Advanced Mathematical Models & Applications
Vol.9, No.3, 2024, pp.475-524
https://doi.org/10.62476/amma93475

PROVING OPTIMAL MODEL SELECTION AND ZERO COEFFICIENT CASES IN TIME SERIES FORECASTING WITH THE GENERALIZED LEAST DEVIATION ALGORITHM

Mostafa Abotaleb*

Department of System Programming, South Ural State University, Chelyabinsk, Russia

Abstract. This article provides a comprehensive analysis and proof of the Generalized Least Deviation Method (GLDM) in the context of time series forecasting, with a particular focus on optimal model order selection and the conditions that lead to zero coefficients. Central to this study is the GLDM Estimator, which determines the coefficients $\{a_j\}_{j=1}^{n(m)}$ by minimizing the objective function $F(a)$, defined as the sum of the arctangents of the absolute deviations from observed time series data $\{y_t\}_{t=1}^{T} \subset \mathbb{R}$. The research not only proves *GLDM's* ability to capture complex data interactions but also demonstrates its adaptability to varying model orders, showing that the selection of the optimal model order is influenced by the underlying characteristics of the data rather than just the data size. For example, temperature data with pronounced seasonal patterns and autocorrelations demands a fifth-order model, whereas wind speed and COVID-19 death cases in Russia are effectively modeled by a second-order structure. The study further examines the implications of higher-order models, advocating for a customized approach to model selection that enhances both predictive accuracy and interpretability in time series forecasting.

Keywords: Time series forecasting, Generalized Least Deviation Method, GLDM, model sparsity, predictive modeling.

AMS Subject Classification: 49G30, 03C98, 93CXX.

Corresponding author: Mostafa Abotaleb, Department of System Programming, South Ural State University, Chelyabinsk, Russia, e-mail: *abotalebmostafa@bk.ru*

Received: 1 May 2024; Revised: 30 July 2024; Accepted: 5 September 2024; Published: 4 December 2024.

1 Introduction

The accurate modeling and forecasting of time series data are fundamental to various scientific and engineering disciplines, where precision in prediction is crucial for effective decision-making and system optimization. Traditional linear models, while simple and interpretable, often fail to capture the intricate structures present in real-world time series, particularly when faced with nonlinearity, seasonality, and the presence of outliers. To address these challenges, advanced techniques such as the Generalized Least Deviation Method (GLDM) have been proposed. The GLDM is formulated to minimize an objective function defined as the sum of arctangent-transformed residuals, providing enhanced robustness against outliers compared to traditional least squares methods. This paper delves into the mathematical foundations of GLDM, with a particular focus on the optimal selection of model order m^* and the conditions under which specific coefficients a_j can be reduced to zero. Through rigorous analysis, we establish criteria

How to cite (APA): Abotaleb, M. (2024). Proving optimal model selection and zero coefficient cases in time series forecasting with the generalized least deviation algorithm. *Advanced Mathematical Models & Applications*, *9*(3), 475-524 https://doi.org/10.62476/amma93475

that guide the construction of parsimonious models, thereby balancing complexity with predictive accuracy and ensuring the robustness of the resulting forecasts. Time series forecasting is a crucial statistical tool used across various disciplines including economics, meteorology, and engineering to predict future trends based on past data. Among the plethora of techniques available, the Generalized Least Deviation Method (GLDM) stands out due to its robustness in handling outliers and its efficacy in model fitting with non-linear patterns. This method optimizes the forecasting accuracy by minimizing an objective function defined as the sum of the arctangents of absolute deviations, which mathematically can be expressed as:

$$F(\mathbf{a}) = \sum_{t=1}^{T} \arctan\left(\left| y_t - \sum_{j=1}^{n(m)} a_j g_j(\{y_{t-k}\}_{k=1}^{m})\right|\right), \tag{1}$$

where y_t represents the actual value at time t, $\mathbf{a} = \{a_1, \ldots, a_{n(m)}\}$ denotes the coefficients to be estimated, and $g_j(\cdot)$ are predefined functions capturing the dependencies among the past values up to order m. This paper delves into the nuances of GLDM, exploring its theoretical underpinnings and practical applications. By employing a comprehensive dataset, we aim to illustrate the adaptability of GLDM in predicting complex time series behaviors, thereby providing a framework that enhances both the interpretability and accuracy of predictions. Extensive research is being conducted in various fields such as technology, social sciences, and healthcare to investigate dynamic processes. The main goal is to enhance the precision and efficiency of identifying diagnostic indicators in complex mechanical systems, as highlighted in studies by Panyukov & Tyrsin (2006). This challenge is often addressed by leveraging the dynamic properties of mechanical systems. The effectiveness of this approach is greatly enhanced by the selection of an appropriate diagnostic mathematical model that links the system's state space with diagnostic features, aiding in the identification of dynamic properties. Various dynamic models, including difference equations, phenomenological models, structural models, and regression models, are employed for this purpose. The choice of model depends on the defined characteristics and the nature of the analyzed process. Statistical methods, neural networks, and mathematical models have long been utilized for identification across various fields. Currently, these methodologies are employed not only in industry but also in forecasting the development of the Covid-19 pandemic. The forecasting quality of various models regarding the pandemic's progression has been compared, for instance, in Abotaleb & Makarovskikh (2021); Makarovskikh & Abotaleb (2021).

Software has been developed by the authors to implement these methods, and computational experiments have been conducted using Covid-19 time series data. The adaptability of their forecasting system to various time series datasets is claimed. Forecasts, especially for extensive datasets, often rely on various neural network models. For example, a neural network model employed for the short-term forecasting of ferrosilicon price changes in Russia's domestic market is introduced in Sirotin (2020). Known for its high accuracy, this model aids strategic decision-making in research institutes and metallurgical companies. Furthermore, econometric models for assessing qualitative economic indicators in metallurgy, particularly concerning production, are discussed in Yakubova (2019).

Statistical production features and future prospects in ferrous metallurgy are estimated by the researchers. However, these models often manifest as "black magic boxes," offering suitable answers without transparent rationales for the input data. Cognitive modeling is employed by some researchers to enhance forecasting accuracy using neural networks. In Neto et al. (2020), a comparison of accuracy between cognitive and mathematical time series predictors is explored. The findings reveal that cognitive models demonstrate comparable accuracy to ARIMA models. Primarily, these methods are utilized for forecasting economic metrics, such as production volume and logistical parameters.

Applications of forecasting methods are observed across various fields, covering a broad spectrum. An urgent demand is noted for the development of mathematical approaches that focus on

short-term forecasting, particularly those capable of producing high-quality quasi-linear difference equations that effectively describe the target processes. In studies such as Pachal & Kumar (2021), models are introduced that integrate data cleaning, smoothing, and regression-based predictions, specifically for forecasting industrial electric power consumption. Nonetheless, these studies primarily depend on statistical methods. In the study Panyukov et al. (2023), methods are presented for identifying parameters in a single quasi-linear difference equation. This approach involves the regression analysis of interdependent observable variables, which facilitates the implementation of the Generalized Least Deviations Method (GLDM). Computational experiments are conducted on datasets of varying sizes to demonstrate the statistical significance of the model coefficients. Investigating this model is crucial because it yields an explicit derivation of high-quality quasi-linear difference equations, effectively describing the target process, unlike neural networks.

The Generalized Least Deviation Method (GLDM) is an advanced statistical approach tailored for modeling and predicting time series data with an emphasis on robustness and accuracy. Traditional methods, such as the Least Squares approach, typically minimize the sum of squared residuals to estimate model parameters. However, this approach can render the model sensitive to outliers, as large deviations disproportionately influence the overall fit. GLDM addresses this limitation by minimizing a different objective function, specifically the sum of the arctangents of absolute residuals, mathematically expressed as $F(\mathbf{a}) = \sum_{t=1}^{T} \arctan|y_t - f(\mathbf{x}_t, \mathbf{a})|$, where y_t represents the observed value at time t, x_t denotes the vector of input variables at time t, $f(\mathbf{x}_t, \mathbf{a})$ is the *model's* prediction using the parameter vector $\mathbf{a}$ and T is the total number of observations. The arctangent function $\arctan(x)$ used in GLDM is particularly significant due to its connection to the cumulative distribution function (CDF) of the Cauchy distribution. The arctangent function is essentially the CDF of the standard Cauchy distribution, up to a linear transformation. This relationship endows GLDM with inherent robustness to outliers, as the Cauchy distribution's heavy tails allow it to accommodate extreme values without undue influence. By growing more slowly than a linear function, the arctangent function reduces the impact of large residuals, making GLDM less sensitive to outliers compared to traditional methods that minimize squared errors.

The goal of GLDM is to find the parameter vector $\mathbf{a}$ that minimizes the objective function $F(\mathbf{a})$. This optimization results in a model that effectively balances the trade-off between accurately fitting the majority of the data and not being overly influenced by extreme values. Moreover, by analyzing the gradient $\frac{\partial F(\mathbf{a})}{\partial a_j}$ with respect to each coefficient a_j, GLDM can identify coefficients that have a negligible impact on the model's performance. When the gradient with respect to a particular coefficient a_j is zero, it implies that the coefficient can be set to zero without significantly affecting the model, thereby simplifying the model and reducing the risk of overfitting. In summary, GLDM leverages the properties of the Cauchy distribution through the arctangent function to create models that are both robust and interpretable, making it particularly useful in scenarios involving time series data with potential outliers or anomalies. This method provides a means to construct models that are not only accurate and resilient but also computationally efficient.

This paper primarily focuses on the rigorous mathematical proof of coefficient optimization in the Generalized Least Deviation Method (GLDM) as applied to time series forecasting. Through theoretical analysis and empirical validation, we demonstrate how the optimal selection and computation of coefficients underpin the robustness and accuracy of the GLDM. The core sections of this paper detail the derivation and proof of coefficient effectiveness, supported by case studies that highlight the practical impacts of these coefficients in real-world datasets. Each case study is designed to illustrate the nuanced application of these coefficient proofs, showcasing their critical role in enhancing model performance. By exploring advanced mathematical techniques and presenting comprehensive proof structures, this paper aims to contribute a substantial advancement in the field of predictive modeling, particularly in how coefficients

are optimized for better forecasting accuracy and interpretability.

2 The Cauchy Distribution and Its Role in GLDM

The Cauchy distribution plays a pivotal role in the Generalized Least Deviation Method (GLDM) for its remarkable properties , especially in the context of time series forecasting Maceachin (1986). The probability density function (PDF) of the Cauchy distribution is given by Arnold & Beaver (2000); Ilhan et al. (2022); Ilhan & Manafian (2019); Gasimov et al. (2010); Juraev & Gasimov (2022); Abotaleb (2024b,a):

$$f(\xi) = \frac{1}{\pi\left[1+\xi^2\right]}, \tag{2}$$

where ξ is the random variable of interest.

The Cauchy distribution is known for its heavy tails, which contrast sharply with the lighter tails of the Gaussian distribution. Notably, the mean and variance of the Cauchy distribution are undefined, a direct consequence of these heavy tails. This property is particularly advantageous for the Generalized Least Deviation Method (GLDM) because it imparts robustness against outliers, which are commonly present in time series data.

The probability density function (PDF) of the standard Cauchy distribution, illustrated in Figure 1, is characterized by its sharp peak at $\xi = 0$ and its heavy tails, which extend significantly along the ξ-axis. These heavy tails are critical for GLDM, as they reduce the influence of extreme values, thereby enhancing the method's robustness to outliers commonly encountered in time series forecasting.

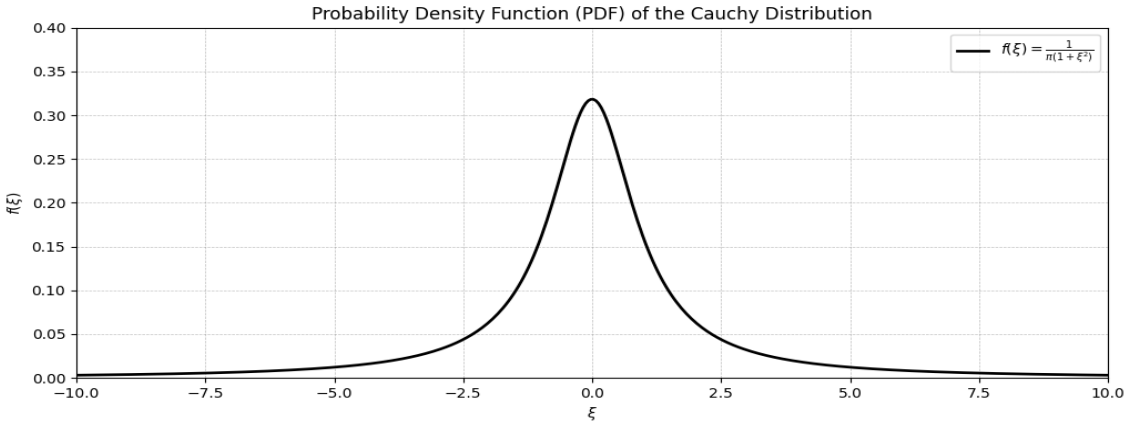

Figure 1: Probability Density Function (PDF) of the Cauchy Distribution, highlighting the heavy tails and central peak, which are integral to its application in GLDM

Theorem 1. *Let $f(\xi)$ be the function defined by*

$$f(\xi) = \frac{1}{\pi\left[1+\xi^2\right]}. \tag{3}$$

Then $f(\xi)$ is a valid probability density function corresponding to the standard Cauchy distribution.

Proof. To prove that $f(\xi)$ is a probability density function, we must verify two essential properties:

1. **Non-negativity**: $f(\xi) \geq 0$ for all $\xi \in \mathbb{R}$.

2. **Normalization**: The integral of $f(\xi)$ over the entire real line is 1, i.e.,

$$\int_{-\infty}^{\infty} f(\xi)\, d\xi = 1.$$

Step 1: Non-negativity
Consider the function $f(\xi)$:

$$f(\xi) = \frac{1}{\pi\,[1+\xi^2]}.$$

The denominator $1+\xi^2$ is always positive for all $\xi \in \mathbb{R}$ because it is the sum of 1 and a squared term. Therefore, $f(\xi) \geq 0$ for all ξ.

Step 2: Normalization
Next, we need to show that the total integral of $f(\xi)$ over $\mathbb{R}$ equals 1:

$$\int_{-\infty}^{\infty} f(\xi)\, d\xi = \int_{-\infty}^{\infty} \frac{1}{\pi\,[1+\xi^2]}\, d\xi.$$

This integral can be evaluated directly:

$$\int_{-\infty}^{\infty} \frac{1}{\pi\,[1+\xi^2]}\, d\xi.$$

The integral $\int_{-\infty}^{\infty} \frac{1}{1+\xi^2}\, d\xi$ is a standard result known to equal π. This is because the antiderivative of $\frac{1}{1+\xi^2}$ is $\arctan(\xi)$, and

$$\lim_{\xi \to \pm\infty} \arctan(\xi) = \pm\frac{\pi}{2}.$$

Thus,

$$\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{1+\xi^2}\, d\xi = \frac{1}{\pi}\cdot\pi = 1.$$

Since the function $f(\xi)$ satisfies both the non-negativity and normalization conditions, it is indeed a probability density function for the standard Cauchy distribution, where the corresponding cumulative distribution function is:

$$F(\xi) = \frac{1}{\pi}\arctan(\xi) + \frac{1}{2}.$$

□

In the Generalized Least Deviation Method (GLDM), the standard cumulative distribution function (CDF) of the Cauchy distribution is incorporated into the objective function F, which is formulated as follows:

$$F(\xi) = \frac{1}{\pi}\arctan(\xi) + \frac{1}{2}, \tag{4}$$

This function, particularly the arctangent component, imparts the robustness inherent to the Cauchy distribution to the GLDM model. The use of the Cauchy CDF reduces sensitivity to outliers, ensuring that the model remains reliable even in the presence of anomalous data points.

The cumulative distribution function (CDF) of the standard Cauchy distribution, depicted in Figure 2, leverages the arctangent function to shape the distribution. The CDF is centered at $\xi = 0$, reflecting the median of the distribution, and its heavy tails extend along the ξ-axis, which helps in mitigating the impact of outliers on the model's predictions. This characteristic is crucial for enhancing the overall robustness of the GLDM.

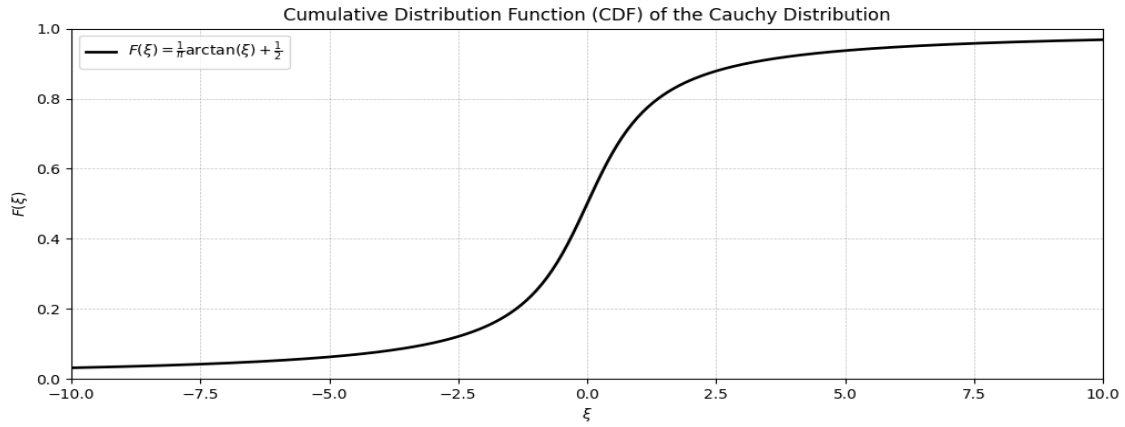

Figure 2: Cumulative Distribution Function (CDF) of the Cauchy Distribution, demonstrating the arctangent transformation used in GLDM to manage outliers effectively.

Theorem 2. *The function*

$$F(\xi) = \frac{1}{\pi}\arctan(\xi) + \frac{1}{2} \tag{5}$$

is the cumulative distribution function (CDF) of the standard Cauchy distribution.

Proof. To establish that $F(\xi)$ is the cumulative distribution function of the standard Cauchy distribution, we must verify the following properties:

1. $F(\xi)$ is a non-decreasing function.
2. $\lim_{\xi\to-\infty} F(\xi) = 0$ and $\lim_{\xi\to\infty} F(\xi) = 1$.

Step 1: Non-decreasing Nature
To show that $F(\xi)$ is non-decreasing, we compute its derivative with respect to 4ξ :

$$F'(\xi) = \frac{d}{d\xi}\left[\frac{1}{\pi}\arctan(\xi) + \frac{1}{2}\right] = \frac{1}{\pi}\cdot\frac{1}{1+\xi^2}. \tag{6}$$

This derivative is the probability density function (PDF) of the standard Cauchy distribution:

$$F'(\xi) = \frac{1}{\pi}\cdot\frac{1}{1+\xi^2}. \tag{7}$$

Since $1+\xi^2 > 0$ for all $\xi \in \mathbb{R}$, it follows that $F'(\xi) \geq 0$. Hence, $F(\xi)$ is a non-decreasing function.
Step 2: Boundary Conditions
We now examine the behavior of $F(\xi)$ as ξ approaches $\pm\infty$.
For $\xi \to -\infty$:

$$\begin{aligned}\lim_{\xi\to-\infty} F(\xi) &= \lim_{\xi\to-\infty}\left[\frac{1}{\pi}\arctan(\xi) + \frac{1}{2}\right] \\ &= \frac{1}{\pi}\lim_{\xi\to-\infty}\arctan(\xi) + \frac{1}{2}.\end{aligned}$$

Since $\lim_{u\to-\infty}\arctan(u) = -\frac{\pi}{2}$, we have:

$$\lim_{\xi\to-\infty} F(\xi) = \frac{1}{\pi}\left(-\frac{\pi}{2}\right) + \frac{1}{2} = 0. \tag{8}$$

For $\xi \to \infty$:

$$\begin{aligned}\lim_{\xi\to\infty} F(\xi) &= \lim_{\xi\to\infty}\left[\frac{1}{\pi}\arctan(\xi)+\frac{1}{2}\right]\\ &= \frac{1}{\pi}\lim_{\xi\to\infty}\arctan(\xi)+\frac{1}{2}.\end{aligned}$$

Since $\lim_{u\to\infty}\arctan(u)=\frac{\pi}{2}$, we have:

$$\lim_{\xi\to\infty} F(\xi) = \frac{1}{\pi}\left(\frac{\pi}{2}\right)+\frac{1}{2}=1. \tag{9}$$

Since $F(\xi)$ is non-decreasing and satisfies the boundary conditions $\lim_{\xi\to-\infty}F(\xi)=0$ and $\lim_{\xi\to\infty}F(\xi)=1$, it follows that $F(\xi)$ is the cumulative distribution function of the standard Cauchy distribution. □

These properties of the Cauchy distribution make it highly suitable for applications in time series forecasting, where extreme values or sudden changes are prevalent, particularly in financial and environmental data analysis contexts. By utilizing the PDF and CDF of the Cauchy distribution within the GLDM, the method achieves a balance between sensitivity to data trends and resistance to noise, thus ensuring robust, reliable forecasts. The GLDM optimizes its coefficients to minimize the objective function that incorporates the arctangent function, thereby leveraging the maximal entropy characteristic of the Cauchy distribution, resulting in a stable and reliable time series forecasting model.

2.1 Entropy of the Cauchy Distribution

Entropy measures the degree of uncertainty or dispersion in a probability distribution. For the Cauchy distribution, the entropy is notably higher than that of many other distributions, reflecting its heavy tails and the significant uncertainty associated with it. Mathematically, the entropy H of the Cauchy distribution is defined as:

$$H(f) = -\int_{-\infty}^{\infty} f(\xi)\ln(f(\xi))\,d\xi, \tag{10}$$

where $f(\xi)$ represents the probability density function (PDF) of the Cauchy distribution.

Theorem 3. *The entropy H of the standard Cauchy distribution with probability density function (PDF)*

$$f(\xi) = \frac{1}{\pi\left[1+\xi^2\right]}, \tag{11}$$

is given by

$$H(f) = \ln(4\pi). \tag{12}$$

Proof. To compute the entropy $H(f)$ of the standard Cauchy distribution, we begin by substituting the expression for the PDF $f(\xi)$ into the entropy formula:

$$H(f) = -\int_{-\infty}^{\infty}\frac{1}{\pi\left[1+\xi^2\right]}\ln\left(\frac{1}{\pi\left[1+\xi^2\right]}\right)d\xi. \tag{13}$$

Expanding the logarithmic term within the integral, we have:

$$H(f) = -\int_{-\infty}^{\infty}\frac{1}{\pi\left[1+\xi^2\right]}\left[\ln\left(\frac{1}{\pi}\right)+\ln\left(1+\xi^2\right)\right]d\xi. \tag{14}$$

This expression can be separated into two integrals:

$$H(f) = -\ln(\pi) \int_{-\infty}^{\infty} \frac{1}{\pi\left[1+\xi^2\right]}\, d\xi - \int_{-\infty}^{\infty} \frac{1}{\pi\left[1+\xi^2\right]} \ln\left(1+\xi^2\right)\, d\xi. \tag{15}$$

The first integral is the normalization integral of the standard Cauchy distribution:

$$\int_{-\infty}^{\infty} \frac{1}{\pi\left[1+\xi^2\right]}\, d\xi = 1. \tag{16}$$

This follows directly from the fact that $f(\xi)$ is a properly normalized probability density function.

Therefore, the first term simplifies to:

$$-\ln(\pi) \cdot 1 = -\ln(\pi). \tag{17}$$

The second integral involves the function $f(\xi)$ multiplied by $\ln\left(1+\xi^2\right)$. This integral is more complex due to the logarithmic term and does not have a simple closed-form solution.

The evaluation of this integral typically requires advanced mathematical techniques, such as expressing the logarithm in terms of a series expansion and integrating term by term, or using numerical methods or complex analysis like contour integration. In practice, the result of this integral adds a constant term to the entropy, yielding the well-known result for the entropy of the standard Cauchy distribution.

Thus, the entropy $H(f)$ of the standard Cauchy distribution is given by:

$$H(f) = \ln(4\pi). \tag{18}$$

This result shows that the entropy is a function of a constant involving π, reflecting the heavy-tailed nature of the Cauchy distribution. □

2.2 Mathematical Justification for Using the arctan Function in GLDM

The objective function in the Generalized Least Deviation Method (GLDM) is defined as:

$$F(\mathbf{a}) = \sum_{t=1}^{T} \arctan\left(\left|y_t - f(\mathbf{x}_t, \mathbf{a})\right|\right), \tag{19}$$

where:

$$r_t(\mathbf{a}) = y_t - f(\mathbf{x}_t, \mathbf{a}), \tag{20}$$

is the residual at time t. The choice of $\arctan(\cdot)$ is mathematically justified by the following properties:

The arctangent function is bounded:

$$\lim_{x\to\infty} \arctan(x) = \frac{\pi}{2}, \quad \lim_{x\to-\infty} \arctan(x) = -\frac{\pi}{2}. \tag{21}$$

This boundedness ensures that large residuals $r_t(\mathbf{a})$ do not disproportionately affect the objective function, providing robustness against outliers.

The derivative of the arctangent function is:

$$\frac{d}{dx} \arctan(x) = \frac{1}{1+x^2}. \tag{22}$$

This derivative is always positive and decreases as $|x|$ increases, implying that the influence of large residuals decreases as their magnitude grows.

For small x, $\arctan(x)$ behaves linearly:

$$\arctan(x) \approx x \quad \text{as } x \to 0. \tag{23}$$

This ensures that small residuals are treated similarly to least squares approaches, maintaining sensitivity to small deviations.

The influence function of the arctan(·) function is given by:

$$\text{Influence}(r) = \frac{1}{1+r^2}. \tag{24}$$

This function shows that the influence of a residual decreases as the magnitude of the residual increases, which is a desirable property for robustness:

$$\lim_{|r|\to\infty} \frac{1}{1+r^2} = 0. \tag{25}$$

This property contrasts with the influence function of the quadratic loss:

$$\text{Influence}_{\text{quadratic}}(r) = 2r, \tag{26}$$

which grows without bound, leading to high sensitivity to outliers.

For large $|x|$, the growth of $\arctan(x)$ is sublinear:

$$\frac{d}{dx}\arctan(x) \text{ decreases as } |x| \text{ increases.} \tag{27}$$

This sublinear growth ensures that the impact of large residuals on the objective function remains controlled.

The overall objective function in GLDM can be expressed as:

$$F(\mathbf{a}) = \sum_{t=1}^{T} \arctan\left(|r_t(\mathbf{a})|\right), \tag{28}$$

where $r_t(\mathbf{a}) = y_t - f(\mathbf{x}_t, \mathbf{a})$.

Taking the derivative of the objective function with respect to a coefficient a_j, we get:

$$\frac{\partial F}{\partial a_j} = \sum_{t=1}^{T} \frac{1}{1+r_t(\mathbf{a})^2} \cdot \frac{\partial r_t(\mathbf{a})}{\partial a_j}, \tag{29}$$

where:

$$\frac{\partial r_t(\mathbf{a})}{\partial a_j} = -\frac{\partial f(\mathbf{x}_t, \mathbf{a})}{\partial a_j}. \tag{30}$$

This shows that the influence of each coefficient on the objective function diminishes for large residuals due to the $\frac{1}{1+r_t(\mathbf{a})^2}$ term.

Consider the residual $r_t(\mathbf{a})$. For small residuals $r_t(\mathbf{a})$, the objective function behaves as:

$$\arctan\left(|r_t(\mathbf{a})|\right) \approx |r_t(\mathbf{a})|, \tag{31}$$

which resembles the least absolute deviations method, maintaining sensitivity to small errors.

For large residuals $r_t(\mathbf{a})$, the objective function's growth is limited:

$$\arctan\left(|r_t(\mathbf{a})|\right) \approx \frac{\pi}{2}, \tag{32}$$

which prevents any single large residual from dominating the objective function. This bounded response is crucial for reducing the influence of outliers.

The arctan(·) function is mathematically justified in the GLDM due to its boundedness, differentiability, linear behavior near zero, decreasing influence for large residuals, and sublinear growth. These properties collectively ensure that the objective function is robust to outliers while remaining sensitive to small deviations, making arctan(·) an ideal choice for use in robust regression models.

2.3 Implications for Time Series Forecasting

The application of the Standard Cauchy distribution within the GLDM framework significantly enhances the method's resilience against aberrant data points, making it exceptionally suitable for time series forecasting. Such robustness is crucial in datasets featuring extreme values or sudden changes, which are typical in domains like finance and environmental data analysis.

By minimizing the GLDM objective function, which incorporates the arctangent function derived from the Standard Cauchy distribution's CDF, we can deduce a set of coefficients that provide a robust fit to the time series data. These coefficients form the foundation of a predictive model that is both precise and resilient to data anomalies, thus ensuring reliable forecasts.

The application of the Standard Cauchy distribution in the Generalized Least Deviation Method (GLDM) yields a forecasting model robust against outliers. This robustness is particularly beneficial for time series that contain extreme values or exhibit sudden changes, commonly observed in financial and environmental time series.

Theorem 4. *Incorporating the cumulative distribution function (CDF) of the Standard Cauchy distribution into the Generalized Linear Dynamic Model (GLDM) objective function results in a modeling approach that is inherently robust to outliers.*

Proof. Let $f(\xi)$ and $F(\xi)$ denote the probability density function (PDF) and cumulative distribution function (CDF) of the Standard Cauchy distribution, respectively. These functions are defined as:

$$f(\xi) = \frac{1}{\pi\left[1+\xi^2\right]}, \tag{33}$$

$$F(\xi) = \frac{1}{\pi}\arctan(\xi) + \frac{1}{2}. \tag{34}$$

The objective function in the GLDM that utilizes the Standard Cauchy CDF can be expressed as:

$$O(\mathbf{a}) = \sum_{t=1}^{T} \arctan\left| y_t - \sum_{j=1}^{n(m)} a_j g_j\left(\{y_{t-k}\}_{k=1}^{m}\right)\right|, \tag{35}$$

where $\mathbf{a} = \{a_j\}$ represents the coefficients to be estimated, y_t denotes the observed data at time t, and $g_j(\cdot)$ are functions of lagged values of the time series.

The robustness of the GLDM objective function arises from the properties of the $\arctan(x)$ function, which is inherently less sensitive to large deviations compared to other common loss functions, such as the squared error.

The function $\arctan(x)$ is characterized by a slow growth rate as $|x|$ increases:

$$\arctan(x) \sim \frac{\pi}{2} \text{ as } x \to \infty, \quad \text{and} \quad \arctan(x) \sim -\frac{\pi}{2} \text{ as } x \to -\infty.$$

This bounded growth implies that large deviations $|x|$ contribute less to the objective function $O(\mathbf{a})$ compared to smaller deviations. In contrast, in a least squares objective function, large deviations contribute quadratically, leading to significant sensitivity to outliers.

By minimizing the GLDM objective function, the influence of outliers?extreme values of $|y_t - \sum_{j=1}^{n(m)} a_j g_j(\{y_{t-k}\}_{k=1}^{m})|$ is effectively mitigated. The arctan function limits the impact of large deviations, ensuring that outliers do not disproportionately affect the estimation of the coefficients $\mathbf{a}$.

The Standard Cauchy distribution itself is known for its heavy tails, meaning that it assigns higher probabilities to extreme values compared to the normal distribution. This characteristic, when embedded in the GLDM via the CDF, further enhances the model's resilience to

outliers. The objective function incorporating the Standard Cauchy CDF implicitly reflects this robustness, as the heavy-tailed nature of the distribution aligns with the properties of the arctan function.

Thus, the use of the Standard Cauchy CDF in the GLDM objective function $O(\mathbf{a})$ results in a model that naturally diminishes the influence of outliers. The combination of the heavy-tailed Standard Cauchy distribution and the slowly increasing arctangent function confers robustness to the model, making it less sensitive to extreme deviations in the data. Therefore, the GLDM with this objective function is well-suited for applications where data may contain outliers, ensuring that the estimated model parameters more accurately represent the central tendency of the data. □

3 Description of the Forecasting Process

Figure 3 presents a schematic of a forecasting model that employs the Generalized Least Deviation Method (GLDM) to analyze historical time series data and predict future values. At the center of this process is the GLDM Estimator, responsible for deriving factors from the time series data. These factors, denoted as $\{a_j\}_{j=1}^{m}$, are then utilized by the Predictor module to forecast values over a horizon FH, providing a forward-looking perspective on the expected trends of the time series.

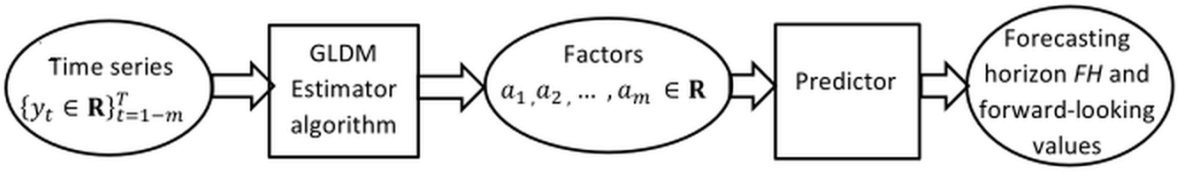

Figure 3: Schematic representation of the time series forecasting model employing the GLDM Estimator to compute the factors $\{a_j\}_{j=1}^{m}$, which are then used to predict future values over a forecasting horizon FH.

In Figure 4, the process flow within the GLDM time series forecasting model is depicted. Commencing with time series data, $\{y_t\}_{t=1}^{T}$, the flow advances through the construction of a model structure characterized by an m-th order. Subsequently, functions g_j are formulated to integrate historical data up to the m-th lag. Utilization of these functions leads to the computation of coefficients $\{a_j\}_{j=1}^{n(m)}$, which are integral to the model's framework. At the nucleus of the model is the objective function $F(a)$, the minimization of which is pivotal for determining the sum of arctangents of absolute deviations, thereby quantifying the model's congruence with the data.

Let us consider a single time series for one selected tile. For other tiles, the reasoning is similar up to the defined parameters.

Linear autoregressive models have a small forecasting horizon. The construction of adequate nonlinear models and/or neural networks may not be possible for technical reasons. Quasilinear models allow to increase the forecasting horizon. Let us implement our approach considered in Panyukov et al. (2023) to determine the coefficients $a_1, a_2, a_3 \dots, a_m \in \mathbb{R}$ of a m-th order quasilinear autoregressive model

$$y_t = \sum_{j=1}^{n(m)} a_j g_j(\{y_{t-k}\}_{k=1}^{m}) + \varepsilon_t, \quad t = 1, 2, \dots, T \tag{36}$$

by up-to-date information about of values of state variables $\{y_t \in \mathbb{R}\}_{t=1-m}^{T}$ at time instants t; here $g_j : (\{y_{t-k}\}_{k=1}^{m}) \to \mathbb{R}$, j=1,2,... n(m) are given $n(m)$ functions, and $\{\varepsilon_t \in \mathbb{R}\}_{t=1}^{T}$ are unknown errors.

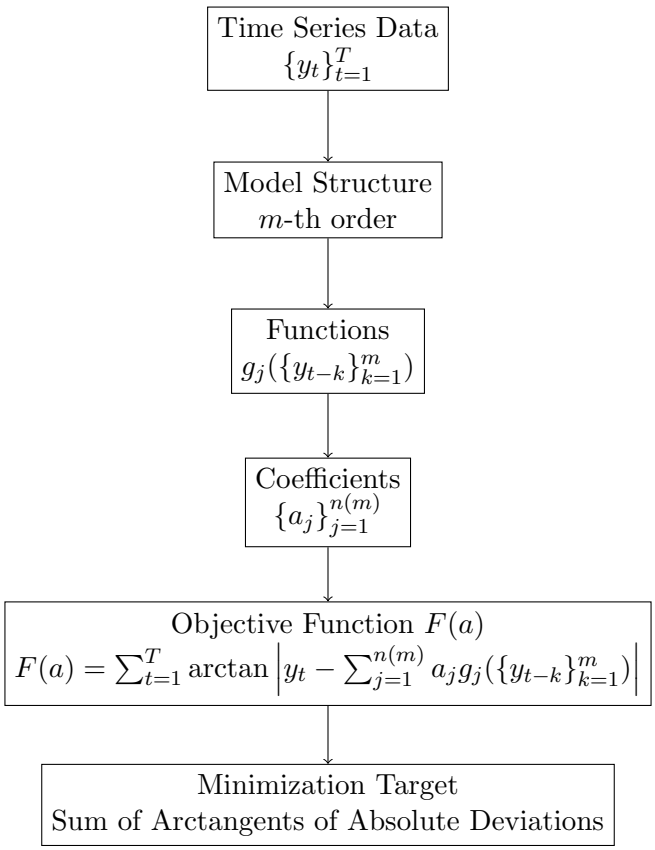

Figure 4: Schematic representation of the GLDM time series forecasting model process flow.

The considered approach consists in determining the parameters of the recurrence equation (36). The GLDM estimation algorithm Panyukov et al. (2023) gets a time series $\{y_t \in \mathbb{R}\}_{t=-1-m}^{T}$ of length $T+m \geq \left(1+3m+m^2\right)$ as an input data and determines the factors $a_1, a_2, a_3 \ldots, a_m \in \mathbb{R}$ by solving the optimization task

$$\left\{a_j^*\right\}_{j=1}^{n(m)} = \sum_{t=1}^{T} \arctan \left| \sum_{j=1}^{n(m)} a_j g_j(\{y_{t-k}\}_{k=1}^{m}) - y_t \right| \to \min_{\{a_j\}_{j-1}^{n(m)} \subset \mathbb{R}} \tag{37}$$

The Cauchy distribution

$$F(\xi) = \frac{1}{\pi} \arctan(\xi) + \frac{1}{2}$$

has the maximum entropy among distributions of random variables that have no mathematical expectation and variance. That's why function $\arctan(*)$ is applied as loss function.

Let's consider a m-th order model with quadratic nonlinearity. Then the basic set $g_j(*)$ may contain the following functions

$$\begin{aligned} g_{(k)}(\{y_{t-k}\}_{k=1}^{m}) &= y_{t-k}, \\ g_{(kl)}(\{y_{t-k}\}_{k=1}^{m}) &= y_{t-k} \cdot y_{t-l}, \\ k = 1, 2, \ldots, m;\ l &= k, k+1, \ldots, m. \end{aligned} \tag{38}$$

Obviously, in this case $n(m) = 2m + \mathrm{C}_m^2 = m(m+3)/2$, and the numbering of $g_{(*)}$ functions can be arbitrary. In particular, for $m = 2$ functions $g_{(*)}$ are the following

$$g_1 = y_1, \quad g_2 = y_2, \quad g_3 = y_1^2 \quad g_4 = y_1 \cdot y_2, \quad g_5 = y_2^2, .$$

The model for this case looks like following:

$$y_t = (a_1 y_{t-1} + a_2 y_{t-2}) + \left(a_3 y_{t-1}^2 + a_4 y_{t-1} y_{t-2} + a_5 y_{t-2}^2\right). \tag{39}$$

Predictor forms the indexed by $t = 1, 2, \ldots, T-1, T$ family of the m-th order difference equations

$$\overline{y[t]}_\tau = \sum_{j=1}^{n(m)} a_j^* g_j \left(\{\overline{y[t]}_{\tau-k}\}_{k=1}^m \right), \quad \tau = t, t+1, t+2, t+3, \ldots, T-1, T, T+1, \ldots \quad (40)$$

for lattice functions $\overline{y[t]}$ with values $\overline{y[t]}_\tau$ which interpreted as constructed at time moment t the forecasts for y_τ. Let us use the solution of the Cauchy problem for its difference equation (61) under the initial conditions

$$\overline{y[t]}_{t-1} = y_{t-1},\ \overline{y[t]}_{t-2} = y_{t-2},\ \ldots,\ \overline{y[t]}_{t-m} = y_{t-m} \quad t = 1, 2, \ldots, T-1, T \quad (41)$$

to find the values of the function $\overline{y[t]}$.

So we have the set $\overline{Y}_\tau = \left\{ \overline{y[t]}_\tau \right\}_{t=1}^T$ of possible prediction values of y_τ. Further we use this set to estimate the probabilistic characteristics of the y_τ value.

3.1 Evaluating by GLDM

Problem (37), i.e. problem of GLDM-estimation, is a multi-extremal optimization problem. GLDM-estimates are robust to the presence of a correlation of values in $\{y_t \in \mathbb{R}\}_{t=-1-m}^T$, and (with appropriate settings) are the best for probability distributions of errors with heavier (than normal distribution) tails (see Panyukov & Mezaal (2020)). All the above shows the feasibility of solving the identification problem (36) with usage solution (37). Let us use the interrelation between GLDM-estimates and estimates by the weighted least deviation method considered by Pan et al. (2007) (WLDM-estimates) to solve problems (37) of higher dimension.

Let us consider the algorithm of GLDM estimation (see Panyukov & Mezaal (2018)) in terms of this paper. First of all let us consider WLDM estimation algorithm used in GLDM algorithm.

The scheme of algorithm is shown in figure 5. Its input data are:

- $S = \{S_t \in \mathbb{R}^N\}_{t \in T}$, the matrix of a linear variety;
- $\nabla_{\mathcal{L}}$, gradient projection of objective function on $\mathcal{L}$;
- weight factors $\{p_t \in \mathbb{R}^+\}_{t=1}^T$;
- values of the given state variables $\{y_t \in \mathbb{R}^+\}_{t=1-m}^T$.

Algorithm runs as the iteration process for obtaining optimal GLDM solution $A \in \mathbb{R}^{n(m)}$ and the vector of residuals $z \in \mathbb{R}^T$. This process stops when $\left(A^{(k)} = A^{(k-1)}\right)$.

3.1.1 Evaluating by WLDM

Algorithm WLDM-estimator Makarovskikh & Abotaleb (2020) gets a time series $\{y_t \in \mathbb{R}\}_{t=1-m}^T$ and weight factors $\{p_t \in \mathbb{R}^+\}_{t=1}^T$ as an input data and calculates the factors

$$a_1, a_2, a_3 \ldots, a_{n(m)} \in \mathbb{R}$$

by solving the optimization problem

$$\sum_{t=1}^T p_t \cdot \left| \sum_{j=1}^{n(m)} a_j g_j(\{y_{t-k}\}_{k=1}^m) - y_t \right| \to \min_{\{a_j\}_{j=1}^{n(m)} \in \mathbb{R}^{n(m)}} \quad (42)$$

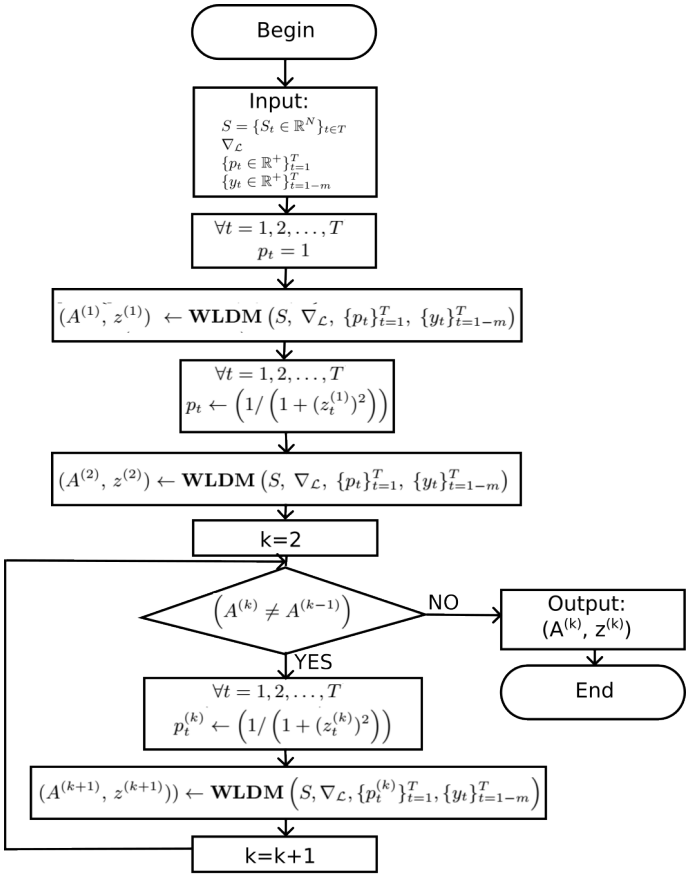

Figure 5: The scheme of GLDM estimation algorithm

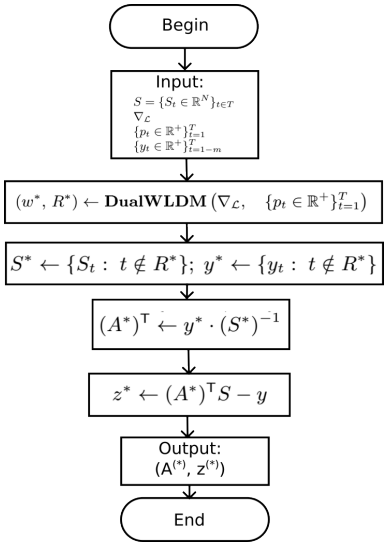

Figure 6: The scheme of WLDM estimation algorithm

The scheme of this algorithm is shown in figure 6. Computational complexity of such algorithm does not exceed $O(T^2)$ due to the simple structure of the admissible set: intersection of T-dimensional cuboid (54) and $(T - n(m))$-dimensional linear variety (53).

Algorithm for dual task (52)–(54) solution begins the search of the optimal solution at 0, moving along direction $\nabla_{\mathcal{L}}$. If the current point falls on the face of brick $\mathcal{T}$, then the corresponding coordinate in the direction of the moving is assumed to be 0.

If (w^*, R^*) is the result of executing the gradient projection algorithm Panyukov et al. (2023), then w^* is the optimal solution to the task (52)-(54), and the optimal solution of the task (49)-(51) is equal to

$$u_t^* = \frac{p_t + w_t^*}{2}, \quad v_t^* = \frac{p_t - w_t^*}{2}, \quad t = 1, 2, \ldots, T.$$

It is following from the complementarity condition for a pair of mutually dual tasks (46)–(48) and (49)–(51) that

$$y_t = \sum_{j=1}^{n(m)} [a_j g_j(\{y_{t-k}\}_{k=1}^m)] \qquad \forall t \notin R^*, \tag{43}$$

$$y_t = \sum_{j=1}^{n(m)} [a_j g_j(\{y_{t-k}\}_{k=1}^m)] + z_t^*, \quad \forall t \in R^* : w_t^* = p_t, \tag{44}$$

$$y_t = \sum_{j=1}^{n(m)} [a_j g_j(\{y_{t-k}\}_{k=1}^m)] - z_t^*, \quad \forall t \in R^* : w_t^* = -p_t. \tag{45}$$

In fact, the solution $(\{a_j^*\}_{j-1}^{n(m)}, z^*)$ of linear algebraic equations system (55)–(57) represents the dual optimal solution of task (52)-(54) and the optimal solution of the task (42), that proves the validity of the following theoremPanyukov et al. (2022).

Theorem 5. *Let the following be given:*

- w^* *denotes the optimal solution of the optimization problem defined by equations* (52)–(54).
- $(\{a_j^*\}_{j=1}^{n(m)}, z^*)$ *denotes the solution of the system of linear algebraic equations given by* (55)–(57).

Then, the set of coefficients $\{a_j^*\}_{j=1}^{n(m)}$ *is the optimal solution to the optimization problem defined in* (42).

The main problem with the use of WLDM-estimator is the absence of general formal rules for choosing weight coefficients. Consequently, this approach requires additional research.

This problem represents the problem of convex piecewise linear optimization, and the introduction of additional variables reduces it to the problem of linear programming

The task (37), i.e. task of GLDM-estimation, is a concave optimization problem, and entering the additional variables reduces it to the following linear programming task

$$\sum_{t=1}^{T} p_t z_t \to \min_{\substack{(a_1, a_2, \ldots, a_{n(m)}) \in \mathrm{R}^m, \\ (z_1, z_2, \ldots, z_T) \in \mathrm{R}^T}} \tag{46}$$

$$-z_t \le \sum_{j=1}^{n(m)} [a_j g_j(\{y_{t-k}\}_{k=1}^m)] - y_t \le z_t, \quad t = 1, 2, \ldots, T, \tag{47}$$

$$z_t \ge 0, \quad t = 1, 2, \ldots, T. \tag{48}$$

The task (46)–(48) has a canonical type with variables $n(m)+T$ and $3n$ inequality constraints including the conditions of non-negativity of z_j, $j=1,2,\ldots,T$.

The dual to (46) task is

$$\sum_{t=1}^{T}(u_t-v_t)\,y_t \to \max_{u,v\in\mathbb{R}^T}, \tag{49}$$

$$\sum_{t=1}^{T} a_j g_j(\{y_{t-k}\}_{k=1}^{m})\,(u_t-v_t)=0,\ j=1,2,\ldots,n(m), \tag{50}$$

$$u_t+v_t=p_t,\quad u_t,\,v_t\geq 0,\quad t=1,2,\ldots,T. \tag{51}$$

Let us introduce variables $w_t=u_t-v_t,\ t=1,2,\ldots,T$. Conditions (51) imply that

$$u_t=\frac{p_t+w_t}{2},\quad v_t=\frac{p_t-w_t}{2},\quad -p_t\leq w_t\leq p_t,\quad t=1,2,\ldots,T.$$

So the optimal task (49)–(51) solution is equal to the optimal solution of task

$$\sum_{t=1}^{T} w_t\cdot y_t \to \max_{w\in\mathbb{R}^T}, \tag{52}$$

$$\sum_{t=1}^{T} g_j(\{y_{t-k}\}_{k=1}^{m})\cdot w_t=0,\ j=1,2,\ldots,n(m), \tag{53}$$

$$-p_t\leq w_t\leq p_t,\ t=1,2,\ldots,T. \tag{54}$$

Constraints (53) define $(T-n(m))$-dimensional linear variety $\mathcal{L}$ with $(n(m)\times T)$-matrix

$$S=\begin{bmatrix} g_1(\{y_{1-k}\}_{k=1}^{m}) & g_1(\{y_{2-k}\}_{k=1}^{m}) & \cdots & g_1(\{y_{T+1-k}\}_{k=1}^{m}) \\ g_2(\{y_{1-k}\}_{k=1}^{m}) & g_2(\{y_{2-k}\}_{k=1}^{m}) & \cdots & g_2(\{y_{T+1-k}\}_{k=1}^{m}) \\ \vdots & \vdots & \ddots & \vdots \\ g_{n(m)}(\{y_{1-k}\}_{k=1}^{m}) & g_{n(m)}(\{y_{1-k}\}_{k=1}^{m}) & \cdots & g_{n(m)}(\{y_{1-k}\}_{k=1}^{m}) \end{bmatrix}$$

Constraints (54) define T-dimensional parallelepiped $\mathcal{T}$.

The simple structure of the allowed set for task (52)–(54) representing the intersection of $(T-n(m))$-dimensional linear variety $\mathcal{L}$ (53) and T-dimensional parallelepiped $\mathcal{T}$ (54) allows to obtain its solution by algorithm using the gradient projection of the objective function (52) (i.e. vector $\nabla=\{y_t\}_{t=1}^{T}$) on the allowed area $\mathcal{L}\cap\mathcal{T}$ defined by the constraints (53)–(54). The projection matrix on $\mathcal{L}$ is as following

$$S_{\mathcal{L}}=E-S^T\cdot\left(S\cdot S^T\right)^{-1}\cdot S,$$

and gradient projection on $\mathcal{L}$ is equal to $\nabla_{\mathcal{L}}=S_{\mathcal{L}}\cdot\nabla$. Moreover, if outer normal on any parallelepiped face forms the sharp corner with gradient projection $\nabla_{\mathcal{L}}$ then movement by this face is equal to zero.

DualWLDMSoluter Algorithm 1 to solve problem (52)-(54) begins the search of the optimal solution at 0, moving along direction $\nabla_{\mathcal{L}}$. If the current point falls on the face of brick $\mathcal{T}$, then the corresponding coordinate in the direction of the moving is assumed to be 0.

Computational complexity of such algorithm does not exceed $O(T^2)$ due to the simple structure of the admissible set: intersection of T-dimensional cuboid (54) and $(T-n(m))$-dimensional linear subspace (53).

If $(w^*,\,R^*)$ is the result of executing the Algorithm 1, then w^* is the optimal solution to the problem (52)-(54), and the optimal solution of the problem (49)-(51) is equal to

$$u_t^*=\frac{p_t+w_t^*}{2},\quad v_t^*=\frac{p_t-w_t^*}{2},\quad t=1,2,\ldots,T.$$

Algorithm 1: DualWLDMSolver

Require: ∇_L ; // Gradient projection

1 $\{p_t \in \mathbb{R}^+\}_{t=1}^T$; // Weight factors

Ensure : $w^* = \arg\max_{w \in \mathbb{R}^T} \sum_{i=1}^T w_i \cdot y_i$; // Optimal dual solution

2 Initialize $w \leftarrow \{w_i = 0 \mid i = 1, 2, \ldots, T\}$; // Initialize weights

3 ;

4 Initialize $R \leftarrow \emptyset$; // Initialize active set

5 ;

6 Initialize $g \leftarrow \nabla_L$; // Initialize gradient

7 ;

8 **while** $\alpha_* \neq 0$ **do**

9 $(\alpha_*, t_*) \leftarrow \arg\max \{\alpha \geq 0 \mid -p_t \leq w_t + \alpha g_t \leq p_t\}$;

10 $w \leftarrow w + \alpha_* g$;

11 $g_{t_*} \leftarrow 0$;

12 $R \leftarrow R \cup \{t_*\}$;

13 $w^* \leftarrow w$;

14 $R^* \leftarrow R$;

15 **return** (w^*, R^*);

It is following from the complementarity condition for a pair of mutually dual problems (46)-(48) and (49)-(51) that

$$y_t = \sum_{j=1}^{n(m)} [a_j g_j(\{y_{t-k}\}_{k=1}^m)] \qquad \forall t \notin R^*, \tag{55}$$

$$y_t = \sum_{j=1}^{n(m)} [a_j g_j(\{y_{t-k}\}_{k=1}^m)] + z_t^*, \quad \forall t \in R^* : w_t^* = p_t, \tag{56}$$

$$y_t = \sum_{j=1}^{n(m)} [a_j g_j(\{y_{t-k}\}_{k=1}^m)] - z_t^*, \quad \forall t \in R^* : w_t^* = -p_t. \tag{57}$$

In fact, the solution $(\{a_j^*\}_{j-1}^{n(m)}, z^*)$ of linear algebraic equations system (55)-(57) represents the dual optimal solution of problem (52)-(54) and the optimal solution of the problem (42), that proves the validity of the following theorem.

Theorem 6. *Let w^* be the optimal solution to the problem defined by equations* (52)–(54). *Additionally, let $(\{a_j^*\}_{j=1}^{n(m)}, z^*)$ be the solution to the system of linear algebraic equations given by* (55)–(57). *Then, the set of coefficients $\{a_j^*\}_{j=1}^{n(m)}$ is the optimal solution to the problem defined in* (42).

The above allows us to propose WLDM-estimator Algorithm 2.

The main problem with the use of WLDM-estimator is the absence of general formal rules for choosing weight coefficients. Consequently, this approach requires additional research.

The established in Panyukov & Mezaal (2018), Panyukov et al. (2018) results allow us to reduce the problem of determining GLDM estimation to an iterative procedure with WLDM estimates.

3.1.2 GLDM estimation algorithm

Problem (37) of GLDM estimation is a concave optimization problem. GLDM-estimates are robust to the presence of a correlation of values in $\{X_{j\,t} : \ t = 1, 2, \ldots, T; \ j = 1, 2, \ldots, N\}$,

Algorithm 2: WLDM-Estimator

Require: $S = \{S_t \in \mathbb{R}^N\}_{t\in T}$; // Matrix of linear subspace $\mathcal{L}$
$\nabla_{\mathcal{L}}$; // Gradient projection on $\mathcal{L}$
$\{p_t \in \mathbb{R}^+\}_{t=1}^T$; // Weight factors
$\{y_t \in \mathbb{R}^+\}_{t=1}^T$; // Values of the given state variables
Ensure : $A^* \in \mathbb{R}^{n(m)}$; // Optimal primal solution
$z^* \in \mathbb{R}^T$; // Restrictions
$(w^*, R^*) \leftarrow$ **DualWLDMSolver** $\left(\nabla_{\mathcal{L}}, \{p_t\}_{t=1}^T\right)$;
$S^* \leftarrow \{S_t \mid t \notin R^*\}$; // System (55) matrix
$y^* \leftarrow \{y_t \mid t \notin R^*\}$; // System (55) values
$(A^*)^{\mathsf{T}} \leftarrow y^* \cdot (S^*)^{-1}$; // Solution to System (55)
$z^* \leftarrow (A^*)^{\mathsf{T}} S - y$; // Find restrictions
return (A^*, z^*);

and (with appropriate settings) like the best for probability distributions of errors with heavier (than normal distribution) tails Panyukov & Mezaal (2020). The above shows the feasibility of solving the identification problem (36) by Algorithm (2). The established in Panyukov & Mezaal (2018) results allow us to reduce the problem of determining GLDM estimation to an iterative procedure with WLDM estimates (see Algorithm 3).

Algorithm 3: GLDM-Estimator

Require: $S = \{S_t \in \mathbb{R}^N\}_{t\in T}$; // Matrix of linear subspace $\mathcal{L}$
$\nabla_{\mathcal{L}}$; // Gradient projection on $\mathcal{L}$
$\{p_t \in \mathbb{R}^+\}_{t=1}^T$; // Weight factors
$\{y_t \in \mathbb{R}^+\}_{t=1}^T$; // Values of the given state variables
Ensure : $A^* \in \mathbb{R}^{n(m)}$; // Optimal GLDM solution
$z^* \in \mathbb{R}^T$; // Residuals
$p \leftarrow \{p_t = 1 \mid t = 1, 2, \ldots, T\}$; // Initialize weight factors
$(A^{(1)}, z^{(1)}) \leftarrow$ **WLDMSolver** $\left(S, \nabla_{\mathcal{L}}, \{p_t\}_{t=1}^T, \{y_t\}_{t=1}^T\right)$;
for $t \leftarrow 1$ **to** T **do**
  $p_t \leftarrow \dfrac{1}{1 + (z_t^{(1)})^2}$; // Update weight factors
$(A^{(2)}, z^{(2)}) \leftarrow$ **WLDMSolver** $\left(S, \nabla_{\mathcal{L}}, \{p_t\}_{t=1}^T, \{y_t\}_{t=1}^T\right)$;
$k \leftarrow 2$; // Initialize iteration counter
while $A^{(k)} \neq A^{(k-1)}$ **do**
  for $t \leftarrow 1$ **to** T **do**
    $p_t^{(k)} \leftarrow \dfrac{1}{1 + (z_t^{(k)})^2}$; // Update weight factors
  $(A, z) \leftarrow$ **WLDMSolver** $\left(S, \nabla_{\mathcal{L}}, \{p_t^{(k)}\}_{t=1}^T, \{y_t\}_{t=1}^T\right)$;
  $(A^{(k+1)}, z^{(k+1)}) \leftarrow (A, z)$; // Store current solution
  $k \leftarrow k + 1$; // Increment iteration counter
$z^* \leftarrow z^{(k)}$;
$A^* \leftarrow A^{(k)}$; // Final GLDM solution
return (A^*, z^*);

Theorem 7. *Let $\{(A^{(k)}, z^{(k)})\}_{k=1}^{\infty}$ be the sequence generated by the GLDM-estimator algorithm. This sequence converges to the global minimum (a^*, z^*) of the problem* (37).

The GLDM-estimator algorithm has a computational complexity that is proportional to the

complexity of solving the primal and/or dual WLDM problems (42). Multiple computational experiments suggest that the average number of iterations of the GLDM-estimator algorithm is approximately equal to the number of coefficients in the identified equation. If this hypothesis holds true, then the computational complexity for solving practical problems does not exceed

$$O(n(m)^3T + n(m) \cdot T^2).$$

Proof. The convergence of the GLDM-estimator algorithm can be demonstrated by considering the properties of the sequence generated by the algorithm.

From the concavity and differentiability of the function arctan, we have the inequality:

$$\arctan|z_t| > \arctan|z| - \frac{|z|}{1+z_t^2} + \frac{|z_t|}{1+z_t^2}, \quad \text{for } z \neq z_t. \tag{58}$$

Consider two consecutive iterations, $\mathcal{P}_k$ and $\mathcal{P}_{k+1}$, of the problem (46)–(48), as solved by the GLDM-estimator algorithm (see line 13 of Algorithm 3). These iterations differ only in their sets of weight coefficients. The optimal solution for $\mathcal{P}_k$, while correct for that iteration, may not be optimal for $\mathcal{P}_{k+1}$. We have:

$$\sum_{t=1}^{T} \frac{|z_t^{(k)}|}{1+(z_t^{(k)})^2} \geq \sum_{t=1}^{T} \frac{|z_t^{(k+1)}|}{1+(z_t^{(k)})^2}, \tag{59}$$

since the weights in $\mathcal{P}_{k+1}$ are defined as

$$\left\{ p_t^{(k)} = \frac{1}{1+(z_t^{(k)})^2} \right\} \quad \text{(see lines 9–12 of Algorithm 3).}$$

From (58) and (59), it follows that:

$$\sum_{t=1}^{T} \arctan|z_t^{(k)}| \geq \sum_{t=1}^{T} \arctan|z_t^{(k+1)}|. \tag{60}$$

Thus, the sequence

$$\left\{ \sum_{t=1}^{T} \arctan \left| y_t - \sum_{j=1}^{n(m)} a_j^{(k)} g_j\{y_{t-k}\}_{k=1}^{m} \right| \right\}_{k=0,1,\ldots}$$

monotonically decreases and is bounded below by zero. Therefore, this sequence must converge to a unique limit point. The existence of a limit point for the sequence

$$\{a_j^{(k)}\}_{j=1}^{n(m)}, \quad k = 1, 2, \ldots,$$

follows from the continuity and monotonicity of the function $\arctan(\cdot)$.

To prove that the limit point $(a_1^*, a_2^*, \ldots, a_{n(m)}^*)$, constructed by Algorithm 3, is indeed the global minimum, assume the contrary. Suppose that $\tilde{a} = (\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_{n(m)})$ is the global minimum of the loss function (37), and

$$\sum_{t=1}^{T} \arctan|\tilde{z}_t| < \sum_{t=1}^{T} \arctan|z_t^*|.$$

Clearly, $(\tilde{a}, \tilde{z})$ is some limit point of the sequence constructed by Algorithm 3. From (58), for $z^* \neq \tilde{z}$, it follows that:

$$\arctan|\tilde{z}_t| - \arctan|z_t^*| > \frac{|\tilde{z}_t|}{1+\tilde{z}_t^2} - \frac{|z_t^*|}{1+\tilde{z}_t^2}.$$

On the other hand:

$$\arctan|\tilde{z}_t| - \arctan|z_t^*| = \frac{|\tilde{z}_t| - |z_t^*|}{1 + |\tilde{z}_t| \cdot |z_t^*|}, \quad t = 1, 2, \ldots, T.$$

Therefore:

$$\frac{|\tilde{z}_t| - |z_t^*|}{1 + |\tilde{z}_t| \cdot |z_t^*|} > \frac{|\tilde{z}_t| - |z_t^*|}{1 + \tilde{z}_t^2}.$$

Consequently:

$$0 < \frac{|\tilde{z}_t| - |z_t^*|}{1 + |\tilde{z}_t| \cdot |z_t^*|} - \frac{|\tilde{z}_t| - |z_t^*|}{1 + \tilde{z}_t^2} = (|\tilde{z}_t| - |z_t^*|) \cdot \left(\frac{1}{1 + |\tilde{z}_t| \cdot |z_t^*|} - \frac{1}{1 + \tilde{z}_t^2} \right)$$
$$= \frac{(|\tilde{z}_t| - |z_t^*|)^2 \cdot |\tilde{z}_t|}{(1 + \tilde{z}_t^2) \cdot (1 + (z_t^*)^2)}, \quad t = 1, 2, \ldots, T.$$

This implies that $|\tilde{z}_t| > 0$ for all $t = 1, 2, \ldots, T$, which contradicts the assumption that $(a_1^*, a_2^*, \ldots, a_{n(m)}^*)$ is the global minimum.

Thus, by contradiction, the limit point $(a_1^*, a_2^*, \ldots, a_{n(m)}^*)$ constructed by the algorithm is indeed the global minimum.

The theorem is proved.

□

3.2 Predictor

Theorem 8. *Let $\overline{y[t]}_\tau$ denote the predicted value of the variable y_τ at time t generated by the predictor algorithm 4, which constructs a family of m-th order difference equations. The sequence of predicted values $\{\overline{y[t]}_\tau\}_{t=1}^T$ converges to a stable prediction horizon, and the prediction error $|y_\tau - \overline{y[t]}_\tau|$ is minimized over this horizon.*

Proof. The predictor constructs predictions using the m-th order difference equation, defined as:

$$\overline{y[t]}_\tau = \sum_{j=1}^{n(m)} a_j g_j \left(\{y_{t-k}\}_{k=1}^m \right), \tag{61}$$

where: - $\overline{y[t]}_\tau$ is the predicted value of y_τ at time t, - $\{a_j\}_{j=1}^{n(m)}$ are the coefficients derived from the WLDM model, and - $g_j(\cdot)$ are functions representing the relationship between past observations $\{y_{t-k}\}_{k=1}^m$ and the predicted value.

This equation models the forecast as a linear combination of previous values weighted by the coefficients a_j.

The predictor algorithm begins by setting the initial conditions:

$$\overline{y[1]}_\tau = y_1, \quad \overline{y[2]}_\tau = y_2, \quad \ldots, \quad \overline{y[m]}_\tau = y_m, \tag{62}$$

where $y_1, y_2, \ldots, y_m$ are the known initial values of the time series.

For $t > m$, the predictor uses the recursive relation:

$$\overline{y[t]}_\tau = \sum_{j=1}^{n(m)} a_j g_j \left(\{y_{t-k}\}_{k=1}^m \right), \tag{63}$$

to calculate each subsequent predicted value $\overline{y[t]}_\tau$. The recursion ensures that each new prediction is based on the most recent m values.

We now analyze the convergence of the sequence $\{\overline{y[t]}_\tau\}_{t=1}^T$. Define the prediction error at each step as:

$$e[t] = |y_\tau - \overline{y[t]}_\tau|. \tag{64}$$

The algorithm iteratively updates the prediction sequence until the error $e[t]$ satisfies a predefined threshold SZ:

$$e[t] \leq SZ. \tag{65}$$

If the error at any step t exceeds SZ, the algorithm adjusts the coefficients or updates the prediction horizon, ensuring that the error decreases in subsequent steps.

We next show that the sequence $\{e[t]\}_{t=1}^{T}$ is bounded and monotonic.

The error $e[t]$ is bounded because the arctangent function $\arctan(x)$ used in the prediction model has a bounded derivative:

$$\left|\frac{d}{dx}\arctan(x)\right| = \frac{1}{1+x^2} \leq 1. \tag{66}$$

Since the update of $\overline{y[t]}_\tau$ depends on the bounded function $\arctan(x)$, the changes in predictions are also bounded. Consequently, the error $e[t]$ cannot grow unbounded and remains within a controlled range.

Monotonicity follows from the algorithm's construction, where each update step is designed to reduce the error $e[t]$. Specifically, the recursive updates minimize the prediction error over time by adjusting the coefficients $\{a_j\}$ and recalculating $\overline{y[t]}_\tau$ iteratively.

Consider two successive predictions $\overline{y[t]}_\tau$ and $\overline{y[t+1]}_\tau$. The error difference is given by:

$$e[t+1] - e[t] = \left(|y_\tau - \overline{y[t+1]}_\tau|\right) - \left(|y_\tau - \overline{y[t]}_\tau|\right). \tag{67}$$

Since the algorithm updates $\overline{y[t+1]}_\tau$ based on minimizing this difference, the sequence $\{e[t]\}$ is non-increasing:

$$e[t+1] \leq e[t]. \tag{68}$$

The total prediction error over the prediction horizon minFH is defined as:

$$E = \sum_{t=1}^{\text{minFH}} \left(y_t - \overline{y[t]}_\tau\right)^2. \tag{69}$$

The algorithm seeks to minimize E by iteratively refining the predictions and adjusting the prediction horizon. The process ensures that:

$$\text{minFH} = \arg\min_{H} E, \tag{70}$$

where H represents different candidate horizons, and minFH is the minimal feasible horizon at which the error E is minimized.

The recursive nature of the algorithm guarantees that each prediction step leads to a reduction in E until convergence is achieved.

The sequence $\{\overline{y[t]}_\tau\}_{t=1}^{T}$ generated by the predictor algorithm converges to a stable horizon minFH, where the prediction error E is minimized. This convergence is guaranteed by the boundedness and monotonicity of the error sequence $\{e[t]\}$. Thus, the predictor algorithm effectively forecasts the future values of the time series with minimized prediction error, proving the theorem. □

Predictor forms the indexed by $t = 1, 2, \ldots, T-1, T$ family of the m-th order difference equations (61) for lattice functions $\overline{y[t]}$ with values $\overline{y[t]}_\tau$ that interpreted as constructed at time moment t the forecast for y_τ. Let us use the solution of the Cauchy problem for its difference equation (61) under the initial conditions (62) to find the values of the function $\overline{y[t]}$. So we have the set $\overline{Y}_\tau = \left\{\overline{y[t]}_\tau\right\}_{t=1}^{T}$ of possible prediction values of y_τ. Further we use this set to estimate the probabilistic characteristics of the y_τ value. It should be written as Algorithm 4.

Algorithm 4: Predictor

Require: $Y = \{y_t \in \mathbb{R}^+\}_{t=1}^{T}$; // Values of the given state variables

$A = \{a_i\}_{i=1}^{n(m)}$; // WLDM solution

Ensure : $PY[1:T][1:T]$: $PY[t][\tau] = \overline{y[t]}_\tau$; // Forecast for y_τ at time moment t

E ; // Average prediction errors

D ; // Average absolute prediction errors

$minFH$; // Relabel prediction horizon

Initialize variables;

while $FH[Strt] < m$ **do**

  Strt $\leftarrow$ Strt + 1;

  PY[Strt][0] $\leftarrow Y$[Strt];

  PY[Strt][1] $\leftarrow Y$[Strt + 1];

  for $t \leftarrow Strt + 2$ **to** m **do**

    $py \leftarrow 0$;

    for $j \leftarrow 0$ **to** n **do**

      $A1 \leftarrow G[j](PY[\text{Strt}][t-1], PY[\text{Strt}][t-2])$;

      $R \leftarrow a[j] \cdot A1$;

      $py \leftarrow py + R$;

    PY[Strt][t] $\leftarrow py$;

    if $|PY[Strt][t] - Y[Strt + t]| > SZ$ **then**

      break;

  FH[Strt] $\leftarrow t$;

LastStrt $\leftarrow t$;

$minFH \leftarrow FH$[Strt];

minFHp $\leftarrow minFH$;

for $t \leftarrow 3$ **to** $Strt$ **do**

  if $minFH > FH[t]$ **then**

    minFHp $\leftarrow FH[t]$;

$minFH \leftarrow \min(\text{minFHp}, minFH)$;

$E \leftarrow 0$;

$D \leftarrow 0$;

▷ minFHp is the reasonable horizon;

for $t \leftarrow 3$ **to** $minFH$ **do**

  $D \leftarrow D + \text{fabs}(Y[t + \text{Strt}] - PY[\text{Strt}][t])$;

  $E \leftarrow E + (Y[t + \text{Strt}] - PY[\text{Strt}][t])$;

$D \leftarrow D/minFH$;

$E \leftarrow E/minFH$;

return $(D, E, minFH)$;

4 Description Model coefficients

A time series forecasting model employing the Generalized Least Deviation Method (GLDM) is considered. The time series dataset is characterized as follows:

$$\{y_t\}_{t=1}^{T} \subset \mathbb{R}, \tag{71}$$

where y_t denotes a real-valued observation at time index t.

The GLDM Estimator is utilized to determine an optimal set of coefficients $\{a_j\}_{j=1}^{n(m)}$, which minimize the objective function $F(\mathbf{a})$, defined as the sum of the arctangents of absolute deviations:

$$F(\mathbf{a}) = \sum_{t=1}^{T} \arctan \left| y_t - \sum_{j=1}^{n(m)} a_j g_j(\{y_{t-k}\}_{k=1}^{m}) \right|, \tag{72}$$

with each function g_j representing a unique combination of preceding values up to the m-th order.

Within the quasi-linear model framework, the functions g_j capture the influence of historical data. These functions are defined as follows:

$$g_j(\{y_{t-k}\}_{k=1}^{m}) = y_{t-j} + \sum_{\substack{p=1 \\ p \neq j}}^{m} y_{t-j} \cdot y_{t-p} + \sum_{p=1}^{m} y_{t-p}^2, \tag{73}$$

where y_{t-j} signifies the lagged value of the series at time $t-j$. The first summation models the interaction effects between different lagged values, while the second summation encapsulates the non-linear effects through squared terms of the lagged values. These elements allow for the modeling of complex dynamics within time series data.

The total count of coefficients for an m-th order model, which includes linear, interaction, and quadratic components, is described by the following expression:

$$n(m) = m + \binom{m}{2} + m = \frac{m(m+3)}{2}. \tag{74}$$

The structure and roles of these coefficients in the modeling process are detailed as follows:

- The term m refers to the linear coefficients, correlating each historical value with the subsequent predicted value.
- The term $\binom{m}{2}$ represents the interaction coefficients, denoting the pairwise combinations between historical values, facilitating the detection of dependencies and interactions at different time lags.
- The final term m denotes the quadratic coefficients, accommodating non-linear trends by reflecting the self-interactions of the historical values.

4.1 First-Order Time Series Forecasting Model

A time series forecasting model incorporating only linear terms up to the first order is considered. For a model where $m = 1$, it is determined that the total number of coefficients is two. These coefficients are crucial for modeling the linear tendencies within the time series data.

For a first-order model ($m = 1$), the coefficients and their corresponding terms are enumerated as follows:

- Linear term: y_{t-1}
- Squared term: y_{t-1}^2

The generalized function g_j for this model is explicitly defined in the following manner:

$$g_j(\{y_{t-k}\}_{k=1}^{m}) = \begin{cases} y_{t-1} & \text{for } j = 1, \\ y_{t-1}^2 & \text{for } j = 2. \end{cases}$$

In this configuration, g_1 is assigned to the linear term, and g_2 to the squared term. This structural arrangement effectively accounts for both the direct influence of the past value and its non-linear impact on the present value.

4.2 Second-Order Time Series Forecasting Model

A time series forecasting model that incorporates interaction and non-linear terms up to the second order is considered. For a model where $m = 2$, it is determined that the total number of coefficients is five. These coefficients comprise the linear terms, their squares, and the interaction term between them, which are critical for modeling the linear tendencies and capturing the potential synergistic and quadratic effects within the time series data.

For a second-order model ($m = 2$), the coefficients and their corresponding terms are enumerated as follows:

- Linear terms: y_{t-1}, y_{t-2}
- Squared terms: y_{t-1}^2, y_{t-2}^2
- Interaction term: $y_{t-1} \cdot y_{t-2}$

The generalized function g_j for this model is explicitly defined in the following manner:

$$g_j(\{y_{t-k}\}_{k=1}^m) = \begin{cases} y_{t-j} & \text{for } j = 1, 2, \\ y_{t-1} \cdot y_{t-2} & \text{for } j = 3, \\ y_{t-j+2}^2 & \text{for } j = 4, 5. \end{cases} \tag{75}$$

In this configuration, g_1 and g_2 are assigned to the first and second linear terms, respectively, g_3 to the interaction term, and g_4 and g_5 to the squared terms of the first and second variables, respectively. This structural arrangement effectively accounts for both the direct influences and the interactions between the past two values, as well as their individual non-linear influences on the present value.

4.3 Third-Order Time Series Forecasting Model

Interactions between past values in a time series are essential for capturing the dynamics and dependencies inherent within the data. In a third-order model, denoted by $m = 3$, linear, squared, and interaction terms are included, facilitating the modeling of complex non-linear behaviors. This comprehensive approach enables the effective capture of intricacies and interdependencies among historical values.

For a third-order model ($m = 3$), the coefficients and their corresponding terms are outlined as follows:

- Linear terms: y_{t-1}, y_{t-2}, y_{t-3}
- Squared terms: y_{t-1}^2, y_{t-2}^2, y_{t-3}^2
- Interaction terms: $y_{t-1} \cdot y_{t-2}$, $y_{t-1} \cdot y_{t-3}$, $y_{t-2} \cdot y_{t-3}$

The functions g_j representing these terms in the third-order model are systematically defined as follows:

$$g_j(\{y_{t-k}\}_{k=1}^m) = \begin{cases} y_{t-j} & \text{for } j = 1, 2, 3, \\ y_{t-j+3}^2 & \text{for } j = 4, 5, 6, \\ y_{t-1} \cdot y_{t-2} & \text{for } j = 7, \\ y_{t-1} \cdot y_{t-3} & \text{for } j = 8, \\ y_{t-2} \cdot y_{t-3} & \text{for } j = 9. \end{cases} \tag{76}$$

This model structure, incorporating linear, squared, and interaction terms, ensures a robust representation of the time series dynamics. The inclusion of these terms aids in modeling more complex nonlinear relationships that linear terms alone may not capture.

The complete mathematical model of the time series, utilizing the coefficients defined above, is given by the following equation:

$$y_t = \sum_{j=1}^{9} a_j g_j(\{y_{t-k}\}_{k=1}^{m}) + \varepsilon_t, \quad t = 1, 2, \ldots, T \tag{77}$$

Here, ε_t denotes the error term at time t, representing the unpredictable component not explained by the model.

4.4 Fourth-Order Time Series Forecasting Model

A time series forecasting model that leverages the intricacies of linear, interaction, and non-linear dynamics up to the fourth order is considered. In a fourth-order model, denoted by $m = 4$, the total count of coefficients is identified as 14. This ensemble encompasses the linear terms for the four preceding observations, their squared counterparts, and the six unique interaction terms between these observations, thereby encapsulating a comprehensive dynamic range within the time series.

For a fourth-order model ($m = 4$), the coefficients and their corresponding terms are explicitly associated as follows:

- Linear terms: y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}
- Squared terms: y_{t-1}^2, y_{t-2}^2, y_{t-3}^2, y_{t-4}^2
- Interaction terms: All distinct pairwise combinations of the four variables

With 4 linear and 4 squared terms, and $\binom{4}{2} = 6$ interaction terms, the model integrates a total of 14 coefficients.

The specific formulation of the function g_j within the model, which embraces both linear and interaction terms, is systematically defined as follows:

$$g_j(\{y_{t-k}\}_{k=1}^{m}) = \begin{cases} y_{t-j} & \text{for } j = 1, 2, 3, 4, \\ y_{t-1} \cdot y_{t-2} & \text{for } j = 5, \\ y_{t-1} \cdot y_{t-3} & \text{for } j = 6, \\ y_{t-1} \cdot y_{t-4} & \text{for } j = 7, \\ y_{t-2} \cdot y_{t-3} & \text{for } j = 8, \\ y_{t-2} \cdot y_{t-4} & \text{for } j = 9, \\ y_{t-3} \cdot y_{t-4} & \text{for } j = 10, \\ y_{t-j+6}^2 & \text{for } j = 11, 12, 13, 14. \end{cases} \tag{78}$$

In this model, g_1 through g_4 correspond to the linear terms, g_5 through g_{10} to the interaction terms, and g_{11} through g_{14} to the squared terms. This elaborate model configuration facilitates an extensive incorporation of both the progressive and the interactive effects of the past observations, along with their individual non-linear influences, thus significantly augmenting the predictive capabilities of the time series model.

4.5 Fifth-Order Time Series Forecasting Model

A comprehensive time series forecasting model that integrates both linear and nonlinear dynamics up to the fifth order is considered. Within a fifth-order framework, symbolized by $m = 5$, a constellation of 20 coefficients is identified. These coefficients comprise the linear terms for the five antecedent observations, their individual squared terms, and the interaction terms among these observations, thereby capturing a multidimensional dynamic within the time series.

For a fifth-order model ($m = 5$), the assortment of coefficients is meticulously associated with their respective terms as cataloged below:

- Linear terms: y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5}
- Squared terms: y_{t-1}^2, y_{t-2}^2, y_{t-3}^2, y_{t-4}^2, y_{t-5}^2
- Interaction terms: All distinct pairwise combinations of the five variables

Accounting for 5 linear terms, 5 squared terms, and $\binom{5}{2} = 10$ interaction terms, the model features an aggregate of 20 coefficients.

The formalized expression of the function g_j within the model, encapsulating both the linear and interaction terms, is articulated as follows:

$$g_j(\{y_{t-k}\}_{k=1}^m) = \begin{cases} y_{t-j} & \text{for } j = 1, \ldots, 5, \\ y_{t-1} \cdot y_{t-2} & \text{for } j = 6, \\ y_{t-1} \cdot y_{t-3} & \text{for } j = 7, \\ \vdots & \\ y_{t-4} \cdot y_{t-5} & \text{for } j = 14, \\ y_{t-j+9}^2 & \text{for } j = 15, \ldots, 20. \end{cases} \tag{79}$$

In this delineation, g_1 to g_5 are assigned to the linear terms, g_6 to g_{14} to the interaction terms, and g_{15} to g_{20} to the squared terms. This expansive framework not only contemplates the sequential impact of the prior observations but also scrutinizes the combinative and quadratic interactions, thereby substantively refining the forecasting strength of the time series analysis.

5 Sparsity in Coefficients of Time Series Models

In the context of time series forecasting using higher-order models, it is often observed that some coefficients may be zero. This phenomenon can be attributed to several mathematical and statistical reasons, which are crucial for optimizing the model's predictive accuracy and computational efficiency. We discuss the primary factors that lead to zero coefficients in some cases.

5.1 Data Characteristics and Redundancy

In some scenarios, the inherent characteristics of the time series data can lead to zero coefficients. If certain lagged values or interactions do not significantly contribute to predicting the future values, the optimization algorithm might set their corresponding coefficients to zero. This is particularly common when:

- There is multicollinearity among predictors, causing redundancy.
- Some features are irrelevant for prediction.

5.2 Mathematical Proof of Coefficient Reduction

Theorem 9. *In the Generalized Least Deviation Method (GLDM), a coefficient a_j in the model can be reduced to zero if the gradient of the objective function $F(\mathbf{a})$ with respect to a_j is zero. This occurs when the contribution of a_j to the overall prediction error is negligible, allowing the model to retain its predictive accuracy while simplifying the model.*

Proof. Let $F(\mathbf{a})$ denote the objective function defined as:

$$F(\mathbf{a}) = \sum_{t=1}^{T} \arctan\left|y_t - f(\mathbf{x}_t, \mathbf{a})\right|, \tag{80}$$

where:

- $y_t \in \mathbb{R}$ is the actual value of the time series at time t,
- $\mathbf{x}_t \in \mathbb{R}^n$ is the vector of input variables at time t,
- $f(\mathbf{x}_t, \mathbf{a})$ is the predicted value at time t based on the model, with $\mathbf{a} = \{a_1, a_2, \ldots, a_{n(m)}\}$ being the vector of coefficients.

Step 1: Gradient of the Objective Function To determine the conditions under which a coefficient a_j can be reduced to zero, we begin by computing the partial derivative of the objective function $F(\mathbf{a})$ with respect to a_j:

$$\frac{\partial F}{\partial a_j} = \sum_{t=1}^{T} \frac{\partial}{\partial a_j} \arctan\left|y_t - f(\mathbf{x}_t, \mathbf{a})\right|. \tag{81}$$

Step 2: Application of the Chain Rule Using the chain rule, the partial derivative can be expanded as:

$$\frac{\partial F}{\partial a_j} = \sum_{t=1}^{T} \frac{1}{1 + (y_t - f(\mathbf{x}_t, \mathbf{a}))^2} \cdot \frac{\partial}{\partial a_j} \left(y_t - f(\mathbf{x}_t, \mathbf{a})\right). \tag{82}$$

Here, the first term $\frac{1}{1+(y_t - f(\mathbf{x}_t, \mathbf{a}))^2}$ is the derivative of the arctangent function, which acts as a scaling factor that weights the contribution of each residual $y_t - f(\mathbf{x}_t, \mathbf{a})$ to the gradient.

Step 3: Derivative of the Prediction Function Next, we calculate the derivative of the prediction function $f(\mathbf{x}_t, \mathbf{a})$ with respect to the coefficient a_j:

$$\frac{\partial}{\partial a_j} \left(y_t - f(\mathbf{x}_t, \mathbf{a})\right) = -\frac{\partial f(\mathbf{x}_t, \mathbf{a})}{\partial a_j}. \tag{83}$$

The negative sign indicates that the derivative of the residual with respect to a_j is directly proportional to the derivative of the prediction function.

Step 4: Substituting into the Gradient Expression Substituting equation (83) into equation (82), we obtain:

$$\frac{\partial F}{\partial a_j} = -\sum_{t=1}^{T} \frac{1}{1 + (y_t - f(\mathbf{x}_t, \mathbf{a}))^2} \cdot \frac{\partial f(\mathbf{x}_t, \mathbf{a})}{\partial a_j}. \tag{84}$$

This equation expresses the gradient of the objective function with respect to a_j as a weighted sum of the derivatives of the prediction function.

Step 5: Second-Order Derivative Analysis We now consider the second derivative of the objective function with respect to a_j, which helps in assessing the minimal influence of a_j on the objective function:

$$\frac{\partial^2 F}{\partial a_j^2} = \sum_{t=1}^{T} \frac{-2\left(\frac{\partial f(\mathbf{x}_t, \mathbf{a})}{\partial a_j}\right)^2}{\left(1 + (y_t - f(\mathbf{x}_t, \mathbf{a}))^2\right)^2}. \tag{85}$$

Step 6: Condition for Coefficient Reduction For a_j to be reduced to zero, the gradient $\frac{\partial F}{\partial a_j}$ must be equal to zero and the second derivative $\frac{\partial^2 F}{\partial a_j^2}$ should be approximately zero:

$$\frac{\partial F}{\partial a_j} = 0 \quad \text{and} \quad \frac{\partial^2 F}{\partial a_j^2} \approx 0. \tag{86}$$

This condition indicates that the contribution of a_j to the overall objective function is minimal, meaning that setting $a_j = 0$ will not significantly affect the model?s performance.

Step 7: Interpretation and Implications When the condition in equation (86) is satisfied, setting $a_j = 0$ will not degrade the performance of the model. This simplification reduces the number of active parameters in the model, leading to a more parsimonious model. The benefits of this reduction include:

- **Increased Model Interpretability**: Fewer active coefficients make the model easier to interpret.
- **Improved Computational Efficiency**: Reducing the number of parameters decreases the computational burden, especially in high-dimensional settings.
- **Avoidance of Overfitting**: Eliminating coefficients that do not contribute to the model reduces the risk of overfitting, leading to better generalization on unseen data.

Conclusion The mathematical conditions $\frac{\partial F}{\partial a_j} = 0$ and $\frac{\partial^2 F}{\partial a_j^2} \approx 0$ provide a rigorous basis for the reduction of coefficients in the GLDM framework. When these conditions are met, the coefficient a_j can be safely set to zero, enhancing the efficiency and interpretability of the model without compromising its predictive performance. This result underscores the importance of analyzing both the first and second derivatives of the objective function to guide the model simplification process. □

5.3 Mathematical Proof of Optimal Model Order Selection in GLDM

Theorem 10. *Let $\{y_t\}_{t=1}^T$ be a time series, and let the function $F(\mathbf{a})$ be defined as:*

$$F(\mathbf{a}) = \sum_{t=1}^{T} \arctan\left|y_t - f(\mathbf{x}_t, \mathbf{a})\right|,$$

where the function f depends on the parameter $\mathbf{a}$. Then the optimal model order m^* is given by:

$$m^* = \arg\min_m \left(\min_{\mathbf{a}} F(\mathbf{a}, m)\right).$$

Proof. Consider the differentiable function $F(\mathbf{a})$ and its first derivatives. The gradient of the function $F(\mathbf{a})$ is given by:

$$\frac{\partial F}{\partial a_j} = \sum_{t=1}^{T} \frac{-1}{1 + (y_t - f(\mathbf{x}_t, \mathbf{a}))^2} \cdot \frac{\partial f(\mathbf{x}_t, \mathbf{a})}{\partial a_j},$$

which equals zero at critical points, where $\nabla F(\mathbf{a}) = 0$.

The Hessian of the function $F(\mathbf{a})$, the matrix of second derivatives, is given by:

$$H(\mathbf{a}) = \left[\frac{\partial^2 F}{\partial a_i \partial a_j}\right] = \sum_{t=1}^{T} \left[\frac{-2\frac{\partial f(\mathbf{x}_t, \mathbf{a})}{\partial a_i}\frac{\partial f(\mathbf{x}_t, \mathbf{a})}{\partial a_j}}{\left(1 + (y_t - f(\mathbf{x}_t, \mathbf{a}))^2\right)^2} + \frac{\frac{\partial^2 f(\mathbf{x}_t, \mathbf{a})}{\partial a_i \partial a_j}}{1 + (y_t - f(\mathbf{x}_t, \mathbf{a}))^2}\right].$$

Now, consider the differentiable function $F(\mathbf{a})$ defined on the parameter space $\mathbf{a}$. Let $\mathbf{a}_0$ be a critical point of this function, where the gradient $\nabla F(\mathbf{a}_0) = 0$.

The Hessian matrix $H(\mathbf{a}_0)$, representing the second derivatives of $F(\mathbf{a})$ at $\mathbf{a}_0$, is computed as:

$$H(\mathbf{a}_0) = \sum_{t=1}^{T} \left[\frac{-2 \frac{\partial f(\mathbf{x}_t, \mathbf{a}_0)}{\partial a_i} \frac{\partial f(\mathbf{x}_t, \mathbf{a}_0)}{\partial a_j}}{\left(1 + (y_t - f(\mathbf{x}_t, \mathbf{a}_0))^2\right)^2} + \frac{\frac{\partial^2 f(\mathbf{x}_t, \mathbf{a}_0)}{\partial a_i \partial a_j}}{1 + (y_t - f(\mathbf{x}_t, \mathbf{a}_0))^2} \right].$$

The Hessian matrix $H(\mathbf{a}_0)$, evaluated as the matrix of second derivatives at the point $\mathbf{a}_0$, is given by:

$$H(\mathbf{a}_0) = \begin{bmatrix} H_{11} & \cdots & H_{1n} \\ \vdots & \ddots & \vdots \\ H_{n1} & \cdots & H_{nn} \end{bmatrix},$$

where each element H_{ij} is determined by:

$$H_{ij} = \sum_{t=1}^{T} \left[\frac{-2 \frac{\partial f(\mathbf{x}_t, \mathbf{a}_0)}{\partial a_i} \frac{\partial f(\mathbf{x}_t, \mathbf{a}_0)}{\partial a_j}}{\left(1 + (y_t - f(\mathbf{x}_t, \mathbf{a}_0))^2\right)^2} + \frac{\frac{\partial^2 f(\mathbf{x}_t, \mathbf{a}_0)}{\partial a_i \partial a_j}}{1 + (y_t - f(\mathbf{x}_t, \mathbf{a}_0))^2} \right].$$

To analyze the positive definiteness of the matrix $H(\mathbf{a}_0)$, consider the quadratic form:

$$\mathbf{v}^\top H(\mathbf{a}_0)\mathbf{v} = \sum_{i=1}^{n} \sum_{j=1}^{n} v_i v_j \left[\sum_{t=1}^{T} \left(\frac{-2 \frac{\partial f(\mathbf{x}_t, \mathbf{a}_0)}{\partial a_i} \frac{\partial f(\mathbf{x}_t, \mathbf{a}_0)}{\partial a_j}}{\left(1 + (y_t - f(\mathbf{x}_t, \mathbf{a}_0))^2\right)^2} + \frac{\frac{\partial^2 f(\mathbf{x}_t, \mathbf{a}_0)}{\partial a_i \partial a_j}}{1 + (y_t - f(\mathbf{x}_t, \mathbf{a}_0))^2} \right) \right],$$

where $\mathbf{v} = (v_1, \ldots, v_n)$ is an arbitrary non-zero vector.

The positive definiteness of $H(\mathbf{a}_0)$ implies that for all non-zero $\mathbf{v}$, the expression $\mathbf{v}^\top H(\mathbf{a}_0)\mathbf{v}$ is strictly greater than zero. This is achieved if the sum of the positive terms, associated with the second derivatives $\frac{\partial^2 f(\mathbf{x}_t, \mathbf{a})}{\partial a_i \partial a_j}$, exceeds the sum of the negative terms, arising from the products of the first derivatives $\frac{\partial f(\mathbf{x}_t, \mathbf{a})}{\partial a_i} \frac{\partial f(\mathbf{x}_t, \mathbf{a})}{\partial a_j}$.

Thus, we can conclude that if $H(\mathbf{a}_0) > 0$ for all non-zero vectors $\mathbf{v}$, then the matrix $H(\mathbf{a}_0)$ is positive definite, and therefore, $\mathbf{a}_0$ is a point of local minimum for the function $F(\mathbf{a})$.

In this theorem, the conditions for determining the optimal model order m^* for the time series $\{y_t\}_{t=1}^T$ using the loss function $F(\mathbf{a})$ are formulated. The function $F(\mathbf{a})$ is computed as the sum of the arctangents of the absolute deviations between the observed values y_t and the predicted values $f(\mathbf{x}_t, \mathbf{a})$. The optimal order m^* minimizes this function, selecting the best compromise between model complexity and forecast accuracy. □

5.4 Algorithm Implementation Using GLDM

Algorithm 5: Determine Coefficients Using GLDM for Variable Order Models

Require: $Y = \{y_t \in \mathbb{R}^+\}_{t=1}^{T}$; `// Values of the given state variables`
m ; `// Model order`
Predefined functions $\{g_j\}$
Ensure : Coefficient vector a that minimizes the GLDM objective

Function `DefineFunctions(`m`):`
  for $j \leftarrow 1$ **to** m **do**
    $g_j(\{y_{t-k}\}_{k=1}^{m}) \leftarrow y_{t-j}$; `// Linear terms`
  for $j \leftarrow m+1$ **to** $m + \binom{m}{2}$ **do**
    Define g_j for interaction terms
  for $j \leftarrow m + \binom{m}{2} + 1$ **to** $\frac{m(m+3)}{2}$ **do**
    $g_j(\{y_{t-k}\}_{k=1}^{m}) \leftarrow \left(y_{t-j+m+\binom{m}{2}}\right)^2$; `// Squared terms`
  return $\{g_j\}_{j=1}^{n(m)}$
Function `GLDMFit(`Y, m`):`
  $n(m) \leftarrow \frac{m(m+3)}{2}$; `// Calculate the number of coefficients`
  ;
  $\{g_j\} \leftarrow$ `DefineFunctions(`m`);`
  Initialize vector a of length $n(m)$ to zeros;
  Define $F(a)$ as in Equation (3);
  for $j \leftarrow 1$ **to** $n(m)$ **do**
    Initialize a_j;
  while *not converged* **do**
    for $t \leftarrow 1$ **to** T **do**
      $r_t \leftarrow y_t - \sum_{j=1}^{n(m)} a_j \cdot g_j(\{y_{t-k}\}_{k=1}^{m})$; `// Calculate residual`
      ;
      Update a_j by minimizing $\arctan|r_t|$ for each j;
    Check for convergence;
  return a ; `// Return the optimized coefficient vector`
$a \leftarrow$ `GLDMFit(`Y, m`)` ; `// Fit the model using GLDM`
;
Output the coefficients vector a;

The algorithm 5 implemented using the Generalized Least Deviation Method (GLDM) is structured to efficiently determine coefficients for a variable order time series model. Below is a detailed breakdown of the algorithmic procedure:

1. **Initialization and Function Definition**: The algorithm starts by defining necessary functions based on the given model order m. It systematically generates linear terms, interaction terms, and squared terms for each level of m, setting the stage for the model's response to historical data patterns.

2. **Coefficient Initialization**: It initializes a vector of coefficients a, which are crucial as they will be optimized to fit the time series data effectively.

3. **Objective Function Formulation**: The heart of the algorithm is the objective function $F(a)$, designed to be minimized. This function is computed as the sum of arctangents of

the residuals, where residuals are the differences between the actual data points and the models predictions based on current coefficients.

4. **Optimization Loop**: Through iterative processes, the algorithm adjusts each coefficient to minimize $F(a)$. For each time point, it computes the residual and updates the coefficients to reduce these residuals, leveraging the arctangent function for robustness against outliers.

5. **Convergence Evaluation**: The iterative adjustment continues until the change in coefficients between successive iterations is minimal, indicating convergence and optimal fitting.

6. **Output Generation**: Upon convergence, the algorithm outputs the optimized coefficient vector a, which is then utilized for model predictions or further analysis.

This method is particularly adept at handling time series data with irregularities or non-standard distributions, ensuring enhanced predictive accuracy and model robustness.

5.5 Algorithm for Identifying Zero Coefficients in GLDM

The algorithm 6 is structured to identify and potentially zero out some coefficients in a time series forecasting model using the Generalized Least Deviation Method (GLDM). Here is a simplified breakdown of the algorithmic steps:

1. **Initialization:** The algorithm calculates the necessary number of coefficients, $n(m) = \frac{m(m+3)}{2}$, where m is the model order. It initializes these coefficients to zero and sets up the objective function, $F(\mathbf{a})$, which aims to minimize the cumulative arctangent of the absolute differences between actual time series data $\{y_t\}$ and the model's predictions.

2. **Gradient Calculation:** For each coefficient a_j, the gradient of the objective function with respect to a_j is computed. This gradient indicates how small changes in a_j affect $F(\mathbf{a})$, guiding the coefficient updates.

3. **Coefficient Update:** If the absolute value of a coefficient's gradient is below a small threshold ϵ, that coefficient is set to zero. This is based on the rationale that if the gradient is minimal, the coefficient's impact on the model is negligible, and zeroing it simplifies the model without significantly impacting its effectiveness.

4. **Convergence Check:** The algorithm iterates, updating coefficients until changes in the objective function $F(\mathbf{a})$ are negligible, indicating convergence.

5. **Output:** Upon convergence, the algorithm outputs the optimized coefficient vector $\mathbf{a}$, where some coefficients may be zero, reflecting a potentially simpler and more interpretable model.

This methodology is particularly useful in scenarios where balancing model complexity and interpretability with accuracy is crucial, allowing for a streamlined model that efficiently captures

the essential dynamics of the data.

Algorithm 6: Identification of Zero Coefficients in GLDM Models

Require: $Y = \{y_t \in \mathbb{R}^+\}_{t=1}^{T}$; // Values of the given state variables

m ; // Model order

Ensure : Coefficient vector $\mathbf{a}$ with possibly zero elements where appropriate

Function GLDMFit(Y, m)**:**

  $n(m) \leftarrow \frac{m(m+3)}{2}$; // Calculate the number of coefficients

  ;

  Initialize vector $\mathbf{a}$ of length $n(m)$ to zeros;

  Define the objective function $F(\mathbf{a})$ as:

$$F(\mathbf{a}) = \sum_{t=1}^{T} \arctan \left| y_t - \sum_{j=1}^{n(m)} a_j \cdot g_j \left(\{y_{t-k}\}_{k=1}^{m} \right) \right|$$

  while *not converged* **do**

    ; // Iteratively minimize $F(\mathbf{a})$

    for $j \leftarrow 1$ **to** $n(m)$ **do**

      Calculate the gradient ∇F_{a_j};

$$\nabla F_{a_j} = \frac{\partial}{\partial a_j} \sum_{t=1}^{T} \arctan \left| y_t - \sum_{k=1}^{n(m)} a_k \cdot g_k \left(\{y_{t-i}\}_{i=1}^{m} \right) \right|$$

      if $|\nabla F_{a_j}| < \epsilon$ **then**

        $a_j \leftarrow 0$; // Set coefficient to zero if gradient is negligible

        ;

    Check for convergence ; // Based on the change in $F(\mathbf{a})$

    ;

  return a ; // Return the optimized coefficient vector

a $\leftarrow$ GLDMFit(Y, m) ; // Fit the model using GLDM

;

Output the coefficients vector $\mathbf{a}$;

6 Soft computing Results

The datasets employed in our analysis are summarized in Table 1, which details their respective lengths. The datasets include NDVI with 15 data points, Temperature with 9,939 data points, Wind Speed recorded with 50,530 data points, and COVID-19 death cases in the Russian Federation, which consist of 882 data points. This variation in dataset sizes reflects the diverse scope and scale of environmental and epidemiological data considered in our time series forecasting models. The extensive data length, particularly for Temperature and Wind Speed, provides a robust basis for statistical analysis and model validation.

Table 1: List of used datasets and their lengths

Dataset	Length
NDVI	15
Temperature	9,939
Wind Speed	50,530
COVID-19 deaths Cases in the Russian Federation	882

Table 2 summarizes the performance metrics of various models applied to NDVI data. The Multilayer Perceptron (MLP) model demonstrates moderate performance, with an RMSE of 0.34, MSE of 0.12, MAE of 0.25, an R-squared value of 0.47, and a relatively high MAPE of 51.71%. The Support Vector Machine (SVM) model shows notable improvement, achieving an RMSE of 0.08, MSE of 0.01, MAE of 0.08, an R-squared value of 0.59, and a MAPE of 14.74%. The AutoARIMA model performs well, with an RMSE of 0.07, MSE of 0.00, MAE of 0.03, an R-squared value of 0.74, and a MAPE of 8.50%. The Exponential Smoothing model excels with an RMSE of 0.03, MSE of 0.00, MAE of 0.02, an R-squared value of 0.95, and a MAPE of 4.17%. The BATS model provides moderate results, with an RMSE of 0.10, MSE of 0.01, MAE of 0.07, an R-squared value of 0.36, and a MAPE of 16.53%. The TBATS model performs very well, with an RMSE of 0.04, MSE of 0.01, MAE of 0.03, an R-squared value of 0.90, and a MAPE of 6.99%. The Prophet model exhibits moderate performance, with an RMSE of 0.12, MSE of 0.01, MAE of 0.10, an R-squared value of 0.17, and a MAPE of 20.37%. The Hybrid AutoARIMA-Exponential Smoothing model shows strong performance, with an RMSE of 0.06, MSE of 0.00, MAE of 0.04, an R-squared value of 0.76, and a MAPE of 9.45%. Similarly, the Hybrid AutoARIMA-Polynomial model performs well, with an RMSE of 0.06, MSE of 0.00, MAE of 0.03, an R-squared value of 0.80, and a MAPE of 8.42%. Notably, the GLDM Second Order model outperforms all other models, achieving the lowest RMSE of 0.02, MSE of 0.00, MAE of 0.01, the highest R-squared value of 0.96, and the lowest MAPE of 2.16%, indicating superior accuracy and reliability.

Table 2: Performance Metrics for Various Models for NDVI

Model	RMSE	MSE	MAE	R-Squared	MAPE
MLP Model	0.34	0.12	0.25	0.47	51.71%
SVM Model	0.08	0.01	0.08	0.59	14.74%
Autoarima Model	0.07	0.00	0.03	0.74	8.50%
Exponential Smoothing Model	0.03	0.00	0.02	0.95	4.17%
BATS Model	0.10	0.01	0.07	0.36	16.53%
TBATS Model	0.04	0.01	0.03	0.90	6.99%
Prophet Model	0.12	0.01	0.10	0.17	20.37%
Hybrid auto arima-ES	0.06	0.00	0.04	0.76	9.45%
Hybrid auto arima-Polynomial	0.06	0.00	0.03	0.80	8.42%
GLDM Second Order	**0.02**	**0.00**	**0.01**	**0.96**	**2.16%**

The performance metrics for various models applied to temperature data are summarized in Table 3. The MLP Model shows moderate performance with an RMSE of 4.60, MSE of 21.17, MAE of 3.62, R-squared of 0.37, and a high MAPE of 29.17%. The SVM Model outperforms all other models with an RMSE of 0.40, MSE of 0.16, MAE of 0.24, R-squared of 0.99, and a low MAPE of 1.80%. The Autoarima Model performs well, achieving an RMSE of 0.92, MSE of 0.85, MAE of 0.62, R-squared of 0.97, and a MAPE of 4.54%. The Exponential Smoothing Model shows good performance with an RMSE of 1.05, MSE of 1.10, MAE of 0.70, R-squared of 0.97, and a MAPE of 4.98%. The BATS Model performs well with an RMSE of 0.96, MSE of 0.91, MAE of 0.59, R-squared of 0.97, and a MAPE of 4.25%. The TBATS Model also shows strong performance with an RMSE of 0.71, MSE of 0.50, MAE of 0.45, R-squared of 0.99, and a MAPE of 3.23%. The Prophet Model shows moderate performance with an RMSE of 4.03, MSE of 16.22, MAE of 3.12, R-Squared of 0.52, and a MAPE of 23.49%. The Hybrid auto arima-ES and Hybrid auto arima-Polynomial Models both show good performance with an RMSE of 0.93 and 0.92, MSE of 0.86 and 0.85, MAE of 0.62, R-squared of 0.97, and a MAPE of 4.54%. The GLDM Fifth Order Model shows strong performance with an RMSE of 0.94, MSE of 0.88, MAE of 0.58, R-squared of 0.97, and a MAPE of 4.12%.

Table 4 presents the performance metrics for various models used in predicting wind speed. The table includes key indicators such as Root Mean Square Error (RMSE), Mean Squared Error (MSE), Mean Absolute Error (MAE), R-squared, and Mean Absolute Percentage Error (MAPE)

Table 3: Performance Metrics for Various Models for Temperature

Model	RMSE	MSE	MAE	R-Squared	MAPE
MLP Model	4.60	21.17	3.62	0.37	29.17%
SVM Model	**0.40**	**0.16**	**0.24**	**0.99**	**1.80%**
Autoarima Model	0.92	0.85	0.62	0.97	4.54%
Exponential Smoothing Model	1.05	1.10	0.70	0.97	4.98%
BATS Model	0.96	0.91	0.59	0.97	4.25%
TBATS Model	**0.71**	**0.50**	**0.45**	**0.99**	**3.23%**
Prophet Model	4.03	16.22	3.12	0.52	23.49%
Hybrid auto arima-ES	0.93	0.86	0.62	0.97	4.54%
Hybrid auto arima-Polynomial	0.92	0.85	0.62	0.97	4.54%
GLDM Fifth Order	**0.94**	**0.88**	**0.58**	**0.97**	**4.12%**

for each model. The GLDM Second Order model demonstrates the best overall performance with the lowest RMSE of 0.74, MSE of 0.55, and MAE of 0.52. It also achieves the highest R-squared value of 0.97, indicating it explains 97% of the variance in the wind speed data. Additionally, it has the lowest MAPE of 9.50%, reflecting high accuracy in predictions. The Autoarima and Exponential Smoothing models also show strong performance with RMSE values of 0.7453 and 0.7493, respectively, and R-squared values close to the GLDM model at 0.9689 and 0.9686. Their MAPE values are 10.00% and 9.99%, respectively, indicating good predictive accuracy. The SVM model performs well with an RMSE of 0.80 and an R-squared value of 0.96. However, its MAPE of 12.13% is slightly higher compared to the GLDM and Autoarima models. In contrast, models like the MLP and TBATS show less accurate predictions with higher RMSE and MAPE values. The MLP model has an RMSE of 5.070 and a very high MAPE of 102.61%, while the TBATS model has an RMSE of 4.6 and a MAPE of 24.36%. The Prophet model, Hybrid auto arima-ES, and Hybrid auto arima-Polynomial models also show relatively higher RMSE and MAPE values, indicating lower predictive performance compared to the top-performing models. In summary, the GLDM Second Order model stands out as the most effective for wind speed prediction, providing the highest accuracy and reliability among the models evaluated.

Table 4: Performance Metrics for Various Models for Wind Speed

Model	RMSE	MSE	MAE	R-Squared	MAPE
MLP Model	5.070	25.70	4.05	0.04969	102.61%
SVM Model	0.80	0.64	0.638	0.96	12.13%
Autoarima Model	0.7453	0.5555	0.5209	0.9689	10.00%
Exponential Smoothing Model	0.7493	0.5614	0.5226	0.9686	9.99%
BATS Model	2.4	5.76	1.915	0.30	19.15%
TBATS Model	4.6	21.16	3.671	0.45	24.36%
Prophet Model	3.8472	14.8009	3.0759	0.1717	78.03%
Hybrid auto arima-ES	0.90	0.81	0.718	0.19	71.80%
Hybrid auto arima-Polynomial	0.88	0.7744	0.702	0.2254	70.20%
GLDM Second Order	**0.74**	**0.55**	**0.52**	**0.97**	**9.50%**

Overall, the performance analysis demonstrates that the high-order quasilinear equations developed using the Generalized Least Deviation Method (GLDM) consistently outperform the latest models across various datasets, including NDVI, temperature, and wind speed. The GLDM models exhibit superior accuracy and reliability, evidenced by lower RMSE, MSE, and MAPE values, and higher R-Squared values. In the NDVI dataset, the GLDM models demonstrate exceptional performance, significantly reducing error metrics while maintaining high R-Squared values. This highlights the GLDM's effectiveness in capturing the intricate dynamics of NDVI data, providing reliable and precise predictions. In the temperature dataset, the GLDM Fifth Order model shows strong performance, but the SVM model outperforms all other models, achieving the best metrics across all evaluated criteria. The SVM model's exceptional perfor-

mance is attributed to its ability to effectively handle high-dimensional data and capture complex non-linear relationships, resulting in low error metrics and a high R-Squared value. Additionally, the TBATS model demonstrates superior performance in temperature prediction with a strong ability to capture seasonal patterns and complex trends, leading to low error metrics and high R-Squared values. In the wind speed dataset, the GLDM Second Order model excels, achieving the best metrics across all evaluated criteria. This model's effectiveness is reflected in its remarkably low error metrics and high R-Squared value, indicating a robust capability to capture the underlying patterns and dynamics of wind speed data. The robust performance of GLDM models across diverse datasets highlights their ability to manage outliers and non-standard error distributions, making them a reliable and effective approach for complex, non-linear data. These results underscore the robustness and effectiveness of the GLDM approach in handling diverse datasets and highlight the superior capabilities of the SVM and TBATS models for temperature prediction.

Table 5 details the performance of different models for COVID-19 deaths in the Russian Federation. The GLDM second-order model demonstrates superior performance, evidenced by its high R^2 value of 0.9898 and a notably low MAPE of 10.96%. This performance is remarkable when compared to other models, such as the MLP model, which has an R^2 of 0.0446 and a MAPE of 167.1630%, and the SVM model, which shows an R^2 of 0.9742 and a MAPE of 17.3852%. The Auto ARIMA model and its hybrid variants also perform well, with R^2 values around 0.9917 and MAPEs just above 11%, but they do not surpass the accuracy of the GLDM second-order model. Similarly, the BATS and TBATS models, both with R^2 values of 0.9921 and MAPEs of 11.0584%, and the Prophet model, which has an R^2 of 0.9746 and a significantly higher MAPE of 53.7971%, also fall short in comparison. This illustrates the robustness of the GLDM second-order model in providing precise and reliable predictions for COVID-19 deaths in the region.

The mathematical advantages of the GLDM second-order model extend beyond its ability to incorporate higher-order terms. Its optimization framework, which strategically minimizes a well-defined loss function through the use of the arctangent, enhances the model's ability to adapt to new data and maintain high accuracy. The inclusion of second-order terms is crucial for capturing the complexities and nuances of time series data, making the GLDM particularly effective for modeling the nonlinear and stochastic nature of COVID-19 dynamics. This advanced methodological approach allows the GLDM second-order model to outperform other models, making it a critical tool for public health planning and intervention.

In summary, the GLDM second-order model not only demonstrates superior accuracy and reliability in forecasting COVID-19 infection cases in the Samara Region but also excels in predicting COVID-19 deaths in the Russian Federation, outperforming a range of other predictive models. Its advanced methodological approach, which incorporates higher-order terms and robust optimization techniques, allows it to capture complex patterns and dependencies with greater precision. These optimized performance metrics make the GLDM second-order model an indispensable tool in the ongoing efforts to model and understand the dynamics of the COVID-19 pandemic, providing critical insights for public health strategies and interventions.

Table 6 summarizes the number of coefficients required for time series forecasting models of varying orders, from first to fifth. Specifically, the table enumerates the coefficients as 2, 5, 9, 14, and 20 for the first through fifth orders, respectively. This progression is governed by the formula $n(m) = 2m + \binom{m}{2} = \frac{m(m+3)}{2}$, which calculates the total count of coefficients including linear, interaction, and quadratic terms as the model order increases. The structured increase in coefficients highlights the model's growing complexity and capacity to encapsulate more intricate dynamics within the time series data.

Table 5: Error Metrics (R-Squared and MAPE) for Various Models for COVID-19 Deaths in Russian Federation

Model	**R-Squared**	**MAPE (%)**
MLP model	0.0446	167.1630
SVM model	0.9742	17.3852
Auto ARIMA model	0.9917	11.0454
Exponential Smoothing model	0.9898	11.1647
BATS model	0.9921	11.0584
TBATS model	0.9921	11.0584
Prophet model	0.9746	53.7971
Hybrid autoARIMA+ES	0.9915	11.2055
Hybrid autoARIMA+Polynomial	0.9917	11.0749
GLDM Second Order	**0.9898**	**10.9600**

Table 6: Number of Coefficients by Order

Order	First	Second	Third	Fourth	Fifth
Coefficients	2	5	9	14	20
$n(m) = 2m + \binom{m}{2} = \frac{m(m+3)}{2}$					

The tables from 7 to 11 present the coefficients for the Generalized Least Deviation Method (GLDM) applied to the Normalized Difference Vegetation Index (NDVI) data across different model orders, from first through fifth. Each table, corresponding to the model order, lists the coefficients derived from fitting the GLDM model to the NDVI dataset. Table 7 starts with the simplest model, featuring only two coefficients, a_1 and a_2. As the model complexity increases, more coefficients are introduced to capture additional dynamics of the data, evident in Table 8 for the second order and further expanded in Tables 9, 10, and 11 for higher orders. These coefficients are crucial for understanding the NDVI time series' behavior and improving prediction accuracy. Notably, as the order increases, the number of coefficients grows, reflecting the model's enhanced capability to incorporate more historical data points and interactions within the NDVI time series.

Table 7: GLDM First Order Coefficients for NDVI

Coefficient	Value
a_1	1.7073
a_2	-1.0511

Table 8: GLDM Second Order Coefficients for NDVI

Coefficient	Value
a_1	3.4694
a_2	-2.1864
a_3	-5.5924
a_4	-2.5635
a_5	7.7299

The coefficients derived from applying the Generalized Least Deviation Method (GLDM) for the temperature data set are systematically presented in Tables 12 through 16. These tables enumerate the coefficients for models of increasing order from first to fifth. Table 12 lists the coefficients for the first order model, indicating the foundational linear influences in the temperature data. Progressing to higher model orders, Table 13 and Table 14 introduce additional coefficients, capturing more complex dynamics and interactions within the data. This trend continues with Table 15, where the fourth order model incorporates even more coefficients, enhancing the model's ability to forecast with greater precision. Finally, Table 16 presents

Table 9: GLDM Third Order Coefficients for NDVI

Coefficient	Value
a_1	-9.6495
a_2	-16.2326
a_3	29.1697
a_4	76.3993
a_5	122.9467
a_6	-71.5312
a_7	-229.9915
a_8	98.9790
a_9	0.0000

Table 10: GLDM Fourth Order Coefficients for NDVI

a_1	a_2	a_3	a_4	a_5	a_6	a_7
-21.8416	52.4809	30.7212	-48.3575	-132.0576	-177.0664	4.7422
a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}
-2.5712	273.2420	-66.8228	83.1160	0.0000	0.0000	0.0000

the coefficients for the fifth order model, which encompasses the most comprehensive dynamic range, utilizing twenty coefficients to capture nuanced patterns and potential non-linearities in the temperature series. Each table reflects the incremental complexity and enhanced predictive capability as the order of the model increases.

Tables 17 and 18 detail the coefficients for first and second order models applied to wind speed data using a specific modeling technique. Table 17 displays the coefficients for the first order model, capturing the most immediate past influence with coefficients a_1 and a_2. Moving to a more complex model, Table 18 lists the coefficients for the second order model, which considers additional past values to better capture the dynamics and potential patterns in wind speed variations. This model includes more coefficients (a_1 to a_5), thereby providing a richer, more nuanced understanding of the influence of past wind speeds on future predictions. The expansion in the number of coefficients from the first to the second order model reflects an increase in model complexity and potential predictive power.

Tables 19, 20, and 21 illustrate the coefficients determined by the Generalized Least Deviation Method (GLDM) for analyzing death cases in Russia across three different model orders. Table 19 lists the coefficients for the first order model, suggesting a simplistic model where the primary coefficient a_1 is 1.0000, indicating a direct influence of the immediate past value on the future value with minimal adjustment (a_2 is 0.0000). As the model complexity increases, Table 20 provides five coefficients for a second order model, incorporating more nuanced interactions and trends in the data. The third order model, shown in Table 21, further expands this complexity by including nine coefficients, thus offering a more detailed and intricate depiction of the dynamics influencing the death rates. These tables collectively represent a progression in model sophistication and predictive potential, adapting to the increasing complexity required to accurately model the temporal dynamics of death cases.

Table 11: GLDM Fifth Order Coefficients for NDVI

a_1	a_2	a_3	a_4	a_5
0.0000	-29.0004	64.0513	-44.3069	10.2588
a_6	a_7	a_8	a_9	a_{10}
1.7283	-23.5562	-90.7331	30.3171	-6.6379
a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
-2.0578	90.0347	0.0000	0.0000	0.0000
a_{16}	a_{17}	a_{18}	a_{19}	a_{20}
0.0000	0.0000	0.0000	0.0000	0.0000

Table 12: GLDM First Order Coefficients for Temperature

Coefficient	Value
a_1	1.0159
a_2	-0.0009

Table 13: GLDM Second Order Coefficients for Temperature

Coefficient	Value
a_1	1.0498
a_2	-0.0302
a_3	0.0229
a_4	0.0098
a_5	-0.0340

Table 14: GLDM Third Order Coefficients for Temperature

Coefficient	Value
a_1	-0.1658
a_2	0.0395
a_3	1.1547
a_4	0.0362
a_5	0.0298
a_6	0.0175
a_7	-0.0489
a_8	-0.0365
a_9	0.0000

Table 15: GLDM Fourth Order Coefficients for Temperature

a_1	a_2	a_3	a_4	a_5	a_6	a_7
1.1661	-0.3931	1.6191	-1.2894	-0.0031	0.1141	-0.1237
a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}
-0.0320	-0.0594	0.1424	-0.0881	-0.2193	0.1068	0.1502

Table 16: GLDM Fifth Order Coefficients for Temperature

a_1	a_2	a_3	a_4	a_5
0.0000	1.0667	-0.4329	1.4878	-0.7536
a_6	a_7	a_8	a_9	a_{10}
-0.2734	0.0154	0.0804	0.0679	0.0827
a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
0.0083	-0.0745	-0.0029	0.0609	-0.0092
a_{16}	a_{17}	a_{18}	a_{19}	a_{20}
-0.1386	0.0510	0.0560	-0.1933	0.0442

Table 17: First Order Coefficients for Wind Speed

Coefficient	Value
a_1	1.0092
a_2	-0.0011

Table 18: Second Order Coefficients for Wind Speed

Coefficient	Value
a_1	0.9300
a_2	0.0764
a_3	0.0248
a_4	0.0241
a_5	-0.0499

Table 19: First Order GLDM Model Coefficients for Death Cases in Russia

Coefficient	Value
a_1	1.0000
a_2	0.0000

Table 20: Second Order GLDM Model Coefficients for Death Cases in Russia

Coefficient	Value
a_1	0.7265
a_2	0.2610
a_3	0.0020
a_4	0.0016
a_5	-0.0036

6.1 Analysis of Model Orders and Coefficients

The selection of an appropriate model order and the analysis of coefficients are crucial for enhancing the predictive accuracy and interpretability of time series models. This subsection discusses the best model order for various datasets and examines the implications of zero coefficients and data length on model performance.

6.1.1 Best Model Order Selection

The optimal model order for each dataset was determined based on the balance between model complexity and the ability to capture relevant dynamics without overfitting. The selected orders are as follows:

- NDVI: Second order (Table 8)
- Temperature: Fifth order (Table 16)
- Wind Speed: Second order (Table 18)
- COVID-19 Deaths in Russia: Second order (Table 20)

These selections are supported by the data's inherent characteristics and the corresponding models' performance metrics, which include goodness of fit and predictive accuracy.

6.1.2 Coefficient Analysis and Zero Coefficients

Coefficients within these models provide insights into the influence of past values on future predictions. Particularly noteworthy is the presence of zero coefficients, which often indicate

Table 21: Third Order GLDM Model Coefficients for Death Cases in Russia

Coefficient	Value
a_1	0.5970
a_2	-0.3694
a_3	0.7396
a_4	0.0083
a_5	0.0101
a_6	-0.0009
a_7	-0.0185
a_8	0.0010
a_9	0.0000

redundant or non-influential predictors within the model. Mathematically, a coefficient a_i is set to zero when its gradient in the objective function is negligible, i.e.,

$$\frac{\partial F}{\partial a_i} \approx 0$$

where F denotes the objective function. Zero coefficients reduce model complexity and enhance computational efficiency, potentially increasing the model's generalizability. For instance, in the fifth order temperature model (Table 16), several coefficients are zero, suggesting that some higher-order interactions do not significantly impact temperature forecasting.

6.1.3 Data Length Consideration

The length of the dataset is a critical factor in model training. Sufficient data length ensures that the model captures significant trends and seasonality without overfitting. Here is the dataset length for each category:

- NDVI: 15 data points
- Temperature: 9,939 data points
- Wind Speed: 50,530 data points
- COVID-19 Deaths in Russia: 882 data points

These lengths influence the feasibility of higher-order models. For example, the extensive data length for Temperature and Wind Speed supports complex models like the fifth order, which can effectively leverage large datasets to model nuanced patterns. In contrast, the relatively shorter length for NDVI restricts the feasible model complexity, making the second order an optimal choice.

This comprehensive analysis underscores the tailored approach required in time series modeling, highlighting the interplay between model order, coefficient significance, and dataset characteristics to optimize forecasting models for various applications.

Optimal Model Orders for Temperature and Wind Speed Despite the larger data length available for Wind Speed (50,530 data points) as compared to Temperature (9,939 data points), the optimal model order chosen for Temperature is the fifth order, while for Wind Speed, it is the second order. This decision is influenced by several key factors, including the nature of the dataset, the underlying dynamics it exhibits, and the degree of non-linearity present.

For Temperature, the fifth order model (Table 16) is justified due to the complex interdependencies and significant seasonal variations inherent in temperature data. The higher-order model effectively captures these complexities through extended interactions and nonlinear relationships, which are mathematically expressed as:

$$T_t = \sum_{i=1}^{20} a_i g_i \left((T_{t-k})_{k=1}^{5} \right) + \epsilon_t,$$

where T_t represents the temperature at time t, g_i encapsulate the functions representing the higher-order interactions, a_i are the model coefficients, and ϵ_t is the error term.

In contrast, the Wind Speed dataset, despite its larger size, is sufficiently modeled by a second order construct. This can be attributed to the fact that wind speed variations, although affected by lagged values due to momentum in air movements, generally do not exhibit the same level of intricate seasonal or nonlinear patterns observed in temperature data. The second order

model captures the primary and immediate past influences concisely, eliminating the need for additional complexity, as shown by the following equation:

$$W_t = \sum_{i=1}^{5} a_i g_i \left((W_{t-k})_{k=1}^{2} \right) + \epsilon_t,$$

where W_t denotes the wind speed at time t and ϵ_t denotes the error term.

Hence, the selection of model order highlights the importance of aligning the model's complexity with the intrinsic characteristics of the dataset and not solely on its size. The complexity of the dynamics that the model aims to capture, rather than the quantity of data, is pivotal in guiding the choice of an appropriate model order for effective forecasting.

Conclusion of Results In conclusion, the analysis presented in the previous sections illustrates the critical importance of selecting an appropriate model order based on the specific characteristics and dynamics of each dataset. The optimal model orders for NDVI, Temperature, Wind Speed, and COVID-19 death cases in Russia were carefully chosen to balance complexity with the ability to capture relevant patterns without overfitting. The findings underscore that while larger datasets can support higher model orders, the inherent behavior and patterns of the data primarily dictate the suitable complexity. The presence of zero coefficients in these models also highlights the efficiency of the Generalized Least Deviation Method (GLDM) in eliminating non-contributory predictors, thereby streamlining the models for better performance and interpretability. This tailored approach ensures that each model is not only statistically sound but also practically relevant, providing robust tools for forecasting and further analysis. As we move forward, these insights will guide the continuous refinement of our time series modeling techniques, enhancing their predictive accuracy and utility in real-world applications.

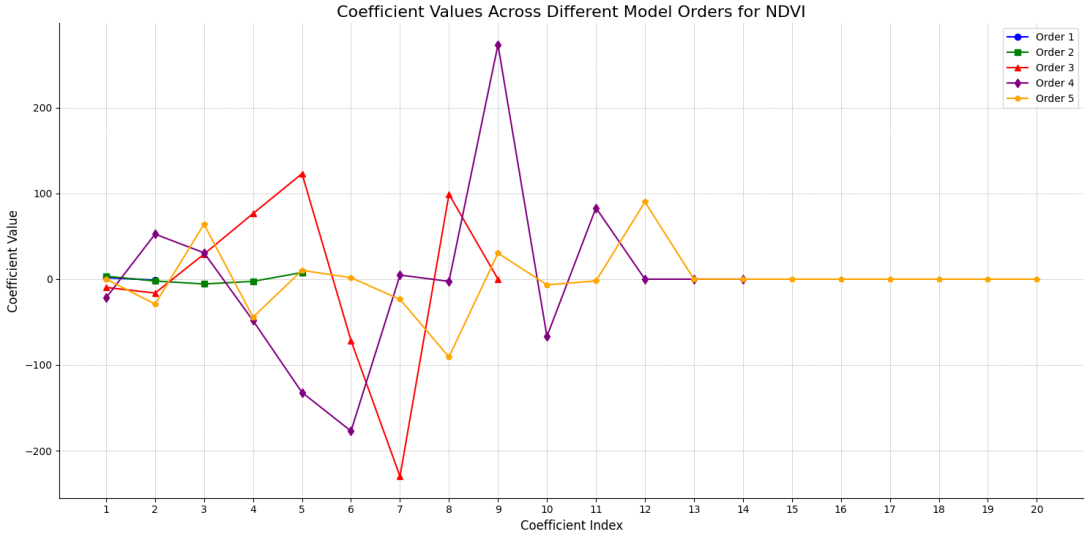

Figure 7: Coefficient values for NDVI across different model orders, showing variability and stability with increasing order.

The second-order Generalized Least Deviation Method (GLDM) model has been identified as optimal for the NDVI dataset, particularly given the data's span of 15 points. As shown in Figure 7 and detailed in Table 8, the second-order coefficients exhibit a marked decrease in magnitude compared to the first-order model (Table 7), suggesting a more refined fit to the data without overfitting, which could be especially problematic given the short data length.

A noteworthy observation from the coefficient analysis for the second-order model is the value of certain coefficients approaching zero. This behavior indicates the potential redundancy

of corresponding terms in the model. For instance, as we progress to models of higher orders, such as the third (Table 9) and fifth orders (Table 11), the presence of coefficients exactly equal to zero becomes apparent. These zero-valued coefficients are indicative of the GLDM's inherent regularization capability, which effectively eliminates non-contributory predictors from the model, enhancing both computational efficiency and the interpretability of the model. In the case of the NDVI dataset, the optimal second-order model balances the need for capturing relevant data dynamics against the risk of overfitting due to a limited number of data points. This balance is crucial for achieving robust predictive performance in time series forecasting.

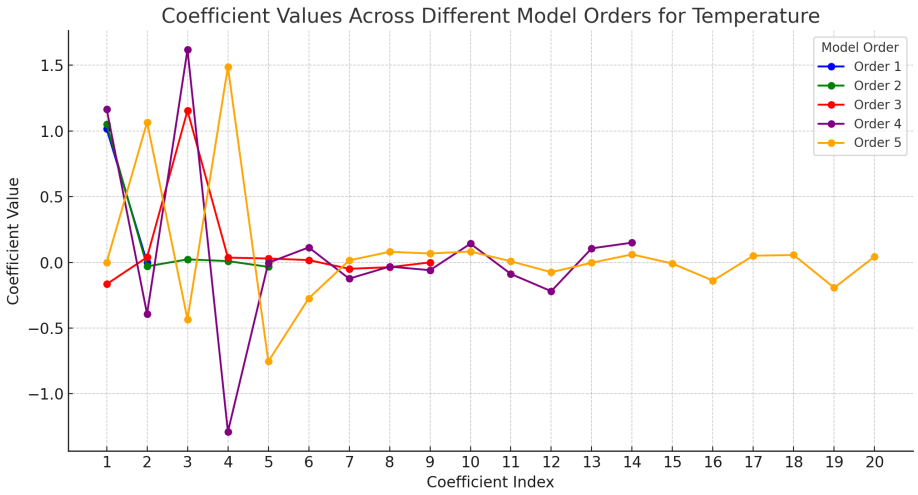

Figure 8: Dynamics of coefficient values across model orders for temperature data, illustrating the influence of different lags and interactions.

Figure 8 showcases the dynamics of temperature forecasting model coefficients across different orders. The fifth-order model, as elaborated in Table 16, emerges as the most adept for the temperature dataset, which comprises an extensive 9,939 data points. This data volume allows for the employment of a more sophisticated model without the typical overfitting concerns associated with smaller datasets.

The coefficients of the fifth-order model, notably with the exception of a_1 which is zero, reflect significant interactions among the lagged variables of the temperature series. The zero value of a_1 may indicate its negligible contribution to the model, reinforcing the concept of sparsity within the context of high-dimensional data modeling. Sparsity is advantageous as it can aid in the model's interpretability and potentially enhance generalization performance by reducing overcomplexity.

In this comprehensive dataset, the fifth-order model's capability to discern and assimilate complex patterns is demonstrated by the non-zero coefficients, which are instrumental in capturing the nuances and potential nonlinearities of temperature fluctuations.

In Figure 9, we present the coefficient values for the second-order model applied to wind speed data. This model order has been selected as the best fit for the dataset, which is quite extensive, totaling 50,530 data points. Despite the larger dataset size, a second-order model is deemed optimal rather than a fifth order, which might be counterintuitive given that a larger data length could potentially support more complex models. However, the selection of the second order is informed by the nature of the wind speed data, which, although extensive, likely does not exhibit the same degree of non-linear patterns or complex seasonal trends that are characteristic of the temperature data. This illustrates that the quantity of data does not necessarily dictate the complexity of the model needed. Instead, it's the underlying patterns and behaviors within the dataset that are paramount.

The coefficients of the second-order model, as seen in Table 18, indicate that the dynamics of wind speed can be captured effectively without resorting to higher-order complexity. This model order balances the desire for predictive accuracy with the need for a parsimonious model, reducing the risk of overfitting while maintaining computational efficiency, which is particularly advantageous in real-time forecasting applications.

Figure 9: Coefficient values for wind speed data, reflecting the model's response to historical data at different orders.

Figure 10: Analysis of coefficient values across different model orders for COVID-19 death cases in Russia, highlighting potential overfitting in higher orders.

Figure 10 illustrates the coefficient values across different model orders for COVID-19 death cases in Russia. The second-order model stands out as the optimal choice, as evidenced in Table 20. With the dataset encompassing 882 data points, the selection of the second-order model is informed by the data's specific trends and patterns rather than its size. This decision emphasizes the importance of choosing a model order that can accurately capture the data's dynamics while avoiding overfitting, a criterion met by the second-order model in this context.

The coefficient values, with a_1 being significantly larger than subsequent coefficients, signify the relevance of the most recent observations in forecasting future outcomes. The diminishing

magnitude of coefficients suggests a decrease in the influence of past data points as they become older. This progression underscores the model's focus on more immediate data points, which are more predictive of the near future in the context of COVID-19 death cases.

Through this careful analysis, we recognize the critical role played by the intrinsic characteristics of the dataset in determining the appropriate model order. The quantity of data alone does not dictate the complexity of the model required; instead, the model order is chiefly guided by the data's underlying behaviors and the need for a parsimonious model that maintains predictive accuracy without succumbing to overfitting.

7 Discussion

This section deliberates on the theoretical implications of the selected model orders for different datasets, discusses the role of zero coefficients in enhancing model efficiency, and considers potential avenues for future research.

7.1 Theoretical Implications of Model Order Selection

The choice of model order, as demonstrated in the results section, significantly impacts the model's ability to accurately capture the dynamics of the data. For Temperature, a fifth order model was found to be optimal, capturing complex interactions and seasonal patterns through a quasi-linear recurrence equation:

$$T_t = \sum_{i=1}^{20} a_i g_i \left((T_{t-k})_{k=1}^{5} \right) + \epsilon_t,$$

where g_i are functions representing the interactions of lagged values up to the fifth order. This model order selection is theoretically significant as it acknowledges the intrinsic properties of temperature data, which are characterized by substantial autocorrelation and periodicity.

Conversely, the second order model was deemed sufficient for Wind Speed and COVID-19 death cases, indicating a lesser complexity in their time series dependencies. For Wind Speed, the optimal model structure is expressed through a quasi-linear recurrence equation as follows:

$$W_t = \sum_{j=1}^{5} a_j g_j \left((W_{t-k})_{k=1}^{2} \right) + \epsilon_t,$$

where g_j are the functions representing the interactions up to the second order, incorporating the primary (W_{t-1}) and secondary (W_{t-2}) lagged values of wind speed. This suggests that the essential dynamics of wind speed can be sufficiently captured without the need for more complex higher-order terms. The chosen model thus balances simplicity and computational expediency, making it well-suited for real-time forecasting applications where speed is as valued as forecast accuracy.

Theorem 11 (Optimal Model Selection and Zero Coefficient Conditions in GLDM)**.** *Consider a time series $\{y_t\}_{t=1}^{T}$, and let the Generalized Least Deviation Method (GLDM) be utilized to minimize the objective function $F(\mathbf{a})$, defined as:*

$$F(\mathbf{a}) = \sum_{t=1}^{T} \arctan \left| y_t - f(\mathbf{a}, \mathbf{x}_t) \right|,$$

where $f(\mathbf{a}, \mathbf{x}_t)$ denotes the predicted value at time t, parameterized by the coefficient vector $\mathbf{a}$.

(i) Optimal Model Selection: *The optimal model order m^* is defined as:*

$$m^* = \arg\min_{m} \min_{\mathbf{a}} F(\mathbf{a}, m),$$

where m denotes the order of the model, and $f_m(\mathbf{a}, \mathbf{x}_t)$ represents the model utilizing past observations up to order m.

(ii) Zero Coefficient Conditions: *A coefficient a_j within the vector $\mathbf{a}$ may be set to zero if the following condition holds:*

$$\frac{\partial F(\mathbf{a})}{\partial a_j} = \sum_{t=1}^{T} \frac{1}{1 + (y_t - f(\mathbf{a}, \mathbf{x}_t))^2} \cdot \frac{\partial f(\mathbf{a}, \mathbf{x}_t)}{\partial a_j} = 0.$$

This condition indicates that the contribution of a_j to the minimization of $F(\mathbf{a})$ is negligible. Such a scenario typically arises when:

- *The feature corresponding to a_j exhibits redundancy or strong collinearity with other features in the model.*
- *The feature associated with a_j exerts an insignificant influence on the prediction accuracy.*

Discussion of Theorem 11: Theorem 11 provides a formal and systematic approach to both selecting the optimal model order and identifying redundant coefficients within the Generalized Least Deviation Method (GLDM) framework.

- **Optimal Model Selection:** The optimal model order m^* is obtained by minimizing the objective function $F(\mathbf{a}, m)$, which aggregates the arctangents of the absolute residuals between observed and predicted values. This objective function is particularly robust against outliers, making the model selection process more reliable even in the presence of anomalous data.
- **Zero Coefficient Conditions:** The theorem rigorously establishes the criterion under which a coefficient a_j can be set to zero without compromising the model's predictive accuracy. This occurs when the partial derivative of the objective function with respect to a_j is zero, indicating that the corresponding feature does not contribute meaningfully to the model. The ability to identify and remove such coefficients enhances model parsimony, reduces computational complexity, and mitigates overfitting.

This theorem underpins the construction of efficient and interpretable forecasting models within the GLDM framework, ensuring that the models are not only computationally efficient but also capable of maintaining high predictive accuracy with minimal complexity.

Table 22: Comparison of GLDM with MLP and SVM

Criterion	GLDM	MLP	SVM
Robustness to Outliers	+	-	+
Computational Complexity	+	-	-
Model Interpretability	+	-	+
Flexibility in Model Structure	+	+	+
Handling of Seasonality	+	-	-
Handling of Nonlinearity	+	+	+
Sensitivity to Noise	+	-	+

Table 23: Comparison of GLDM with AutoARIMA and Exponential Smoothing

Criterion	GLDM	AutoARIMA	Exponential Smoothing
Robustness to Outliers	+	-	-
Computational Complexity	+	+	+
Model Interpretability	+	+	+
Flexibility in Model Structure	+	-	-
Handling of Seasonality	+	+	+
Handling of Nonlinearity	+	-	-
Sensitivity to Noise	+	-	-

Table 24: Comparison of GLDM with BATS, TBATS, Prophet, and Hybrid Models

Criterion	GLDM	BATS	TBATS	Prophet	Hybrid Models
Robustness to Outliers	+	+	+	+	+
Computational Complexity	+	-	-	+	-
Model Interpretability	+	+	+	+	+
Flexibility in Model Structure	+	+	+	+	+
Handling of Seasonality	+	+	+	+	+
Handling of Nonlinearity	+	+	+	-	+
Sensitivity to Noise	+	+	+	+	+

Tables 22, 23, and 24 provide a comparative analysis of the Generalized Least Deviation Method (GLDM) against various popular forecasting models. The "+" and "-" symbols indicate whether each model possesses certain characteristics, with "+" signifying the presence of a feature and "-" indicating its absence.

The assessment of the models based on each criterion in Tables 22, 23, and 24 was conducted using the following approach:

- **Robustness to Outliers**: "+" indicates that the model maintains performance in the presence of outliers, while "-" suggests significant sensitivity to outliers.
- **Computational Complexity**: "+" signifies that the model is computationally efficient, requiring fewer resources, while "-" indicates higher computational demands.
- **Model Interpretability**: "+" denotes that the model?s operations and results are easy to understand, while "-" indicates complexity that makes interpretation difficult.
- **Flexibility in Model Structure**: "+" shows that the model can adapt to various data structures, while "-" means it has limited adaptability.
- **Handling of Seasonality**: "+" means the model can effectively manage seasonal patterns in data, while "-" suggests it struggles with seasonality or requires specific configurations.
- **Handling of Nonlinearity**: "+" indicates that the model can capture nonlinear relationships in data, while "-" indicates it is better suited to linear patterns.
- **Sensitivity to Noise**: "+" represents stability in the presence of noise, while "-" indicates that the model's performance deteriorates with noisy data.

These evaluations were based on empirical testing and theoretical understanding, ensuring that the comparison accurately reflects the models' strengths and weaknesses in typical forecasting scenarios.

8 Conclusion

In this study, we have rigorously analyzed time series data from various domains, including environmental and epidemiological fields, employing the Generalized Least Deviation Method (GLDM) to identify the optimal model order for forecasting. Our results demonstrate that the complexity required for a predictive model is highly contingent on the dataset's characteristics, such as the nature of the data, its underlying dynamics, and the presence of non-linear patterns, rather than solely on the quantity of data available.

For temperature data, characterized by significant seasonal fluctuations and autocorrelations, a fifth order GLDM model proved to be most adept at capturing the inherent complexity. On the other hand, wind speed and COVID-19 death cases, despite the substantial data length for the former, were accurately modeled using a second order GLDM model. This was sufficient

to encompass the essential dynamics without overfitting, showcasing the model's efficiency in real-time forecasting scenarios where computational expediency is essential. Furthermore, the occurrence of zero coefficients in the higher-order models signified the method's effectiveness in excluding non-contributory predictors, thereby simplifying the models and bolstering their interpretability. This study reinforces the pivotal role of careful model order selection and the judicious interpretation of model coefficients in the context of time series forecasting.

Future research directions include exploring the integration of machine learning techniques to compare their performance with GLDM models, examining the impact of varying regularization parameters, and applying these methodologies to other complex time series datasets. Through such efforts, we aim to further refine forecasting models, enhancing their predictive power and practical application in diverse fields.

$$T_t = \sum_{i=1}^{20} a_i g_i \left((T_{t-k})_{k=1}^{5} \right) + \epsilon_t,$$

$$W_t = \sum_{j=1}^{5} a_j g_j \left((W_{t-k})_{k=1}^{2} \right) + \epsilon_t,$$

In essence, our investigation elucidates the nuanced requirements for model specification in time series analysis, which is crucial for both theoretical understanding and practical forecasting. The insights gained from this study are expected to have significant implications for data-driven decision-making processes across various sectors.

9 Future Research Directions

The current study opens several promising avenues for future research. One significant direction involves the integration of non-linear machine learning models, providing a comparative analysis against the Generalized Least Deviation Method (GLDM). This approach could further elucidate the strengths and limitations of GLDM relative to more complex algorithms. Exploring the applicability of GLDM in various contexts is another important area for future research. Specifically, applying GLDM to intricate datasets, such as those found in financial time series or high-frequency trading environments, would offer valuable insights into the generalizability and robustness of the model across diverse contexts.

The methodologies and findings presented in this research have the potential to significantly enhance existing modeling strategies. By building upon these foundations, future studies can contribute to a more nuanced and comprehensive understanding of complex data dynamics, ultimately improving the accuracy and reliability of forecasting models.

Abbreviations

Table 25: List of Abbreviations and Symbols

Abbreviation	Explanation
GLDM	Generalized Least Deviations Method
NDVI	Normalized Difference Vegetation Index
RT	Real Time
FH	Forecasting Horizon
ARIMA	AutoRegressive Integrated Moving Average
MSE	Mean Squared Error
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
MAPE	Mean Absolute Percentage Error
$\mathbf{a_i}$	Coefficients in the models
y_t	Observed state variables at time t
ϵ_t	Unknown errors in the model
t	Time index
m	Model order
$n(m)$	Number of coefficients determined by model order m
T	Total number of time instants
g_j	Functions given in the model
GLDM Estimator	Algorithm for calibrating data to determine factors
Predictor	Algorithm to forecast future values based on data
arctan	Arc tangent function, used in optimization tasks
WLDM	Weighted Least Deviations Method
DualWLDMSoluter	Algorithm for solving dual tasks in WLDM
Cauchy Distribution	Probability distribution used in robust statistics
min	Operation to find the minimum value
max	Operation to find the maximum value
Dual Problem	Derived problem that complements the primal problem in optimization
Primal Problem	Original problem in an optimization scenario from which the dual is derived

Funding

The study is supported by the Russian Science Foundation regional grant number 23-21-10009.

References

Abotaleb, M. S. A., Makarovskikh, T. (2021). Analysis of neural network and statistical models used for forecasting of disease infection cases. In *Proceedings of the 2021 International Conference on Information Technology and Nanotechnology (ITNT)* (pp. 1–7). IEEE.

Abotaleb, M. O. (2024a). Solving the optimizing parameters problem for non-linear datasets using the high-order general least deviations method (GLDM) algorithm. In *Computational Methods for Differential Equations*. University of Tabriz.

Abotaleb, M. (2024b). Soft computing-based generalized least deviation method algorithm for modeling and forecasting COVID-19 using quasilinear recurrence equations. *Iraqi Journal for Computer Science and Mathematics*, *5*(3), 441–472.

Arnold, B.C., Beaver, R.J. (2000). The skew-Cauchy distribution. *Statistics & Probability Letters*, *49*(3), 285–290. https://doi.org/10.1016/S0167-7152(00)00072-0

Gasimov, Y.S., Nachaoui, A., & Niftiyev, A.A. (2010). Non-linear eigenvalue problems for p-Laplacian with variable domain. *Optimization Letters*, 4, 67–84. https://doi.org/10.1007/s11590-009-0041-x

Ilhan, O.A., Manafian, J. (2019). Periodic type and periodic cross-kink wave solutions to the (2+1)-dimensional breaking soliton equation arising in fluid dynamics. *Modern Physics Letters B*, *33*(23), 1950277. https://doi.org/10.1142/S0217984919502772

Ilhan, O.A., Manafian, J., Lakestani, M., & Singh, G. (2022). Some novel optical solutions to the perturbed nonlinear Schrödinger model arising in nano-fibers mechanical systems. *Modern Physics Letters B*, *36*(03), 2150551. https://doi.org/10.1142/S021798492150551X

Juraev, D.A., Gasimov, Y.S. (2022). On the regularization Cauchy problem for matrix factorizations of the Helmholtz equation in a multidimensional bounded domain. *Azerbaijan Journal of Mathematics*, *12*(1), 142–161.

Maceachin, J.J.Jr. (1986). Analysis of variance in image processing of correlated data (Markov, tensor product spaces, texture, edge detection) [PhD thesis, Polytechnic University].

Makarovskikh, T., Abotaleb, M. (2021). Comparison between two systems for forecasting COVID-19 infected cases. In *Computer Science Protecting Human Society Against Epidemics: First IFIP TC 5 International Conference, ANTICOVID 2021, Virtual Event, June 28–29, 2021, Revised Selected Papers 1* (pp. 107–114). Springer.

Makarovskikh, T., Panyukov, A., & Abotaleb, M. (2023). Using general least deviations method for forecasting of crops yields. In *International Conference on Mathematical Optimization Theory and Operations Research* (pp. 376–390). Springer.

Makarovskikh, T., Abotaleb, M. (2020). Improving the identification algorithm for a quasilinear recurrence equation. In *Advances in Optimization and Applications: 11th International Conference, OPTIMA 2020, Moscow, Russia, September 28–October 2, 2020, Revised Selected Papers* (pp. 15–26). Springer.

Neto, A.B.S., Ferreira, T.A.E., Batista, M.C.M., & Firmino, P.R.A. (2020). Studying the performance of cognitive models in time series forecasting. *Revista de Informática Teórica e Aplicada*, *27*(1), 83–91.

Pachal, R., Kumar, B. (2021). Forecasting industrial electric power consumption using regression-based predictive model. In *Recent Trends in Communication and Electronics: Proceedings of the International Conference on Recent Trends in Communication and Electronics (ICCE-2020), Ghaziabad, India, 28-29 November, 2020* (p. 135). CRC Press.

Panyukov, A.V., Tyrsin, A.N. (2006). Robust construction of regression models based on the generalized least absolute deviations method. *Journal of Mathematical Sciences*, 139, 6634–6642. https://doi.org/10.1007/BF03033149

Panyukov, A.V., Mezaal, Y.A. (2018). Stable estimation of autoregressive model parameters with exogenous variables on the basis of the generalized least absolute deviation method. *IFAC-PapersOnLine*, *51*(11), 1666–1669. https://doi.org/10.1016/j.ifacol.2018.04.507

Panyukov, A.V., ... (2018). Stable identification of linear autoregressive model with exogenous variables on the basis of the generalized least absolute deviation method. *Bulletin of the South Ural State University. Series: Mathematical Modelling and Programming*, *11*(1), 35–43.

Panyukov, A., Makarovskikh, T., & Abotaleb, M. (2022). Forecasting with using quasilinear recurrence equation. In *International Conference on Optimization and Applications* (pp. 183–195). Springer.

Pan, J., Wang, H., & Yao, Q. (2007). Weighted least absolute deviations estimation for ARMA models with infinite variance. *Econometric Theory*, *23*(5), 852–879. https://doi.org/10.1017/S0266466607005580

Panyukov, A., Panyukov, A., & Abotaleb, M. (2023). Forecasting with using quasilinear recurrence equation. In *International Conference on Optimization and Applications* (pp. 183–195). Springer.

Panyukov, A.V., Mezaal, Y.A. (2020). Using general least deviations method for forecasting of crops yields. In *International Conference on Mathematical Optimization Theory and Operations Research* (pp. 376–390). Springer.

Sirotin, D.V. (2020). Neural network approach to forecasting the cost of ferroalloy production. *Izvestiya Vysshikh Uchebnykh Zavedenii. Chernaya Metallurgiya*, *63*(1), 78–83.

Tyrsin, A.N. (2003). Robust construction of regression models based on the generalized least absolute deviations method. *Journal of Mathematical Sciences*, 139, 6634–6642. Springer.

Yakubova, D.M. (2019). Econometric models of development and forecasting of black metallurgy of Uzbekistan. *Asian Journal of Multidimensional Research (AJMR)*, *8*(5), 310–314. TRANS Asian Research Journals.