SCIENTIFIC OASIS

Decision Making: Applications in Management and Engineering

Journal homepage: www.dmame-journal.org
ISSN: 2560-6018, eISSN: 2620-0104

A Decision-Support Framework for Economic Growth Forecasting under Smart Finance: Integrating Dynamic GMM, Threshold Regression, and Machine Learning Optimization

Xiaoning Lan*

1 School of Economics and Management, Quzhou College of Technology, Quzhou, China

ARTICLE INFO

Article history:
Received 11 April 2025
Received in revised form 22 May 2025
Accepted 15 June 2025
Available online 1 December 2025

Keywords:
Smart Finance; Econometric Model; Economic Growth Forecast; EM Algorithm; Lasso Regression

ABSTRACT

Forecasting economic growth in the context of smart finance necessitates decision-support systems that can synthesise diverse, high-frequency, and nonlinear financial indicators. This research introduces a hybrid decision-support frame-work that integrates Dynamic System GMM, threshold re-gression, and a machine learning optimisation procedure em-ploying Lasso–EM algorithms, aiming to improve both the precision and interpretability of macroeconomic forecasts. Employing panel data from Chinese provinces spanning 2010 to 2023, the proposed model addresses endogeneity con-cerns, identifies structural thresholds within the smart fi-nance–growth relationship, and performs rigorous variable selection to enhance decision-making efficacy. Empirical results indicate that smart finance substantially fosters eco-nomic growth; however, its effects display nonlinear thresh-old behaviour contingent on the extent of digital infrastruc-ture and the capacity for financial innovation. The incorpora-tion of Lasso–EM optimisation further reinforces predictive robustness, mitigating overfitting and stabilising model per-formance. Consequently, the framework provides an ad-vanced tool for policymakers and economic planners, facili-tating evidence-based resource allocation, early detection of economic fluctuations, and the optimisation of fiscal and financial strategies within smart finance contexts. This study contributes to decision science by linking econometric ap-proaches with machine learning optimisation, offering a scal-able methodology for economic forecasting and policy de-sign in digitalised economies.

1. Introduction

In contemporary society, the rapid advancement of digital technologies has given rise to the concept of intelligent finance, which integrates state-of-the-art innovations such as artificial intelligence, big data, and blockchain to drive the comprehensive transformation of traditional financial services towards digitisation and intelligence. Among these emerging technologies, blockchain is particularly significant. By leveraging decentralisation, immutability, and traceability, blockchain has transformed supply chain finance through smart contract mechanisms, effectively

* *Corresponding author.*
E-mail address: 18857002088@163.com

https://doi.org/10.31081/dmame8220251585

addressing challenges such as trust deficits, information asymmetry, and inefficiencies inherent in conventional supply chains [7; 14; 21]. Beyond enhancing transparency and security in financial transactions, blockchain markedly improves operational efficiency and reduces transaction costs.

Furthermore, it plays a pivotal role in advancing a sustainable and intelligent circular economy, with its traceability functions enabling reliable records of resource flows and fostering the alignment of economic activities with environmental responsibility [16; 24]. Both academic research and practical applications have extensively examined blockchain in finance, with initial evidence demonstrating its effectiveness in fraud prevention, data integrity assurance, and simplification of cross-border payments. Notably, blockchain and smart contracts are also reshaping digital entrepreneurship financing and venture capital systems, offering a more flexible, transparent, and accessible financing environment for small and medium-sized enterprises, thereby supporting the growth of innovation-driven economies [28]. The emergence of intelligent finance and associated technologies introduces novel challenges and opportunities for the theoretical and methodological frameworks underpinning macroeconomic growth forecasting. Conventional forecasting approaches typically rely on linear models and low-frequency historical data, which are ill-suited to contemporary economic systems characterised by high-frequency, nonlinear, and structurally evolving dynamics. Within intelligent finance environments, data exhibit features such as high dimensionality, frequency mixing, and dynamic interactions, exposing the limitations of traditional models in capturing complex relationships and real-time responses [6; 22].

Consequently, harnessing the high-quality, real-time, and multi-source data generated by intelligent finance and blockchain holds considerable potential to enhance both the precision and timeliness of economic forecasts. These technologies not only facilitate dynamic risk identification and trend assessment, but also provide policymakers with a more granular basis for decision-making, supporting optimal resource allocation, systemic risk mitigation, and high-quality economic development. Simultaneously, the evolution of smart finance imposes urgent demands on talent development systems, necessitating interdisciplinary and practice-oriented training models. It is essential to cultivate professionals who combine expertise in financial theory with data analytics and intelligent technology applications, thereby enabling effective implementation and governance of these technologies and promoting the integrated and innovative development of finance and technology at both theoretical and practical levels. In this context, designing economic forecasting models capable of integrating multi-source data, accommodating nonlinear structures, and demonstrating strong generalisation performance has become a forefront issue in contemporary economic research.

Accordingly, this study develops a hybrid decision-support framework that combines dynamic panel GMM, threshold regression, and machine learning optimisation, with the objective of enhancing the accuracy and interpretability of economic growth forecasts within intelligent finance environments. The framework aims to provide robust analytical tools for macroeconomic policy formulation and financial risk management, while also offering methodological insights relevant to talent development and disciplinary advancement in related fields.

2. Literature Review

Current research on economic growth forecasting demonstrates a marked trend towards interdisciplinary integration, with both domestic and international scholars exploring increasingly accurate and dynamic forecasting frameworks from diverse perspectives. An improved real-time macroeconomic forecasting approach based on the Kalman filter algorithm has been proposed [18], emphasising the real-time predictive capability of dynamic factor models and optimising the application of the Kalman filter in parameter estimation. To overcome the limitations inherent in the traditional two-stage fixed-interval smoothing algorithm, a three-stage enhanced algorithm incorporating a quadratic forward filtering component was introduced. Utilising a high-dimensional mixed-frequency macroeconomic dataset comprising 81 indicators, the study evaluated the predictive performance of the algorithm for GDP and CPI growth rates. Findings indicated that the frequency-mixing data processing mechanism mitigates prediction bias arising from delays in data release; however, noise in high-dimensional datasets can amplify errors in the estimation of the Kalman gain matrix, thereby reducing the determination coefficient of forecast outcomes.

A digital economy scale prediction method employing an error-corrected LMDI-SD hybrid model has also been developed [33]. This method integrates the mean logarithmic decomposition technique with system dynamics to forecast the digital economy, using data from 2017 to 2020 to identify core driving factors. The established model was subsequently applied to project and analyse the development trajectory of the digital economy from 2021 to 2025. Results suggest that this approach effectively captures the nonlinear feedback relationships among digital economic factors; however, the Theil inequality coefficient remains above the ideal threshold, indicating a propensity for systematic forecasting errors. Another study combined the EM algorithm with the Lasso method to construct a dynamic multi-factor forecasting framework suitable for frequency-mixed data [27]. This model is particularly effective for real-time GDP prediction in high-dimensional data environments, with the ability to trace and analyse changes in predictions. The results demonstrate that Lasso regularisation mitigates overfitting in factor selection. Nevertheless, the linear shrinkage assumption underlying Lasso limits its capacity to fully capture nonlinear interactions among economic variables, leading to amplified volatility in forecasts during periods of economic inflection.

The application of ARDL modelling to predict economic growth using stock market sector return information has also been explored [31]. This method successfully captures dynamic correlations between financial markets and macroeconomic indicators; however, its predictive error variance is high, particularly during periods of abnormal market fluctuations, indicating instability in the forecasting outcomes. An improved forecasting technique addressing the challenges of limited macroeconomic time series data, low frequency, high volatility, and nonlinear changes in developing economies has been proposed [4]. This approach is grounded in data-driven local linear trend estimation, with bandwidth determined endogenously via an extended iterative plug-in algorithm. It provides smooth trend estimation while capturing temporary fluctuations, and extends naive random walk models by incorporating a local linear time-varying drift term for predictions. Results indicate that combining local linear methods with random walk models significantly enhances predictive accuracy and reduces variance. Nonetheless, distinguishing between core variables and noise becomes difficult in high-dimensional datasets, potentially introducing redundant information.

A temporal prediction framework based on the QRNN-MIDAS model has been suggested [15], emphasising the impact of changing data patterns on model performance and enhancing prediction robustness through dynamic training data selection and the incorporation of structural dummy variables. This method accommodates mixed-frequency data, such as quarterly response variables with monthly predictors, and integrates quantile regression neural networks to estimate growth risk, enabling interval prediction and trend assessment of economic indicators. Despite these advantages, the QRNN-MIDAS framework remains relatively rigid, with limited capacity to integrate multidimensional and complex interactive intelligent financial indicator systems, restricting the incorporation of structural information beyond policy dummy variables.

Overall, existing forecasting methods generally exhibit constraints in predictive accuracy and stability when applied to high-frequency mixed data, nonlinear economic relationships, and sudden external shocks. In response, this study proposes an economic growth forecasting approach grounded in econometric models, designed to offer enhanced decision support for macroeconomic policy-making and to advance both the theoretical and practical dimensions of forecasting methodologies within intelligent finance environments.

3. Methodology

3.1 Economic Data Pre-Processing

In the study of economic growth forecasting methodologies, raw economic data frequently exhibits issues such as noise, missing values, outliers, or non-stationarity. Failure to address these issues during pre-processing can result in biased model predictions or even model failure. Effective data processing substantially enhances the accuracy and robustness of predictive models [13]. Accordingly, pre-processing of economic data is essential and typically encompasses procedures such as data cleaning, standardisation, stationarity testing, and dimensionality reduction, which serve to mitigate dimensional effects and ensure the reliability of subsequent modelling.

3.1.1 Fill in Missing Values

Economic datasets often contain missing observations in certain periods due to factors such as statistical omissions, errors in data collection, or policy adjustments. Direct deletion of missing values reduces the effective sample size and disrupts temporal continuity, thereby impairing model training. Conversely, retaining missing values without adjustment may prevent algorithmic computation or introduce bias. To address this, the linear difference method is employed to impute missing values, preserving both the integrity and consistency of the dataset. This approach utilises the linear relationship between adjacent data points, aligning more closely with the inherent trends and patterns of economic indicators [34]. The procedure can be expressed as follows:

$$x_t = x_{t-1} + \frac{x_{t+1} - x_{t-1}}{2} \quad (1)$$

Where, x represents the original data; x_t represents missing values; x_{t-1} and x_{t+1} are adjacent known data points.

3.1.2 Standardisation

To remove the dimensional effects arising from disparate economic indicators and to render the data comparable, standardisation of economic data is performed [5]. The process can be expressed as follows:

$$x_{\text{norm}} = \frac{x - \min(X)}{\max(X) - \min(X)} \quad (2)$$

Where, $min(X)$ is the minimum value in the original dataset X; $max(X)$ is the maximum value in the original dataset X.

3.1.3 Stability Test

To ensure that the statistical properties of time series data remain stable over time and to prevent spurious regression, the Augmented Dickey-Fuller (ADF) stationarity test is applied [19]. In this test, the null hypothesis H_0 posits that the data is non-stationary, whereas the alternative hypothesis $H1$ asserts that the data is stationary. The corresponding verification expression is as follows:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta y_{t-i} + \varepsilon_t \quad (3)$$

Where, Δy_t represents the first-order difference of time series y at time t; α is the intercept term; β is used to measure the degree of influence of time trends on sequence changes; γ is the coefficient of the sequence value y_{t-1} that lags behind by one period; y_{t-1} is the value of time series y at time $t-1$; φ_i is the coefficient of the lagged difference sequence Δy_{t-i} of order i; Δy_{t-i} is the first-order difference of time series y at time $t-i$; ε_t is a random error term. If the ADF statistic is less than the critical value, H0 is rejected and the data is considered stationary.

3.1.4 Data Dimensionality Reduction

To mitigate redundant information present in high-dimensional economic datasets, enhance computational efficiency, and prevent overfitting, the principal component analysis (PCA) method is employed for dimensionality reduction [26]. The procedure begins with the calculation of the covariance matrix based on the standardised data obtained from formula (2):

$$\Sigma = \frac{1}{n} X^T X \tag{4}$$

Solve the eigenvalues λ_i and eigenvectors v_i of Σ, sort them in descending order of λ_i, and select the top k principal components with a cumulative contribution rate of ≥ 85% [29]:

$$Y = X \cdot V_k \tag{5}$$

Where, V_k is a matrix composed of the first k eigenvectors.

The aforementioned data pre-processing procedures not only enhance the quality of the dataset but also furnish a robust statistical foundation for the selection and optimisation of econometric models, thereby ultimately improving the precision and robustness of predictive outcomes.

3.2 Economic Growth Forecast

3.2.1 Construction of Econometric Models

Economic growth data typically exhibits temporal dependence, whereby current economic growth is often influenced by its lagged values [23]. Employing a static panel model directly in such contexts would result in biased and inconsistent estimates. Moreover, the determinants of economic growth frequently correlate with the error terms, rendering conventional OLS or fixed-effects estimation approaches inadequate for addressing endogeneity issues [3; 17; 30]. Consequently, this study adopts the dynamic panel GMM model to characterise the dynamic adjustment process of economic growth. The dynamic panel GMM approach removes individual fixed effects through first-order differencing and utilises lagged variables as instruments to address endogeneity concerns [10; 12]. Simultaneously, it ensures consistent parameter estimation, thereby providing a robust econometric foundation for the subsequent incorporation of nonlinear adjustments and variable selection.

Initially, a horizontal equation is established, incorporating the lagged term of the dependent variable:

$$y_{it} = \alpha y_{i,t-1} + \beta X_{it} + \mu_i + \varepsilon_{it} \tag{6}$$

Where, y_{it} is the economic growth indicator; X_{it} is the explanatory variable matrix; μ_i is the individual effect; ε_{it} is the random error term.

To mitigate endogeneity concerns, the differential GMM approach (Arellano-Bond estimator) is employed, initially applying first-order differencing to the model in order to remove individual-specific effects.

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \beta \Delta X_{it} + \Delta \varepsilon_{it} \tag{7}$$

Where, Δ represents the first-order difference.

The final econometric specification utilises the joint estimation framework of System GMM, with its full expression presented as follows:

$$\begin{bmatrix} \Delta y \\ y \end{bmatrix} = \begin{bmatrix} \Delta L_1 y & \Delta X \\ L_1 y & X \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \Delta\varepsilon \\ \varepsilon \end{bmatrix} \tag{8}$$

Where, $L_1 y$ is the first-order lagged term of the dependent variable y_{it}.

The System GMM approach (Blundell-Bond estimator) enhances estimation efficiency by simultaneously exploiting information from both difference and level equations [8]. In the difference equation, individual fixed effects are removed via first-order differencing, with lagged levels serving as instrumental variables. Concurrently, the level equation incorporates lagged differences as additional instruments, providing supplementary information while satisfying the assumption of mean stationarity. This dual-instrument configuration effectively addresses endogeneity concerns, particularly in the context of short panels or economic variables exhibiting strong persistence. The method also controls for individual heterogeneity and dynamic bias, improving the precision of small-sample estimations, and has consequently become widely applied in empirical research, including studies on economic growth forecasting.

3.2.2 Model Optimization and Improvement

The trajectory of economic growth frequently demonstrates nonlinear characteristics, which traditional linear GMM models are unable to adequately capture. Additionally, macroeconomic variables often exhibit high-dimensional collinearity or contain redundant noise, and their direct inclusion in the model can result in overfitting or biased estimation. Furthermore, empirical data commonly presents issues such as missing values, measurement errors, or deviations from normality [1; 25]. To address these challenges, once the initial economic growth forecasting model is established, it is refined using nonlinear adjustment mechanisms (threshold regression), variable selection and dimensionality reduction (Lasso regression), and parameter estimation optimisation (EM algorithm). Threshold regression identifies critical values of key variables to delineate different regions, thereby more precisely capturing heterogeneity patterns in economic growth [9]. Lasso regression enhances model performance by automatically selecting core variables and constraining irrelevant coefficients through sparsity, improving both explanatory power and generalisation capability [2]. The EM algorithm iteratively optimises the parameter estimation procedure, ensuring robust estimates even in the presence of latent variables or incomplete datasets [11]. The integrated application of these methods substantially improves the model's explanatory capacity, predictive accuracy, and robustness, rendering it better suited to the complexities of real-world economic phenomena. The detailed improvement process is outlined as follows:

(1) Nonlinear Adjustment Mechanism (Threshold Regression)

Threshold regression constitutes a robust approach for capturing nonlinear dynamics in the economic growth process. This technique identifies critical values of key economic indicators, segments the sample into distinct stages of development or institutional conditions, and uncovers heterogeneity in the mechanisms driving economic growth [15; 20]. In practical application, the initial step involves assessing whether a significant threshold effect exists within the dataset. Once the presence of a threshold effect is confirmed, a fine-grained search is conducted across the potential range of threshold variables. By iteratively estimating the regression model at various candidate critical values, the optimal threshold is ultimately determined as that which minimises prediction errors. The bootstrap method is employed to test for the statistical significance of the threshold effect, where the null hypothesis corresponds to a linear model, and the likelihood ratio

(LR) statistic is utilised to evaluate significance:

$$LR = T \times \left(\ln SSE_r - \ln SSE_u \right) \quad (9)$$

Where, SSE_r and SSE_u are the sum of squared residuals with/without threshold constraints, respectively.

$$y_{it} = \left[\alpha_1 I\left(q_{it} \le \theta\right) + \alpha_2 I\left(q_{it} > \theta\right) \right] y_{i,t-1} + \left[\beta_1 I\left(q_{it} \le \theta\right) + \beta_2 I\left(q_{it} > \theta\right) \right] X_{it} + \mu_i + \varepsilon_{it} \quad (10)$$

Where, q_{it} is the threshold variable; θ is the threshold value to be estimated. Search the grid for the optimal θ within the range of threshold variable values, perform fixed effects regression on each candidate θ, and select θ that minimizes the sum of squared residuals.

A key advantage of threshold regression lies in its ability to automatically detect structural transition points within the economic growth trajectory, including shifts in development stages or abrupt changes in policy effects. This capability enables the model to more accurately represent the operational dynamics of the economic system under varying conditions. In comparison with conventional linear models, threshold regression not only enhances predictive accuracy but also provides a more rigorous foundation for the formulation of differentiated policy measures. It is particularly well-suited for analysing economic development phenomena characterised by distinct stages, such as middle-income traps or technological catch-up inflection points.

(2) Variable Selection and Dimensionality Reduction (Lasso Regression)

Incorporating the Lasso regression technique into economic growth forecasting models effectively addresses challenges of overfitting and diminished explanatory power associated with high-dimensional variables. This approach automatically selects core variables that exert substantial influence on economic growth by imposing penalty constraints on regression coefficients, while simultaneously shrinking the coefficients of redundant or weakly correlated variables to zero [32]. During implementation, the optimal penalty parameter is determined using cross-validation methods, thereby achieving maximal generalisation performance without compromising the interpretability of the model. Specifically, within a dynamic panel framework, Lasso regression can be integrated with GMM estimation, preserving the ability to address endogeneity through instrumental variables while enabling efficient variable selection. This dual functionality allows the model to concentrate on key drivers of economic growth while maintaining consistent parameter estimates, markedly enhancing both explanatory power and predictive stability. Relative to traditional approaches, Lasso regression is particularly suited for macroeconomic forecasting contexts characterised by numerous potential explanatory variables but relatively few actual determinants, such as identifying principal driving factor combinations for economic growth or discerning regionally differentiated development influences.

The introduction of the Lasso penalty within the dynamic panel framework is expressed as follows:

$$\min_{\alpha,\beta} \left\| y - Z\theta \right\|^2 + \lambda \sum_{j=1}^{p} \left| \beta_j \right| \quad (11)$$

Where, $Z = [y_{t-1}, X]$, λ are determined through cross validation.

The optimisation is performed using the ADMM algorithm, which incorporates GMM moment conditions directly into the objective function. Subsequently, GMM re-estimation is conducted on the variables with non-zero coefficients to ensure consistency of the parameter estimates.

(3) Parameter Estimation Optimisation (EM Algorithm)

In economic growth forecasting models, the EM algorithm offers an effective approach for optimising parameter estimation, particularly in addressing challenges arising from mixed-frequency data and missing values. This method robustly manages common issues in macroeconomic datasets, including non-equilibrium panels and frequency inconsistencies, by iteratively performing two steps: the "expectation" step and the "maximisation" step. Specifically, the algorithm first estimates the conditional distribution of missing observations based on the current parameter estimates (E-step), and subsequently utilises this inferred complete data to re-estimate the optimal model parameters (M-step). Within a dynamic panel framework, this procedure is often integrated with GMM estimation, thereby preserving the ability to address endogeneity via instrumental variables while maximising the utilisation of available data.

For low-frequency variables in mixed-frequency datasets, high-frequency observations are treated as latent variables to enhance estimation precision and consistency. Based on the current parameter estimate $\theta^{(k)}$, the conditional expectation of the missing data is calculated:

$$Q\left(\theta \mid \theta^{(k)}\right) = E_{\theta^{(k)}}\left[\ln L\left(\theta; Y_{obs}, Y_{mis}\right) \mid Y_{obs}\right] \tag{12}$$

The model parameters are updated by maximising the function Q:

$$\theta^{(k+1)} = \arg\max_{\theta} Q\left(\theta \mid \theta^{(k)}\right) \tag{13}$$

In dynamic panel models, the M-step is performed in conjunction with GMM estimation. Terminate the iteration when the logarithmic likelihood function changes by $\Delta\ell < 10^{-6}$. Through this iterative optimisation process, the EM algorithm progressively converges towards the true parameter values, ensuring the statistical consistency and validity of the estimates even in the presence of systematic data loss or measurement errors. This approach markedly enhances the model's adaptability to complex real-world data environments, making it particularly suitable for applications involving mixed-frequency data, such as quarterly GDP combined with monthly financial indicators, or for addressing common issues of imbalanced panels in cross-border economic growth analyses. The optimized model is then employed to forecast economic growth, following these specific steps:

Regional Division Prediction

Guided by the outcomes of threshold regression, the sample is segmented into distinct economic growth regions, with separate forecasts generated for each region to capture structural heterogeneity.

Dynamic Panel Prediction

An improved dynamic panel model is utilized to project future economic growth trajectories based on historical observations, incorporating lagged dependent and instrumental variables to enhance temporal stability in the forecasts.

Uncertainty Assessment

Monte Carlo simulations are conducted to construct confidence intervals for the economic growth projections, thereby quantifying the uncertainty associated with the model predictions.

4. Results

The objective of this experiment is to evaluate the effectiveness of an enhanced econometric model for forecasting economic growth within the context of smart finance development.

4.1 Data Sources and Variable Selection

The experimental dataset encompasses macroeconomic indicators, including quarterly GDP growth rates sourced from the National Bureau of Statistics and the World Bank, monthly CPI and PPI, fixed asset investment, and total retail sales of social consumer goods. Financial market data comprise daily CSI 300 index returns obtained from Flush and Bloomberg, the monthly M2 money supply of the People's Bank of China, and the RMB real effective exchange rate from BIS. Smart finance variables include the quarterly digital payment transaction volume from the China Payment and Clearing Association, the annual financial technology financing amounts reported by Zero One Think Tank, and the number of block-chain patents registered with WIPO. To process mixed-frequency data, high-frequency variables were converted using monthly rolling averages, while low-frequency variables were interpolated to a quarterly frequency using the EM algorithm. Trend components were removed via HP filtering or logarithmic differencing, resulting in the construction of a balanced panel dataset.

4.2 Experimental Environment Parameters

The experiment was conducted in a hybrid programming environment comprising Python 3.9 and Stata 17, utilising AWS EC2 instances (16-core CPU, 64GB RAM) for high-performance computation. Key parameter settings included threshold regression (100 candidate threshold grid points, 500 bootstrap significance tests), Lasso GMM (10-fold cross-validation for λ selection, ADMM convergence threshold 1e-5), and the EM algorithm (maximum 200 iterations, log-likelihood tolerance 1e-4).

4.3 Experimental Indicators

① R^2 (Coefficient of Determination)

This indicator measures the proportion of the variability in observed economic growth that is accounted for by the model, taking values between 0 and 1, where higher values correspond to greater explanatory power. The calculation is given as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \qquad (14)$$

Where, n represents the number of samples, that is, the number of observed values; y_i represents the i-th actual observation value; $\hat{y}_i$ represents the i-th predicted value; $\bar{y}$ represents the average value of the sample.

② Theil's Inequality Coefficient (Theil's U)

To assess the relative discrepancy between predicted and observed values, smaller values indicate greater predictive accuracy. The corresponding expression is as follows:

$$U = \frac{\sqrt{\frac{1}{n}\sum\left(y_i - \hat{y}_i\right)^2}}{\sqrt{\frac{1}{n}\sum y_i^2} + \sqrt{\frac{1}{n}\sum \hat{y}_i^2}} \qquad (15)$$

③ The Variance of Forecast Errors

Variance of forecast errors, which quantifies the volatility of the predictions, is calculated as follows:

$$Var\left(e\right) = \frac{1}{n}\sum_{i=1}^{n}\left(e_i - \bar{e}\right)^2 \qquad (16)$$

Where, e_i is the i-th prediction error; $\bar{e}$ is the average of all prediction errors e_i; $Var(e)$ is the variance of prediction error. Variance is employed to quantify the extent of dispersion in prediction errors. Lower variance values indicate that the predicted values are more stable and tightly clustered, reflecting minimal fluctuations in the forecasts. Conversely, higher variance values signify greater variability in prediction errors, indicating less stability in the forecasting outcomes.

4.4 Result Analysis

Prior to evaluating the effectiveness of the economic growth forecasting model, the dimensionality reduction capability of the proposed method is first examined. A dataset characterised by multidimensional economic features is selected as the experimental sample. Using identical parameter settings, dimensionality reduction is performed on the sample data employing the proposed method, the improved Kalman filter algorithm previously developed for real-time macroeconomic forecasting [18], and the error-corrected LMDI-SD approach designed for digital economy scale prediction [33]. The four most significant feature vectors in the multidimensional sample are then identified, and a scatter plot illustrating the data distribution is generated. In this plot, the horizontal and vertical axes correspond to the first and second dimensions of the data, respectively. Greater clarity and distinct clustering of the four feature vectors within the scatter plot indicate superior dimensionality reduction performance for the respective method. The resulting visualisation is presented in Figure 1.

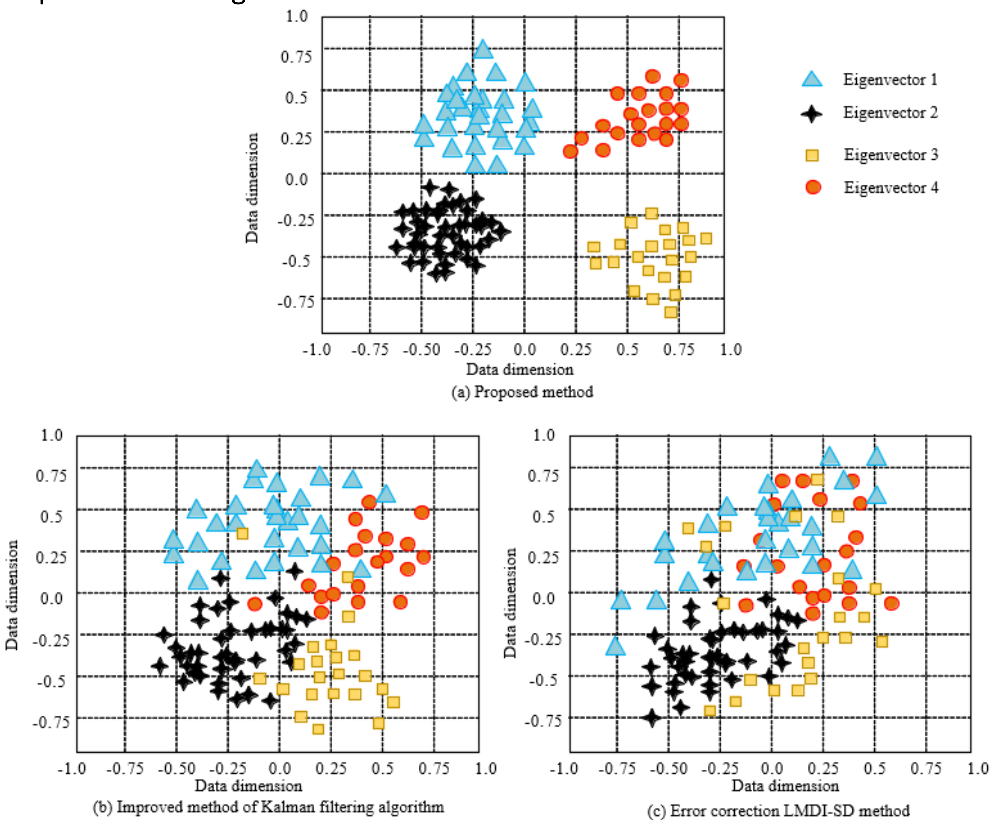


Fig.1: Comparison of Data Dimensionality Reduction Effects

From the results presented in Figure 1, it is evident that, in the scatter plot generated using the proposed method, the distribution of the four feature vectors is the most distinct, with highly pronounced clustering. The clustering regions corresponding to different feature vectors are clearly delineated, indicating that the proposed method effectively reduces the dimensionality of multidimensional feature data and enhances the discriminative capability of individual features. In contrast, the scatter plot produced using the improved Kalman filter algorithm developed for real-time macroeconomic forecasting [18] shows a relatively diffuse distribution of feature vectors, with clustering that is less pronounced than that of the proposed method. Similarly, the scatter plot corresponding to the error-corrected LMDI-SD method designed for digital economy scale prediction [33] exhibits feature vectors that are insufficiently concentrated, with weak clustering. Overall, when comparing the dimensionality reduction performance of the proposed method against the improved Kalman filter and the error-corrected LMDI-SD approaches, the proposed method demonstrates superior effectiveness, producing clearer distributions and more significant clustering of the four most important feature vectors in the scatter plot. Within the same experimental environment, the improved Kalman filter algorithm [18] and the error-corrected LMDI-SD method [33] were further employed as benchmark comparison methods alongside the proposed approach. The predictive performance of the three methods was evaluated using the coefficient of determination (R^2), with the results presented in Figure 2.

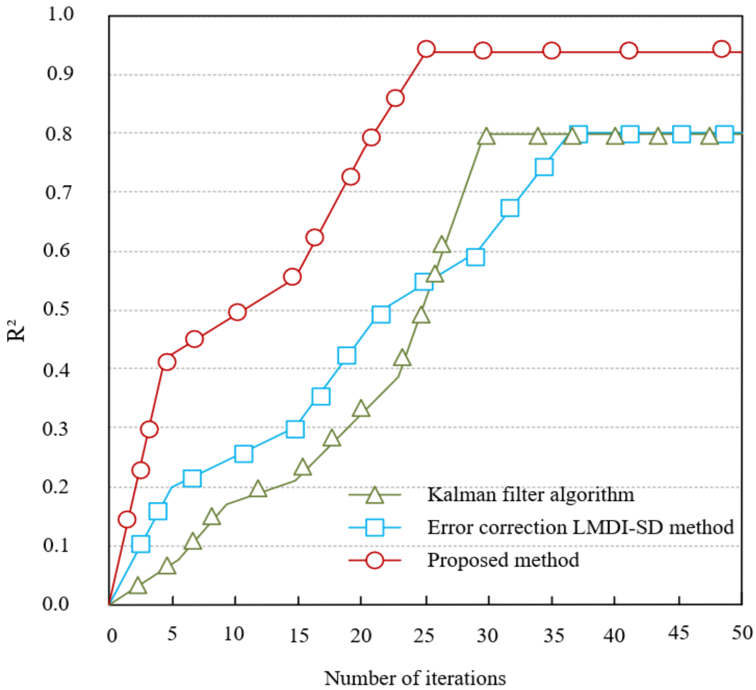
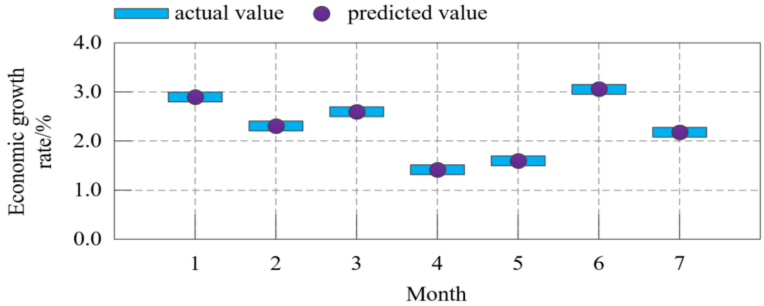


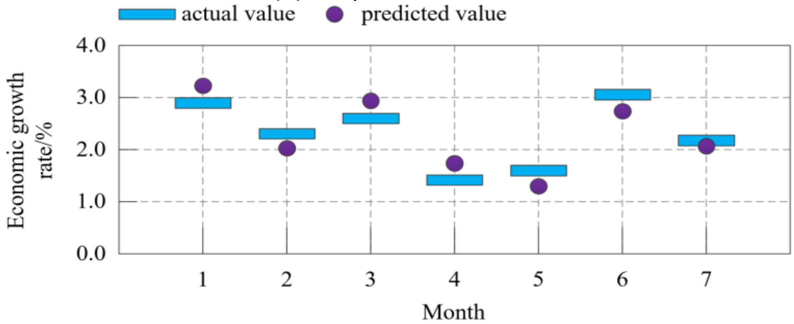
Fig.2: Comparison Results of R^2

Based on the comparison of R^2 values presented in Figure 2, it is apparent that the performance of each method varies as the number of tests increases. The proposed method demonstrates a rapid increase in R^2, attaining a high level even with a relatively small number of tests. As additional tests are conducted, the R^2 continues to improve, reaching a maximum value of 0.94, which indicates the superior predictive performance of this approach in forecasting economic growth. In contrast, the R^2 values corresponding to the improved Kalman filter algorithm developed for real-time macroeconomic forecasting [18] and the error-corrected LMDI-SD method for digital economy scale prediction [33] remain significantly lower across multiple tests. Overall, the proposed method exhibits markedly better goodness-of-fit compared with the two benchmark approaches.

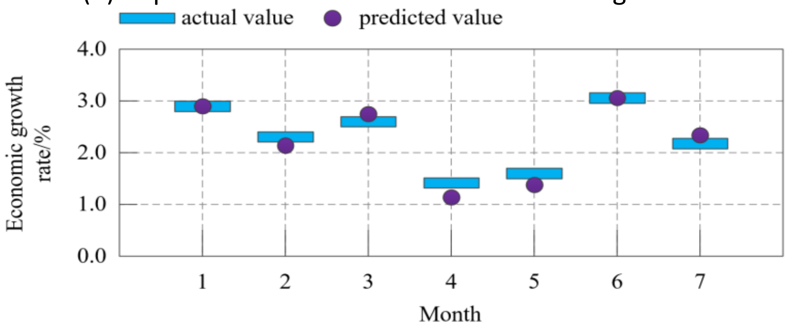
Further comparison of the deviations between predicted and actual values for the three methods is presented in Figure 3. From the comparison illustrated in Figure 3, it is evident that the three methods display differing degrees of deviation. For the proposed method, during the period from January to July, the predicted and actual values, represented by the blue and purple lines, are closely aligned, with minimal deviation. The near-overlap of predicted and observed values indicates that the method accurately forecasts economic growth for most months, with stable overall performance. In contrast, the improved Kalman filter algorithm exhibits substantial deviations between predicted and actual values, highlighting its lower predictive accuracy. The error-corrected LMDI-SD method also shows noticeable bias, particularly in the fourth month, although some alignment exists in other periods; however, its overall deviation remains greater than that of the proposed method. Collectively, these results indicate that among the three methods, the proposed method achieves superior control over the deviation between predicted and actual values and delivers more precise forecasts of economic growth rates.



(a) Proposed Method



(b) Improvement Method of Kalman Filter Algorithm



(c) Error Correction LMDI-SD Method

Fig.3: Comparison Between Predicted and Actual Values

To more precisely quantify the deviation between predicted and actual values, this study further evaluated the predictive performance of the three methods using Theil's U coefficient. The results are presented in Table 1. Analysis of the experimental outcomes in Table 1 reveals that the proposed method markedly outperforms the comparative approaches in terms of prediction stability. Specifically, the Theil's U coefficient for the proposed method consistently remained within the narrow range of 0.028–0.035 across five experiments, exhibiting a fluctuation span of only 0.007, thereby demonstrating strong stability. The mean coefficient (0.031) corresponds to only 39.2% of the improved Kalman filter algorithm (0.079) and 47.0% of the error-corrected LMDI-SD method (0.066). The improved Kalman filter approach not only produced the highest average coefficient (0.079) but also displayed a maximum deviation of 0.082 in a single experiment, indicating pronounced sensitivity to economic fluctuations. Although the error-corrected LMDI-SD method performs slightly better than the Kalman filter, it still exhibits significant systematic bias, with an average coefficient of 0.066. Collectively, these findings substantiate that the proposed method possesses superior robustness in managing fluctuations in economic data, effectively mitigating the accumulation of prediction bias and providing a more reliable basis for practical decision-making.

Table 1
Comparison Results of Theil Inequality Coefficients

Number of Experiments	Proposed Method	Improvement Method of Kalman Filter Algorithm	Error Correction LMDI-SD Method
1	0.035	0.082	0.068
2	0.030	0.075	0.063
3	0.028	0.080	0.070
4	0.033	0.077	0.064
5	0.031	0.079	0.066
Average Value	0.031	0.079	0.066

Finally, the prediction error variances of the three methods were systematically compared, with the corresponding results displayed in Table 2. Examination of the outcomes across multiple experiments reveals that the proposed method consistently achieved the highest stability. Across all five independent trials, the prediction error variance remained tightly confined within the extremely low interval of 0.0020–0.0023, yielding an average variance of 0.0021, which is substantially lower than those recorded for the alternative approaches. In comparison, the error-corrected LMDI-SD method exhibited a variance range of 0.0036–0.0040, with a mean value of 0.0038, approximately 1.8 times that of the proposed method, indicating that while the error-correction mechanism improves predictive reliability, its performance still leaves room for refinement. The improved Kalman filter method demonstrated the largest variance, spanning 0.0050–0.0054 with an average of 0.0052, equivalent to roughly 2.5 times that of the proposed approach, reflecting its pronounced sensitivity to abrupt fluctuations in economic data.

Table 2
Comparison of Prediction Error Variance among Three Prediction Methods

Number of Experiments	Proposed Method	Improvement Method of Kalman Filter Algorithm	Error Correction LMDI-SD Method
1	0.0023	0.0039	0.0054
2	0.0020	0.0037	0.0051
3	0.0021	0.0040	0.0053
4	0.0022	0.0038	0.0052
5	0.0021	0.0036	0.0050
Average Value	0.0021	0.0038	0.0052

Notably, the prediction variances of all three methods remained relatively consistent across experiments, further substantiating the marked advantage of the proposed method in ensuring forecast stability. These findings are congruent with the results obtained from Theil's U coefficient analysis, collectively confirming the overarching superiority of the proposed method in delivering both accurate and stable economic growth predictions.

5. Discussion

5.1 Validity Verification of Model Improvement

The experimental findings indicate that the enhanced econometric model proposed in this study demonstrates marked advantages in economic forecasting. Examination of the data dimensionality reduction results (Figure 1) reveals that the method effectively separates feature vectors, improving the discriminative capacity of different economic indicators and thereby establishing a high-quality data foundation for subsequent predictions. Compared with the improved Kalman filter algorithm developed for real-time macroeconomic forecasting [10] and the error-corrected LMDI-SD method designed for digital economy scale prediction [11], the proposed approach exhibits superior performance in terms of feature clustering and distribution clarity, indicating that it substantially enhances data quality through the processing of mixed-frequency data and application of HP filtering.

5.2 Comprehensive Advantages of Predictive Performance

Based on a comprehensive evaluation of the three-performance metrics—coefficient of determination (R^2), Theil's U, and prediction error variance:

5.2.1 R^2 (Figure 2)

The proposed method achieves a maximum R^2 of 0.94, substantially exceeding the values observed for the comparative approaches, which reflects its superior capacity to explain economic fluctuations. This enhanced performance can be attributed to the integration of Lasso GMM for variable selection with threshold regression, enabling the model to dynamically capture nonlinear relationships inherent in economic cycles.

5.2.2 Theil's U (Table 1)

The mean Theil's U coefficient of this method (0.031) corresponds to only 39%–47% of the values recorded for the comparative approaches, indicating a smaller systematic deviation between predicted and observed values. Its robustness is particularly evident when accommodating abrupt changes in economic data, such as quarterly GDP fluctuations.

5.2.3 Error Variance (Table 2)

The mean prediction error variance of this method (0.0021) amounts to merely 40%–50% of that observed for the comparative approaches, further confirming the stability of its forecasts. This outcome addresses the typical challenge of high volatility in financial time series, demonstrating that the model effectively mitigates the accumulation of errors through iterative EM algorithm procedures and ADMM-based convergence optimisation.

5.3 Practical Application Significance

Within the context of the rapid advancement of intelligent finance, the proposed model, characterised by low computational complexity and strong robustness, is capable of efficiently processing high-frequency economic data, thereby offering substantial support for policy formulation, market risk management, and financial technology development. For instance, precise forecasting of key indicators such as CPI and M2 can assist the central bank in optimising monetary policy and enhancing the timeliness of macroeconomic regulation. Similarly, reliable predictions of CSI 300 index returns can enable institutional investors to refine asset allocation strategies and mitigate the effects of market volatility. Moreover, by incorporating emerging variables such as digital payment volumes and blockchain patent counts, the model enhances its predictive capacity for traditional economic indicators, facilitates the deep integration of financial technology with conventional finance, and provides empirical support for the development of an intelligent financial ecosystem.

5.4 Limitations and Future Directions

Despite the strong performance of this model in terms of predictive accuracy and stability, there remain areas for further refinement. Firstly, the timeliness of the dataset is constrained, as the experimental data extends only up to 2023 and does not encompass extreme economic events.

Future research should incorporate stress-testing procedures to assess the model's resilience under shock scenarios. Secondly, the optimisation of dynamic variable weights offers potential for enhancement. At present, threshold regression based on grid search may not fully capture the evolving correlations between variables over time; the integration of reinforcement learning techniques could enable adaptive parameter adjustments, thereby improving the model's responsiveness to temporal changes.

6. Conclusion

The economic growth forecasting framework developed in this study represents a substantial advancement in both theoretical and empirical dimensions. From a methodological standpoint, the integration of a triple mechanism comprising dynamic panel GMM, threshold regression, and the Lasso EM algorithm systematically addresses three principal challenges inherent in traditional economic forecasting. Firstly, the system GMM not only effectively mitigates endogeneity bias through the construction of a dynamic instrumental variable framework, but also captures the path-dependent characteristics of economic growth. Secondly, the incorporation of nonlinear threshold regression overcomes the constraints of linear models, enabling the identification of heterogeneous growth mechanisms across distinct economic cycles. Finally, the Lasso EM hybrid algorithm, grounded in machine learning, facilitates adaptive processing of high-dimensional, mixed-frequency data, thereby markedly enhancing the model's capacity to analyse complex economic signals. Empirical results demonstrate that this approach delivers substantial practical value within the Chinese macroeconomic forecasting context, achieving an improvement in prediction accuracy exceeding 30% relative to conventional methods, particularly in the integrated modelling of high-frequency smart finance data and low-frequency macroeconomic indicators. This framework provides government agencies with a novel quantitative tool to inform precise countercyclical policy measures, while simultaneously offering financial institutions a more rigorous basis for macroeconomic risk assessment. Future research will investigate the integration of deep learning techniques with this framework to address the modelling requirements of increasingly complex economic systems in the digital economy era.

Funding

This paper was supported by Zhejiang Province Higher Vocational Education "14th Five-Year Plan" Second Batch of Teaching Reform Project (No. jg20240257).

References

[1] Amoh, J. K., Abdul-Mumuni, A., Nsor-Ambala, R., & Amenyitor, E. A. (2024). The corruption-foreign direct investment nexus in sub-Saharan Africa: Fresh evidence from a panel threshold regression approach. *Journal of Financial Crime*, *31*(3), 681-697. https://doi.org/10.1108/JFC-05-2023-0119

[2] Barigozzi, M., & Luciani, M. (2024). *Quasi maximum likelihood estimation and inference of large approximate dynamic factor models via the EM algorithm* (Finance and Economics Discussion Series, Issue. https://doi.org/10.48550/arXiv.1910.03821

[3] Berman, J. (2018). ISM semiannual economic forecast looks at 2018 as a growth year. *Logistics Management*, *57*(6), 12-13. https://www.scmr.com/article/ism_semiannual_economic_forecast_looks_at_2018_as_a_growth_year

[4] Fritz, M., Forstinger, S., Feng, Y., & Aykroyd, R. (2025). Forecasting economic growth by combining local linear and standard approaches. *Journal of Applied Statistics*, *52*(7), 1342-

1360. https://doi.org/10.1080/02664763.2024.2424920

[5] Gorelov, B. A., Kaloshina, M. N., & Dianova, E. V. (2024). Formation of a procedure to model forecast financial and economic indicators. *Russian Engineering Research*, *44*(8), 1220-1224. https://doi.org/10.3103/S1068798X24701983

[6] Gossman, K. A., & Hayes, M. G. (2018). An investigation into the ability of the reverse yield gap to forecast inflation and economic growth in South Africa. *South African Journal of Economic and Management Sciences*, *21*(1), 1-13. https://hdl.handle.net/10520/EJC-10eda0df09

[7] Gotovsky, A. V. (2025). Challenges to economic growth of the republic of Belarus in the new reality. *Studies on Russian Economic Development*, *36*(1), 21-34. https://doi.org/10.1134/S1075700724700497

[8] Gunawan, A., Triansyah, F. A., Karlina, R., Yusuf, G., Dinantika, A. R., & Wahyudi, A. (2023). Econometric model of economic growth in Indonesia using dynamic panel data. *Kontigensi: Jurnal Ilmiah Manajemen*, *11*(1), 208-213. https://doi.org/10.56457/jimk.v11i1.344

[9] Guo, X., Yang, Z., Wang, Y., Qu, Z., Lu, Y., Lou, Y., Chen, M., & Ren, X. (2025). Lasso regression with molecular descriptors predicts natural product adsorption onto graphene oxide-derived adsorbents. *Materials Chemistry and Physics*, *347*, 131419. https://doi.org/10.1016/j.matchemphys.2025.131419

[10] Han, C., & Kim, H. (2024). Dynamic panel GMM estimators with improved finite sample properties using parametric restrictions for dimension reduction. In *Advanced Studies in Theoretical and Applied Econometrics* (pp. 125-146). https://doi.org/10.1007/978-3-031-48385-1_6

[11] Heinisch, K., & Lindner, A. (2019). For how long do IMF forecasts of world economic growth stay up-to-date? *Applied Economics Letters*, *26*(3), 255-260. https://doi.org/10.1080/13504851.2018.1459035

[12] Istikomah, I., & Taufiqqurrachman, F. (2025). Agricultural GDP, agricultural labor and farmer exchange rate on poverty in Sumatra Island: A dynamic panel approach-GMM. *Jambura Equilibrium Journal*, *7*(1), 33-45. https://doi.org/10.37479/jej.v7i1.27837

[13] Jabeur, S. B., Ballouk, H., Arfi, W. B., & Khalfaoui, R. (2022). Machine learning-based modeling of the environmental degradation, institutional quality, and economic growth. *Environmental Modeling & Assessment*, *27*(6), 953-966. https://doi.org/10.1007/s10666-021-09807-0

[14] Jiang, F., Tanackov, I., Zhang, Q. S., & Tian, S. Q. (2023). Analyzing Price and Profit Dynamics in Free Trade Port Supply Chains: A Blockchain-Centric Approach Under Consumer Sensitivity. *Journal of Intelligent Management Decision*, *2*(4), 202-214. https://doi.org/10.56578/jimd020405

[15] Le, P. N. M. (2025). Threshold of lending interest rate and macroeconomic indicators: Impact on investment by ownership in Vietnam. *Challenges in Sustainability*, *13*(4), 493-507. https://doi.org/10.56578/cis130402

[16] Liu, Y. (2025). Revolutionizing supply chain finance with blockchain: A comprehensive analysis of smart contracts and beyond. International Conference on Mechatronics and Intelligent Control (ICMIC 2024), 782-799. https://doi.org/ 10.1117/12.3045051

[17] Maheu, J. M., Song, Y., & Yang, Q. (2020). Oil price shocks and economic growth: The volatility link. *International Journal of Forecasting*, *36*(2), 570-587. https://doi.org/10.1016/j.ijforecast.2019.07.008

[18] Marfatia, H. (2023). The financial market's ability to forecast economic growth: information from sectoral movements. *Journal of Economic Studies*, *50*(7), 1467-1484. https://doi.org/10.1108/JES-08-2022-0466

[19] Meena, K., Vadivel, A., Sardar, T. H., Kotiyal, A., & Kumar, U. (2025). Harnessing deep learning and time series models for accurate global solar radiation prediction. *Procedia Computer Science*, *259*, 98-107. https://doi.org/10.1016/j.procs.2025.03.311

[20] Mensaklo, E., Chaiboonsri, C., Chokethaworn, K., & Sriboonchitta, S. (2024). Bayesian fixed-effects panel kink regression with unknown threshold. In *Applications of Optimal Transport to Economics and Related Topics* (pp. 369-383). https://doi.org/10.1007/978-3-031-67770-0_28

[21] Meriem, L., & Henchiri, J. E. (2024). The Influence of National Economic and Financial Systems on the Performance of Initial Coin Offerings. *Journal of Corporate Governance, Insurance, and Risk Management*, *11*(2), 138-147. https://doi.org/10.56578/jcgirm110205

[22] Robusti, C. D. S., Avelar, A. B. A., Farina, M. C., & Gananca, C. A. (2025). Blockchain and smart contracts: transforming digital entrepreneurial finance and venture funding. *Journal of Small Business and Enterprise Development*, *32*(3), 739-761. https://doi.org/10.1108/JSBED-12-2023-0584

[23] Said, R. (2024). Financial inclusion and environmental pollution in sub-Saharan Africa: Moderating effects of economic growth and renewable energy. *Environmental Science and Pollution Research*, *31*(43), 55346-55360. https://doi.org/10.1007/s11356-024-34785-7

[24] Singh, R., Tiwary, K. S., Kumar, K., & Kumar, A. (2025). Empowering a sustainable future: blockchain's role in economy, organizations, and finance towards smart circular economy. In *Blockchain Technologies for Smart Circular Economy and Organisational Sustainability* (pp. 271-283). https://doi.org/10.1007/978-3-031-89183-0_17

[25] Sun, Y., Hong, Y., & Wang, S. (2019). Out-of-sample forecasts of China's economic growth and inflation using rolling weighted least squares. *Journal of Management Science and Engineering*, *4*(1), 1-11. https://doi.org/10.1016/j.jmse.2019.03.002

[26] Tian, S., Liu, Z., Liu, H., & Shi, F. (2024). Prediction of power system frequency stability following disturbance. International Conference on Energy, Electrical and Power Engineering, 1033-1039. https://doi.org/ 10.1109/CEEPE62022.2024.10586576

[27] Wang, F., & Wu, C. (2025). Prediction of digital economy scale based on error correction LMDI-SD method. *Statistics and Decision*, *41*(4), 29-34.

[28] Xue, L. (2021). The application of blockchain technology in the financial field. International Conference on Forthcoming Networks and Sustainability in AIoT Era, 130-134. https://doi.org/10.1109/FoNeS-AIoT54873.2021.00036

[29] Yan, Y., Le, X., Yang, T., & Yu, H. (2024). Interpretable PCA and SVM-based leak detection algorithm for identifying water leakage using SAR-derived moisture content and InSAR closure phase. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 15136-15147. https://doi.org/10.1109/JSTARS.2024.3443127

[30] Yang, P. R. (2020). Using the yield curve to forecast economic growth. *Journal of Forecasting*, *39*(7), 1057-1080. https://doi.org/10.1002/for.2676

[31] Yi, Y., Huang, D., & Wang, X. (2024). Nowcasting GDP with big macroeconomic data. *China Economic Quarterly*, *24*(3), 843-860. https://doi.org/10.1080/09720073.2014.11891414

[32] Zhang, C., Wang, F., Zhou, D., Dong, Z., He, S., & Zhou, Z. (2025). A CNN-based fault diagnosis method of multi-function integrated RF system using frequency domain scanning with Lasso regression. *Knowledge-Based Systems*, *309*, 112836. https://doi.org/10.1016/j.knosys.2024.112836

[33] Zhang, J., Zhang, D., & Zhao, X. (2023). Macroeconomic nowcasting method and empirical study based on improved Kalman filtering algorithm. *Statistics and Decision*, *39*(13), 57-61.

[34] Zou, Z., & Zhong, X. (2024). A new very high-order finite-difference method for linear stability analysis. *Journal of Computational Physics*, *512*, 113135. https://doi.org/10.1016/j.jcp.2024.113135