case. In fact, by replacing $\eta(x, y)$ with $x - y$, we have the classical form of Hessian.

In the following example, we show that sometimes characterizing the invexity of a function by the second-order condition is easier than using the first order condition.

Example 2.4 (19, Nadi, Zafarani) Consider the following $C^{1,1}$ function $g: \mathbb{R} \to \mathbb{R}$,

$$g(x) = \begin{cases} -x^2 + x, & x \le 0 \\ x^2 + x, & x > 0. \end{cases}$$

Consider, also $\eta(x, y) = x^3 - y^3$. An easy calculation implies that

$$\partial g'(x) = \begin{cases} -2, & x < 0 \\ \{[-2,2]\}, & x = 0 \\ 2, & x > 0, \end{cases}$$

which means that $\langle \eta_x(x, x)v, x^* v \rangle = 3x^2 x^* < 0$, by letting $x = -1$ and any arbitrary $v \in \mathbb{R}$.

Theorem 2.12 (19, Nadi, Zafarani) Let $g: \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function, g and η satisfy Assumptions A and C, $\eta(., y)$ be onto for any $y \in \mathbb{R}^n$ and skew. If $\langle \eta_x(x, x)v, \nabla^2 g(x)v \rangle \ge 0$, for any $x, v \in \mathbb{R}^n$, then g is invex with respect to η.

Optimization

Consider the nonlinear programming (NLP) as follows, with C^1 data ($f, g_i: X \to \mathbb{R}$ for $1 \le i \le n$ are continuously differentiable):

$$\text{minimize } f(x) \text{ subject to}$$
$$g_i(x) = 0, \text{ for } i \in E \text{ and } g_i(x) \le 0 \text{ for } i \in I,$$

Where for the constrains, $E := \{1, \dots, n_1\}$ and $I := \{n_1 + 1, \dots, n_1 + n_2\}$ are finite index sets and $n := n_1 + n_2$. The point x is called a feasible point of the foregoing (NLP) problem if

$$x \in \Gamma := \{y \in X: g_i(y) = 0 \text{ for } i \in E \text{ and } g_i(y) \le 0 \text{ for } i \in I\}.$$

Also, the classical Lagrange function is:

$$L(x, \lambda) := f(x) + \langle \lambda, g \rangle(x), \text{ for } x \in X \text{ and } \lambda \in \mathbb{R}^l.$$

When $\bar{x}$ is a solution for (NLP), the first order necessary condition is that there exist λ_i for $i = 1, \dots n$, which are said to be the Lagrange multipliers, with

$\lambda_i g_i(\bar{x}) = 0$ (for i $= 1, \dots, n$) and $\nabla f(\bar{x}) + \sum_{i=1}^{n} \lambda_i \nabla g_i(\bar{x}) = 0$

and the standard second-order sufficient condition (SSOSC) is that there exists $k > 0$ such that

$\nabla_x^2 L(\bar{x}, \bar{\lambda})(v, v) \geq k \| v \|^2$ with $\bar{\lambda} = (\bar{\lambda}_1, \dots, \bar{\lambda}_l)$ (4)

for all $v \in X$, with $\langle \nabla g_i(\bar{x}), v \rangle = 0$ for $i \in E \cup I^+(\bar{\lambda})$ and $\langle \nabla g_i(\bar{x}), v \rangle \leq 0$ for $i \in I^0(\bar{\lambda})$, where

$I^+(\bar{\lambda}) = \{i \in I: \bar{\lambda}_i > 0\}$ and $I^0(\bar{\lambda}) = \{i \in I: \lambda_i = 0\}$.

Also, when X is finite-dimensional, we can change the inequality (4) as follows:

$\nabla_x^2 L(\bar{x}, \bar{\lambda})(v, v) > 0$ with $\bar{\lambda} = (\bar{\lambda}_1, \dots, \bar{\lambda}_l)$. (5)

Indeed, when X is finite-dimensional, the second-order sufficient condition implies optimality of $\bar{x}$ (the critical point) for Lagrange multipliers $\bar{\lambda}$, when $\nabla_x^2 L(\bar{x}, \bar{\lambda})$ is positive definite on the critical cone of (NLP) at $(\bar{x}, \bar{\lambda})$; it means that

$C(\bar{x}) =: \{v: \langle \nabla g_i(\bar{x}), v \rangle = 0$ for $I^+(\bar{\lambda}) \cup E$ and $\langle \nabla g_i(\bar{x}), v \rangle \leq 0$ for $i \in I^0(\bar{\lambda})\}$.

We continue with the following second-order sufficient condition for optimality of a KKT-point of (NLP). In the following, X is a reflexive Banach space.

Theorem 3.1 (20, Nadi, Zafarani) *(*Point-based sufficient condition) Assume the foregoing stated (NLP) problem with $\bar{z} \in \Gamma$ a KKT-point of (NLP) and Lagrange multipliers $\bar{\lambda}$. Suppose that the second-order condition holds:

$\widehat{D}_- \nabla L(\bar{z}, \bar{\lambda}, v) > 0$ for all $v \in C(\bar{z}) \backslash \{0\}$. (6)

Then $\bar{z}$ is a strictly local minimum for (NLP).

In condition (6), we use the coderivative of the differential mapping and it is more efficient than the other similar second-order optimality conditions which have been introduced by the various kinds of generalized second-order directional derivatives. As illustrated by the following example, the following theorem due to (21, Ben-Tal and Zowe) and its constrained version can not be used for the C^1 data case.

Let $g: \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\bar{x}$. We denote by $g''(\bar{x}, v)$, the second-order directional derivative of g at x in direction $v \in \mathbb{R}^n$ which is defined as an element of $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$; that is

$$g''(\bar{x}, v) := \lim_{t \to +\infty} \frac{2}{t^2} (g(\bar{x} + tv) - g(\bar{x}) - t \nabla g(\bar{x}) v).$$

Theorem 3.2 (21, Ben-Tal and Zowe) Suppose that $g \in C^{1,1}(\mathbb{R}^n)$, $\nabla g(\bar{x}) = 0$ and $g''(\bar{x}, v) > 0$ for all $v \in \mathbb{R}^n \backslash \{0\}$. Then $\bar{x}$ is a strict local minimizer of g.

Example 3.1 (20, Nadi, Zafarani) Consider the function $g: \mathbb{R}^2 \to \mathbb{R}$ defined as

$g(z_1, z_2) := (\max(0, z_2 - 2z_1^{\frac{4}{3}}))^{\frac{3}{2}} + (\max(0, z_1^{\frac{4}{3}} - z_2))^{\frac{3}{2}}$.

One can show that $g''(\bar{x}, v) > 0$ for $\bar{x} = (0,0)$ and all nonzero direction v, but $\bar{x}$ is not a strict local minimum of g since $g(z) = 0$ for all z between the curves $z_2 = z_1^{\frac{4}{3}}$ and $z_2 = 2z_1^{\frac{4}{3}}$.

Letting (z_k) be an arbitrary sequence which converges to zero, we have $(z_k, \frac{3}{2} z_k^{\frac{4}{3}}) \to (0,0)$. It is trivial that $\nabla g(z_k, \frac{3}{2} z_k^{\frac{4}{3}}) = 0$ because g is equal to zero in a neighbourhood of $(z_k, \frac{3}{2} z_k^{\frac{4}{3}})$.

Now, it is easy to see that $0 \in \widehat{D}^* \nabla g(z_k, \frac{3}{2} z_k^{\frac{4}{3}})(v)$ for all $v \in \mathbb{R}^2$, which implies that $\widehat{D}_- \nabla g(\bar{x}, v) \leq 0$. This means that condition (6) in the above theorem does not hold.

Pseudoconvexity of the cost function in addition to the quasiconvexity of constrained functions implies the optimality of the point that satisfies the Karush Kahn-Tucker conditions. More precisely, if the cost function or one of the active constrained functions with positive Lagrange multipliers is pseudoconvex and the rest are quasiconvex, then the Lagrange function is pseudoconvex. Booth of quasiconvexity and pseudoconvexity of constrained functions imply the convexity of the feasible set and optimality of a KKT-point will be obtained. But we know that the convexity of the feasible set is not necessary in (NLP). As mentioned below, the pseudoconvexity of the cost function and quasiconvexity of constraint functions at a KKT-point is sufficient for its optimality.

Theorem 3.3 (Mangasarian) Let the set constraint be open. The functions f and g_i for $i = 1, \dots, n_1$ are the functions defined on X and $\bar{x}$ is a feasible point. Assume that f is pseudoconvex at $\bar{x}$, f and g_i for $i \in I(\bar{x})$ are differentiable at $\bar{x}$, and g_i for $i \in I(\bar{x})$ are quasiconvex at $\bar{x}$. If there exist Lagrange nonnegative multipliers $\lambda_1, \dots, \lambda_{l_1}$ with $\lambda_i g_i(\bar{x}) = 0$ for $i = 1, \dots, n_1$ and $\nabla L(\bar{x}) = 0$ where $L = f + \sum_{i=1}^{l_1} \lambda_i g_i$, then $\bar{x}$ is a global minimizer of (NLP).

The following example shows that the pseudoconvexity at a point for the cost function in the foregoing Theorem is more than what is required.

Example 3.2 (20, Nadi, Zafarani) Consider the following (NLP) with $C^{1,1}$ data:

$$\text{minimize} \quad f(x) := -\frac{1}{2}z_1|z_1| + z_1z_2 - z_1 + z_2 \ \text{ for } z = (z_1, z_2)$$
$$\text{subject to} \quad g_1(z) := z_1^2 + z_1 - z_2 \le 0,$$
$$g_2(z) := z_1 + z_2 - 1 \le 0, \quad g_3(z) := -2z_1 + z_2 \le 0.$$

The Lagrangian function for $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ is

$$L(z, \lambda) = -\frac{1}{2}z_1|z_1| + z_1z_2 + \lambda_1(z_1^2 + z_2 - 1) + \lambda_2(z_1 + z_2 - 2) - \lambda_3 z_1.$$

Now, we can show that $\bar{z} = (0,0)$ is a KKT-point for (NLP) with Lagrange multipliers $\bar{\lambda} = (1,0,0)$. Also, for all $z \in \mathbb{R}^2$ we have

$$\nabla L(z) = (-|z_1| + z_2 + 2z_1, z_1).$$

For $z_1 > 0$ and $v = (v_1, v_2) \in \mathbb{R}^2$,

$$\nabla^2 L(z)(v) = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_1 + v_2 \\ v_1 \end{pmatrix}$$

and for $z_1 < 0$ and $v = (v_1, v_2) \in \mathbb{R}^2$ we have

$$\nabla^2 L(z)(v) = \begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_1 + v_2 \\ v_1 \end{pmatrix}.$$

Thus, for $z_1 > 0$ and $p \in \widehat{D}^*(\nabla L)(z)(v) = \nabla^2 L(z)(v)$ we derive

$$\langle p, v\rangle = v_1^2 + 2v_2v_1.$$

Also, for $z_1 < 0$ and $p \in \widehat{D}^*(\nabla L)(z)(v) = \nabla^2 L(z)(v)$ we deduce

$$\langle p, v\rangle = 3v_1^2 + 2v_1v_2.$$

On the other hand, the set of active indexes in $\bar{z}$ is $I(\bar{z}) = \{1,3\}$ and $I^+(\bar{\lambda}) = \{1\}$ and $I^0(\bar{\lambda}) = \{3\}$. Therefore, by an easy calculation, we conclude that the critical direction cone at $\bar{x}$ is

$$C(\bar{z}) = \{v: \langle \nabla g_1(\bar{z}), v\rangle = 0 \text{ and } \langle \nabla g_3(\bar{z}), v\rangle \le 0\}$$
$$= \{(v_1, v_2): v_1 = v_2 \text{ and } -2v_1 + v_2 \le 0\}$$
$$= \{(v_1, v_2): v_1 = v_2 \text{ and } v_1, v_2 \ge 0\}.$$

This means that $\langle p, v\rangle > 0$ for all $p \in \widehat{D}^*(\nabla L)(z)(v)$ with $z \ne 0$ and $v \in C(\bar{z})\backslash\{0\}$. It is not difficult to see that $\widehat{D}^*(\nabla L)(z)(v) = \emptyset$ for all $z = (0, z_2)$. Therefore, the second-order sufficient condition $\widehat{D}_-(\nabla L)(z, v) > 0$ holds for all $v \in C(\bar{z})\backslash\{0\}$ by our Theorem. Moreover, it is easy to see that the cost function f is strictly pseudoconvex in direction $v = (1,1) \in C(\bar{z})$, because for all $t > 0$:

$$f(\bar{z} + tv) = f(t, t) = -\frac{1}{2}t^2 + t^2 - t + t = \frac{1}{2}t^2 > f(\bar{z}) = 0.$$

But for $u = (1,0) \notin C(\bar{z})$ and all $t > 0$:

$$f(\bar{z} + tu) = f(t, 0) = -\frac{1}{2}t^2 - t < f(\bar{z}) = 0$$

This means that f is not pseudoconvex at $\bar{z}$ in the direction u. Therefore, f is not pseudoconvex at $\bar{z}$, but $\bar{z}$ is a minimizer for (NLP).

Instead of pseudoconvexity and quasiconvexity at a point, we use the pseudoconvexity and quasiconvexity at a point in a direction and present the following extension of Mangasarian's theorem in the case of local solution.

Theorem 3.4 (20, Nadi, Zafarani) Let the set constraint be open. The functions f and g_i for $i = 1, \dots, n_1$ are defined on X and $\bar{z}$ is a feasible point. Suppose that there exist Lagrange nonnegative multipliers $\lambda_1, \dots, \lambda_{l_1}$ with $\lambda_i g_i(\bar{z}) = 0$ for $i = 1, \dots, n_1$ and $\nabla L(\bar{z}) = 0$ where $L = f + \sum_{i=1}^{l_1} \lambda_i g_i$. If f and g_i for $i \in I(z)$ are differentiable at z, f is pseudoconvex at $\bar{z}$ in all critical directions $v \in C(\bar{z})$ and g_i for $i \in I(\bar{z})$ are quasiconvex at $\bar{x}$ in all critical directions $v \in C(\bar{z})$, then $\bar{z}$ is a local minimizer of (NLP).

Now, we give some applications in tilt-stability theory, as an application of our results in classical optimization.

Proposition 3.1 (14, Nadi, Zafarani) Let (PSD) hold for $g: X \to \mathbb{R}$ that is a differentiable function and $\widehat{D}^*(\nabla g)(z)(v)$ be non-empty for any $z, v \in X$. If $\nabla g(\bar{z}) = 0$, then $\bar{z}$ is a global minimizer of g.

Definition 3.1 (22, Tilt Stability, Poliquin-Rockafellar 1998) Given $g: X \to \overline{\mathbb{R}}$, a point $\bar{z} \in \text{dom} f$ is a tilt-stable local minimizer of g, if there is $\gamma > 0$ such that the mapping

$$M_\gamma: z^* \longmapsto \operatorname{argmin}\{f(z) - \langle z^*, z\rangle: z \in B_\gamma(\bar{z})\}$$

is a single-valued ma*pping* and Lipschitz continuous on some vicinity of $0 \in X^*$ with $M_\gamma(0) = \bar{z}$.

Proposition 3.2 (14, Nadi, Zafarani) Let $g: X \longrightarrow \mathbb{R}$ be a strongly convex lower semicontinuous function and X be a Banach space. Then the following conditions hold:

(i) If $\bar{z}$ is a global minimizer for g, then it is the tilt-stable local minimum of g.

(ii) The point $\bar{z}$ is a local minimizer for g when X is an Asplund space. Also, there exist numbers $r \in (0, \frac{1}{\kappa})$ and $\varepsilon > 0$ such that

$$g(x) \ge g(\bar{x}) + \langle \bar{y}, z - \bar{y}\rangle - \frac{r}{2\kappa}\| z - \bar{z} \|^2 \quad \text{whenever } z \in B_\varepsilon(\bar{z}).$$

Proposition 3.3 (20, Nadi, Zafarani) Let $g: \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function which satisfies Assumption **A** with respect to some η and $\nabla g(\bar{z}) = 0$. Moreover, suppose that one of the following holds:

(i) $\langle \eta_z(z, z)v, \nabla^2 g(z)(z)v\rangle \ge 0$, for any $z, v \in \mathbb{R}^n$, where η is skew and satisfies Assumption C and $\eta(., y)$ is onto for any $y \in \mathbb{R}^n$.

(ii) $\langle \eta(y, \bar{z}), \nabla^2 g(\bar{z})\eta(y, \bar{z})\rangle \ge 0$, for any $y \in \mathbb{R}^n$.

Then $\bar{z}$ is a local minimizer of g.

Consider the following constrained optimization problem:

$\min\ g_0(z)$ subject to $g_i(z) \le 0\ \ (i = 1,\dots,m)$,(7)

which $g_0, g_1, \dots, g_m$ are twice differentiable functions defined on $\mathbb{R}^n$.

Let $g(z) = (g_0(z), \dots, g_n(z))$. We know that the existence of a vector $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$ which satisfies the following conditions, (Kuhn-Tucker conditions) is necessary for $\bar{z}$ to solve this problem:

$\nabla g(\bar{z}) + \langle \lambda, \nabla g(\bar{z}) \rangle = 0$ (8)

$\langle \lambda, g(\bar{z}) \rangle = 0$ (9)

$\lambda_1, \dots, \lambda_n \ge 0$. (10)

Hanson 1981 showed that the Kuhn-Tucker conditions are also sufficient for $\bar{z}$ to be a solution of (4), when each g_i is invex with respect to the same η. Indeed, only the invexity in a neighbourhood of $\bar{z}$ for each g_i guarantees that the foregoing conditions are sufficien (Craven 1982).

Now, we give some second-order sufficient conditions for constrained optimization problems, by using our results.

Proposition 3.4 (20, Nadi, Zafarani) Suppose we have the constrained optimization problem (4). If the Kuhn-Tucker conditions hold in $\bar{z}$, each g_i satisfies Assumption **A**, and one of the following second-order conditions holds (with respect to the same η):

(i) $\langle \eta_z(z,z)v, \nabla^2 g_i(z)v \rangle \ge 0$, for any $z, v \in \mathbb{R}^n$, where η is skew and satisfies Assumption **C** and $\eta(.,y)$ is onto for any $y \in \mathbb{R}^n$,

(ii) $\langle \eta(y,\bar{z}), \nabla^2 g_i(z)\eta(y,\bar{z}) \rangle \ge 0$ for any $y \in \mathbb{R}^n$,

then $\bar{z}$ is a solution for the constrained optimization problem (4).

Acknowledgement

This research was in part supported by a grant from IPM (No.1402260047).

References

1. Khanh PD, Mordukhovich BS, Phat VT. Variational convexity of functions and variational sufficiency in optimization. SIAM Journal on Optimization. 2023;33:1121-1158.
2. Benku M, Gfrerer H, Ye JJ, Zhang J, Zhou J. Second-order optimality conditions for general nonconvex optimization problems and variational analysis of disjunctive systems. SIAM Journal on Optimization. 2023;33(4):2625-2653.
3. Gfrerer H. On Second-Order variational analysis of variational convexity of prox-regular functions. Set-Valued Variational Analysis. 2025;33(8).
4. Rockafellar RT. On the maximal monotonicity of subdifferential mappings. Pacific Journal of Mathematics. 1970;33:209-216.
5. Ginchev I, Ivanov VI. Second-order characterizations of convex and pseudoconvex functions. Journal of Applied Analysis. 2003;9:261-273.
6. Mordukhovich BS. Variational analysis and generalized differential I. II: Springer, NewYork, 2006.
7. Mordukhovich BS. Second-order variational analysis in optimization, variational Stability, and Control: Theory, Algorithms, Applications. Series in Operations Research and Financial Engineering, Springer, Cham, 2024.
8. Chieu NH, Huy NQ. Second-order subdifferentials and convexity of real-valued functions. Nonlinear Anallysis. 2011;74:154-160.
9. Ngai HV, Luc DT, Thera M. Approximate convex functions. Journal of Nonlinear and Convex Analysis. 2000;1(2):155-176.
10. Aussel D, Daniilidis A, Thibault L. Subsmooth sets: functional characterizations and related concepts. Transactions of the American Mathematical Society. 2005;357(4):1275-1301.
11. Nadi MT, Yao JC, Zafarani J. Second-order characterization of convex functions and its applications. Journal of Applied Analysis. 2019;25:49-58.
12. Chieu NH, Lee GM, Mordukhovich BS, Nghia TTA. Coderivative characterizations of maximal monotonicity for set-valued mappings. Journal of Convex Analysis. 2016;23:461-480.
13. Crouzeix JP, Ferland JA. Criteria for differentiable generalized monotone maps. Mathematical Programming. 1996;75:399-406.
14. Nadi MT, Zafarani J. Characterizations of quasiconvex and pseudoconvex functions by their second-order regular subdifferentials. Journal of the Australian Mathematical Society. 2020;109:217-229.
15. Nadi MT, Zafarani J. Second-order characterization of convex mappings in Banach spaces and its applications. Journal of Global Optimization. 2023;86:1005-1023.
16. Daniilidis A, Georgiev P. Approximate convexity and submonotonicity. Journal of Mathematical Analaysis and Applictions. 2004;291(1):292–301.
17. Khanh PD, Phat VT. Second-order characterizations of quasiconvexity and pseudoconvexity for differentiable functions with Lipschitzian derivatives. Optimization Letters. 2020(14): 2413-2427.
18. Hanson MA. On sufficiency of Kuhn-Tucker conditions. Journal of Mathematical Analaysis and Applictions. 1981;80:545–550.
19. Nadi MT, Zafarani J. Second-order characterization of invex functions and its applications in optimization problems. Indian Journal of Industrial and Applied Mathematics. 2019;10:59-70.
20. Nadi MT, Zafarani J. Second-order optimality conditions for constrained optimization problems with C^1 data via regular and limiting subdifferentials. Journal of the Optimization Theory and Applications. 2022;193:158-179.
21. Ben-Tal A, Zowe J. Directional derivatives in nonsmooth optimization. Journal of the Optimization Theory and Applications. 1985;47:483-490.
22. Poliquin RA, Rockafellar RT. Tilt stability of a local minimum. SIAM Journal of Optimization. 1998;287-299.

PERSIAN TRANSLATION OF ABSTRACTS

چکیده‌های فارسی

Bayesian Clustering of Spatially Varying Coefficients Zero-Inflated Survival Regression Models

S. Asadi, M. Mohammadzadeh*

Department of Statistics, Tarbiat Modares University, Tehran, Islamic Republic of Iran.
**Email:mohsen_m@modares.ac.ir*

خوشه‌بندی بیزی مدل‌های رگرسیون بقای متورم صفر با ضرایب متغیر فضایی

سپیده اسعدی، محسن محمدزاده*

گروه آمار، دانشکده علوم ریاضی، دانشگاه تربیت مدرس، تهران، جمهوری اسلامی ایران

چکیده

این مطالعه به چالش‌های تحلیل داده‌های زمان تا وقوع رویداد می‌پردازد و به‌ویژه بر ماهیت گسسته مدت زمان‌تا وقوع پیشامد، مانند تعداد سال‌های تا طلاق، تأکید می‌کند. این وضعیت غالباً منجر به داده‌های بقا با توزیع صفر افزوده می‌شود، زیرا فراوانی مشاهدات صفر به طور قابل توجهی بالا است. برای مقابله با این مشکل، مطالعه از مدل رگرسیون وایبول گسسته صفر آماسیده استفاده می‌کند که به عنوان یک چارچوب مناسب برای ارزیابی تأثیر متغیرهای توضیحی در تحلیل بقا عمل می‌کند. با این حال، چالش‌هایی مانند عدم ایستایی در رابطه بین متغیرها و پاسخ‌ها و ناهمگونی فضایی در مناطق جغرافیایی می‌تواند منجر به مدلی با پارامترهای بسیار زیاد شود. برای کاهش این مشکل، ما یک روش خوشه‌بندی فضایی را برای خلاصه‌سازی فضای پارامترها پیشنهاد می‌کنیم. این مقاله از روش‌های بیز نا پارامتری برای بررسی ناهمگونی فضایی ضرایب رگرسیون بهره می‌برد و بر فرآیند رستوران چینی موزون جغرافیایی برای خوشه‌بندی پارامترهای مدل رگرسیون وایبول گسسته صفر آماسیده تمرکز می‌کند. از طریق مطالعات شبیه‌سازی نشان داده می‌شود که فرآیند رستوران چینی موزون جغرافیایی عملکرد بهتری نسبت به الگوریتم‌های خوشه‌بندی بدون نظارت مانند خوشه‌بندی K-میانگین و فرآیند رستوران چینی استاندارد دارد، به‌ویژه در سناریوهایی با اندازه‌های خوشه نامتعادل از دقت و کارایی محاسباتی بالاتری برخوردار است. شاخص‌های رند بالاتر و میانگین مربعات خطای پایین‌تر بر کارایی و دقت مدل پیشنهادی ما تأکید می‌کنند. اعمال این روش بر روی داده‌های زمان تا وقوع پیشامد طلاق، با انباشتگی درپنج سال اول زندگی زناشویی در ایران، منجر به یک خوشه‌بندی بهینه و برآوردهای کارامد از ضرایب رگرسیونی متغیرهای تبیینی دخیل در این امر با تغییرپذیری فضایی شده است.

واژه‌های کلیدی: تحلیل بقا؛ ضریب متغیر فضایی؛ خوشه بندی فضایی

Modeling Some Repeated Randomized Responses

M. Tarhani, M. R. Zadkarami*, S. M. R. Alavi

Department of Statistics, Faculty of Mathematical Sciences and Computer, Shahid Chamran University of Ahvaz, Ahvaz, Islamic Republic of Iran
** Email: zadkarami@yahoo.co.uk*

مدل سازی چند روش پاسخ تصادفیده مکرر

مهتاب طرهانی، محمدرضا زادکرمی*، سید محمدرضا علوی

گروه آمار دانشگاه شهید چمران اهواز، اهواز، ایران

چکیده

در برخی تحقیقات اجتماعی محقق با موضوعاتی مواجه است که به دلیل حساسیت بالا، پاسخهای قابل اعتمادی به دست نمی‌آید. روشهای پاسخ تصادفیده برای افزایش سطح محرمانگی و ارائه پاسخ صادقانه به کار می‌روند. با این وجود برآوردهای به دست آمده از این روش واریانس بزرگتری دارند. تکرار پاسخ تصادفیده به ازای هر عضو نمونه باعث افزایش تعداد مشاهدات شده و استفاده از میانگین مشاهدات هر شخص برای برآوردیابی واریانس برآورد را کاهش می‌دهد و مقادیر برآورد را به واقعیت نزدیکتر می‌نماید. در این مقاله با استفاده از پاسخهای تصادفیده‌ی جمعی پیوسته مکرر، میانگین پاسخهای هر شخص در نظر گرفته شده است و در قالب یک مدل خطی به برآورد میانگین متغیر حساس و پارامترهای رگرسیونی پرداخته شده است. برای این منظور داده‌های درآمد خانوار جمع آوری شده از دانشجویان به روش تصادفیده‌ی جمعی مکرر بررسی شده است که از هر شخص تقاضا شده پاسخ خود را پنج بار تصادفیده کند. نتایج به دست آمده با دو روش، یک بار برای مشاهده اول و بار دیگر با در نظر گرفتن میانگین مشاهدات هر شخص در قالب مدل رگرسیون به دست آمده است. نتایج نشان می‌دهد برآوردهای به دست آمده از روش دوم به واقعیت نزدیکتر بوده و واریانس کمتری دارند.

واژه های کلیدی: پاسخ تصادفیده؛ پاسخ تصادفیده مکرر؛ مدل رگرسیون خطی؛ متغیر پیوسته حساس؛ تکرار مشاهدات

Bounds for the Varentropy of Basic Discrete Distributions and Characterization of Some Discrete Distributions

F. Goodarzi*

Department of Statistics, Faculty of Mathematical Sciences, University of Kashan, Kashan, Islamic Republic of Iran.

** Email: f-goodarzi@kashanu.ac.ir*

کران‌هایی برای ورآنتروپی توزیع‌های گسسته پایه و مشخص‌سازی برخی توزیع‌های گسسته

فرانک گودرزی*

گروه آمار، دانشکده علوم ریاضی، دانشگاه کاشان، کاشان، جمهوری اسلامی ایران

چکیده

با توجه به اهمیت ورآنتروپی در نظریه اطلاع و از آنجا که شکل بسته‌ای برای ورآنتروپی برخی توزیع‌های گسسته نمی‌توان یافت، هدف ما تعیین کران‌هایی برای ورآنتروپی این توزیع‌ها و معرفی ورآنتروپی گذشته برای متغیرهای تصادفی گسسته است. در این مقاله، ابتدا کران‌های بالا و پایینی را برای ورآنتروپی توزیع‌های پواسون، دوجمله‌ای، دوجمله‌ای منفی و فوق هندسی **به دست آورده‌ایم. با توجه به آن‌که کران‌های بالای حاصل به‌صورت** امید ریاضی توانهای دوم لگاریتمی **بیان می‌شوند، یک عبارت معادل با استفاده از** ضرایب تفاضل لگاریتمی **ارائه می‌کنیم. به همین ترتیب، کران‌های پایینی را نیز برحسب** ضرایب تفاضل لگاریتمی **ارائه می‌دهیم. علاوه بر این، کران بالایی برای** واریانس تابعی از تابع باقیمانده عمر معکوس گسسته **به دست می‌آید. همچنین، نامساوی‌هایی شامل** گشتاورهای‌های توابعی منتخب **از طریق** نرخ شکست معکوس را **بررسی کرده و برخی توزیع‌های گسسته را با استفاده از** نامساوی کُشی شوارتز **مشخص‌سازی می‌کنیم.**

واژه های کلیدی: ورآنتروپی؛ نرخ شکست معکوس؛ تبدیل دو جمله ای؛ نامساوی کُشی شوارتز

Evaluating Feature Selection Methods for Macro-Economic Forecasting, applied for Iran's macro-economic variables

M. Goldani*

Department of Political Science and Economics, Faculty of Literature and Humanities, Hakim Sabzevari University, Sabzevar, Islamic Republic of Iran
** Email: m.goldani@hsu.ac.ir*

ارزیابی روش‌های انتخاب ویژگی برای پیش‌بینی کلان‌اقتصادی: کاربرد برای متغیرهای کلان‌اقتصادی ایران

مهدی گلدانی*

گروه اقتصاد و علوم سیاسی، دانشکده علوم انسانی، دانشگاه حکیم سبزه‌واری، سبزه‌وار، جمهوری اسلامی ایران

چکیده

این مطالعه به ارزیابی روش‌های مختلف انتخاب ویژگی برای بهبود دقت پیش‌بینی مدل‌های کلان‌اقتصادی می‌پردازد، با تمرکز بر شاخص‌های کلان‌اقتصادی ایران که از داده‌های بانک جهانی استخراج شده‌اند. یک مقایسه جامع با استفاده از ۱۴ تکنیک انتخاب ویژگی انجام شد که در دسته‌های روش‌های فیلتری، روش‌های مبتنی بر پوسته ، روش‌های تعبیه‌شده ، و روش‌های مبتنی بر شباهت طبقه‌بندی شده‌اند. چارچوب ارزیابی شامل معیارهای میانگین مربعات خطا و میانگین مطلق خطا در یک فرآیند اعتبارسنجی متقاطع ۱۰-بخشی بود. یافته‌های کلیدی نشان داد که روش انتخاب گام‌به‌گام ، روش‌های مبتنی بر درخت و تکنیک‌های مبتنی بر شباهت، به‌ویژه فاصله‌های هاسدورف و اقلیدسی، به طور مداوم دقت پیش‌بینی بالاتری از خود نشان دادند. میانگین مقادیر میانگین مطلق خطا برای روش انتخاب گام‌به‌گام ۳۲.۰۳ و برای فاصله هاسدورف ۶۲.۶۹ بود. در مقابل، روش‌های حذف بازگشتی ویژگی و آستانه‌گذاری واریانس عملکرد نسبتاً ضعیفی با مقادیر میانگین مطلق خطا به‌مراتب بالاتر داشتند. روش‌های مبتنی بر شباهت، میانگین رتبه ۹.۱۲۵ را در بین مجموعه داده‌ها کسب کردند که نشان‌دهنده استحکام آن‌ها در مواجهه با داده‌های کلان‌اقتصادی با ابعاد بالا است. این نتایج بر پتانسیل ادغام معیارهای شباهت با روش‌های سنتی انتخاب ویژگی برای بهبود دقت و کارایی مدل‌های پیش‌بینی تأکید دارد. این مطالعه بینش‌های ارزشمندی برای پژوهشگران و سیاست‌گذاران ارائه می‌دهد که به دنبال توسعه ابزارهای پیش‌بینی اقتصادی قابل‌اعتماد هستند.

واژه‌های کلیدی: انتخاب ویژگی؛ دقت پیش‌بینی؛ شاخص‌های بانک جهانی؛ تحلیل کلان‌اقتصادی؛ روش‌های شباهت

Comparison of Adaptive Neural-Based Fuzzy Inference System and Support Vector Machine Methods for the Jakarta Composite Index Forecasting

A. Mutmainnah*, S. A. Thamrin, G. M. Tinungki

Department of Statistics, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Makassar, 90245 Indonesia
Makassar, Indonesia

* *Email: amutmainnah8@gmail.com*

مقایسه سیستم استنتاج فازی مبتنی بر عصبی تطبیقی و روش‌های ماشین بردار پشتیبان برای پیش‌بینی شاخص ترکیبی جاکارتا

آیو موتمینه*، سری آستوتی تامرین، جورجینا ماریا تینونگکی

گروه آمار، دانشکده ریاضیات و علوم طبیعی، دانشگاه حسن الدین، ماکاسار، ۹۰۲۴۵ اندونزی ماکاسار، اندونزی

چکیده

شاخص ترکیبی جاکارتا (JCI) یک معیار اساسی برای ارزیابی عملکرد تمام سهام های فهرست شده در بورس اوراق بهادار اندونزی (IDX) است. با توجه به پیچیدگی ذاتی، غیرخطی و غیرثابت بودن داده‌های بازار سهام، انتخاب روش‌های پیش‌بینی قوی ضروری است. این مطالعه عملکرد سیستم استنتاج عصبی فازی تطبیقی (ANFIS) و ماشین بردار پشتیبان (SVM) را در پیش‌بینی حرکات JCI مقایسه می‌کند. محقق دقت پیش‌بینی را با استفاده از ریشه میانگین مربعات خطا (RMSE) و میانگین درصد مطلق خطا (MAPE) ارزیابی کرد. مرحله آموزش نشان داد که مدل ANFIS بهینه از تابع عضویت زنگ تعمیم‌یافته استفاده می‌کند و از جایگزین‌های ذوزنقه‌ای و گاوسی بهتر عمل می‌کند. به طور همزمان، بهترین پیکربندی SVM از یک هسته خطی (هزینه = ۱۰) استفاده می‌کند که عملکرد برتر را در مقایسه با تابع پایه شعاعی (RBF) و هسته‌های سیگموئید نشان می‌دهد. در مرحله آزمایش، ANFIS به RMSE=39.894 و MAPE=0.4647 دست یافت، در حالی که SVM RMSE=38.728 و MAPE=0.4516 را ثبت کرد. این نتایج بر قابلیت‌های پیش‌بینی برتر SVM تاکید می‌کند و آن را به عنوان ابزاری قابل اعتماد برای پیش‌بینی بازار سهام قرار می‌دهد. یافته‌های این مطالعه بینش‌های ارزشمندی را برای سرمایه‌گذاران و سیاست‌گذاران در جهت‌یابی عدم قطعیت‌های بازار و بهینه‌سازی استراتژی‌های سرمایه‌گذاری فراهم می‌کند.

واژه های کلیدی: پیش بینی؛ ماشین بردار پشتیبانی; شاخص ترکیبی جاکارتا; سیستم استنتاج فازی مبتنی بر عصبی تطبیقی

Second-ordered Characterization of Generalized Convex Functions and Their Applications in Optimization Problems

M. T. Nadi[1*], J. Zafarani[2]

[1] *School of Mathematics, Institute for Research in Fundamental Sciences (IPM), P. O. Box 19395-5746, Tehran, Islamic Republic of Iran*
[2] *Department of Mathematics, Sheikhbahaee University and University of Isfahan, Isfahan, Islamic Republic of Iran*

** Email: mt_nadi@yahoo.com*

مشخص‌سازی مرتبه دوم توابع محدب تعمیم‌یافته و کاربردهای آنها در مسائل بهینه‌سازی

محمد تقی نادی*[۱]، جعفر زعفرانی[۲]

[۱] *پژوهشکده ریاضیات، پژوهشگاه دانشهای بنیادی (IPM)، تهران، جمهوری اسلامی ایران*
[۲] *گروه ریاضی، دانشگاه شیخ بهایی و دانشگاه اصفهان، اصفهان، جمهوری اسلامی ایران*

چکیده

این مطالعه، برخی از توسعه‌های مشخص‌سازی مرتبه دوم توابع محدب تعمیم‌یافته را با استفاده از هم‌مشتق نگاشت زیرمشتق بررسی می‌کند. بطور دقیقتر، مشخص‌سازی مرتبه دوم توابع شبه محدب، محدب نما و محدب پایا ارائه می‌شوند. علاوه بر این بعضی از کاربردهای زیرمشتق‌های مرتبه دوم در مسائل بهینه‌سازی، مانند برنامه‌ریزی غیر خطی مقید و نامقید را ارائه می‌دهد.

واژه‌های کلیدی: زیرمشتق مرتبه دوم؛ ویژگی نیم‌معین مثبت؛ زیرمشتق مرتبه دوم منظم؛ شرایط بهینه مرتبه دوم

JOURNAL OF SCIENCES

ISLAMIC REPUBLIC OF IRAN

SUBSCRIPTION FORM

Please enter my annual subscription to the Journal of Sciences, Islamic Republic of Iran including 4 quarterly issues for the Year …… Vol. …… No. ……

	Iran	Other Countries
☐ Personal	R. 30,000	$ 80.00
☐ Institutional	R. 40,000	$ 100.00
☐ Student	R. 16,000	$ 50.00

☐ Check enclosed ☐ Bill me

Name:
Mailing Address: **City:** **Country:**

Check or money order must be made to order of:
Journal of Sciences, Islamic Republic of Iran, University of Tehran
Enghelab Ave., P.O. Box 13145-478, Tehran, Iran
Payment can be made via our transfer account.
Iran: Account No. 5225458914 Mellat Bank, Branch of University of Tehran, Central Income of Vice President of Research.
Foreign: Account No. 162454986 tejarat Bank, Branch of University of Tehran, Islamic Republic of Iran.
* Please allow 6-8 weeks for delivery.

JOURNAL OF SCIENCES

ISLAMIC REPUBLIC OF IRAN

SUBSCRIPTION FORM

Please enter my annual subscription to the Journal of Sciences, Islamic Republic of Iran including 4 quarterly issues for the Year …… Vol. …… No. ……

	Iran	Other Countries
☐ Personal	R. 30,000	$ 80.00
☐ Institutional	R. 40,000	$ 100.00
☐ Student	R. 16,000	$ 50.00

☐ Check enclosed ☐ Bill me

Name:
Mailing Address: **City:** **Country:**

Check or money order must be made to order of:
Journal of Sciences, Islamic Republic of Iran, University of Tehran
Enghelab Ave., P.O. Box 13145-478, Tehran, Iran
Payment can be made via our transfer account.
Iran: Account No. 5225458914 Mellat Bank, Branch of University of Tehran, Central Income of Vice President of Research.
Foreign: Account No. 162454986 tejarat Bank, Branch of University of Tehran, Islamic Republic of Iran.
* Please allow 6-8 weeks for delivery.

To:
Journal of Sciences
Islamic Republic of Iran

University of Tehran
Enghelab Ave., Tehran
P.O. Box 13145-478
Islamic Republic of Iran

Stamp

گیرنده:
تهران - مجله بین المللی علوم پایه جمهوری اسلامی ایران
تهران، خیابان انقلاب، دانشگاه تهران،
صندوق پستی ۴۷۸ - ۱۳۱۴۵

تمبر

Cover Legend:

"Abu Reyhan al-Biruni (973-1048) was one of the greatest scholars and philosopher in the history of Islamic civilization. He was equally well versed in a number of fields including mathematics, astronomy, physical and natural sciences, geography, history, chronology, and linguistics. George Sarton has called the first half of the 11th century «the age of Biruni». He wrote over 150 works, including 18 volumes on astronomy, 15 volumes on geography, 13 on literature, 12 volumes on astrology, 6 on religious ideology, 6 on astrolabes, 5 on timing devices, 5 on comets and 4 on the physics of light and optics. His scientific contributions also include the accurate determination of the density of eighteen different metals and minerals. He calculated the sine and cosine of angles from zero to 90 degrees and recorded them in a table. He made a remarkably accurate determination of the radius of the earth, which is very close to the latest estimate. He wrote dissertations regarding comets and meteorites, sun rays, light, twilight and dawn, time and space, astrolabes, methods to determine Qibla, the values of numerals, the relativity between precious stones and metals, ways of determining the longitude and latitude of cities, and the distances between them, and devised a method of measuring the earth's circumference, and complex problems of trigonometry. In addition to his native language (Persian) he also mastered Arabic, Sanskrit, Hebrew, Greek, Syriac and Turkish."

"Ref.: Noori-Daloii, M.R., J. Sci. I.R. Iran, Vol. 1, No. 1, pp. 2-3,. Autumn 1988"

JOURNAL OF SCIENCES

ISLAMIC REPUBLIC OF IRAN

University of Tehran

Indexed by Thomson Reuters, Scopus, Islamic World Science Citation Center (ISC), Scientific Information Database (SID) and/Abstracted in

Biological Abstracts
Biosis Selective Coverage Unique
Biosis Previews
Zoological Records
Chemical Abstracts
Methods in Organic Synthesis
Mathematical Reviews
Current Mathematical Publications
MathSciNet