Известия Саратовского университета. Новая серия. Серия: Математика. Механика. Информатика. 2026. Т. 26, вып. 2. С. 296–301
Izvestiya of Saratov University. Mathematics. Mechanics. Informatics, 2026, vol. 26, iss. 2, pp. 296–301
https://mmi.sgu.ru DOI: https://doi.org/10.18500/1816-9791-2026-26-2-296-301
EDN: https://elibrary.ru/RYRBIP

Article

Stochastic framework for macroeconomic scenario forecasting using sparse graph optimization and CIR++ simulations

A. S. Podgornyi

Far Eastern Federal University, 10 Ajax Bay, Russky Island, Vladivostok 690922, Russia

Andrei S. Podgornyi, podgornyy_as@dvfu.ru, https://orcid.org/0009-0005-4211-4633, SPIN: 2675-7756, AuthorID: 1155133

Abstract. The article presents a comprehensive methodology for forecasting macroeconomic indicators for long-term planning in project finance. The purpose of the research is to develop a system of stochastic simulations capable of generating plausible scenarios of economic development, taking into account the relationships between various economic parameters. The methodology includes two key components: an algorithm for selecting significant predictors based on sparse graphs and the minimum Steiner tree, and a system of stochastic simulations integrating the CIR++ model with the Monte Carlo method. The author has developed an efficient algorithm for building regression models that takes into account structural relationships between economic indicators. The research material consisted of historical data on a wide range of Russia's macroeconomic indicators: GDP, inflation, interest rates, real estate price indices, and loan delinquency rates. The results of applying the methodology demonstrate high accuracy of forecasting on historical data and intuitively understandable behavior in the long term. Model validation is based on conceptual validity, systematic output analysis, and business logic verification rather than traditional point forecast metrics, which is appropriate for long-term scenario generation. PCST hyperparameter calibration methodology and extreme scenario modeling for tail risk assessment are presented. The system is capable of generating probabilistic scenarios with a horizon of up to 30 years, which allows assessing various aspects of risks, including extreme scenarios. The modular architecture of the system provides flexibility and adaptability to various economic conditions. The results of the research have practical significance for risk management in financial institutions and strategic planning in project finance.

Keywords: graph sparsity, feature selection, prize-collecting Steiner forest, regression, Monte Carlo method, Cox – Ingersoll – Ross Model, project finance

For citation: Podgornyi A. S. Stochastic framework for macroeconomic scenario forecasting using sparse graph optimization and CIR++ simulations. *Izvestiya of Saratov University. Mathematics. Mechanics. Informatics*, 2026, vol. 26, iss. 2, pp. 296–301. DOI: https://doi.org/10.18500/1816-9791-2026-26-2-296-301, EDN: RYRBIP

Научная статья
УДК 004.942:330.43

Стохастическая модель прогнозирования макроэкономических сценариев с использованием оптимизации разреженных графов и симуляций CIR++

А. С. Подгорный

Дальневосточный федеральный университет, Россия, 690922, г. Владивосток, о. Русский, п. Аякс, д. 10

Подгорный Андрей Сергеевич, старший преподаватель департамента математики Института математики и компьютерных технологий, podgornyy_as@dvfu.ru, https://orcid.org/0009-0005-4211-4633, SPIN: 2675-7756, AuthorID: 1155133

Аннотация. В статье представлена комплексная методология прогнозирования макроэкономических показателей для долгосрочного планирования в проектном финансировании. Целью исследования является разработка системы стохастических симуляций, способной генерировать правдоподобные сценарии экономического развития с учетом взаимосвязей различных экономических параметров. Методология включает два ключевых компонента: алгоритм выбора значимых предикторов на основе разреженных графов и минимального дерева Штейнера, а также систему стохастических симуляций, интегрирующую модель CIR++ с методом Монте-Карло. Автором разработан эффективный алгоритм построения регрессионных моделей, учитывающий структурные взаимосвязи экономических показателей. Материалом исследования послужили исторические данные по широкому спектру макроэкономических показателей России: ВВП, инфляция, процентные ставки, индексы цен на недвижимость, показатели просроченной задолженности по кредитам. Результаты применения методологии демонстрируют высокую точность прогнозирования на исторических данных и интуитивно понятное поведение в долгосрочной перспективе. Валидация моделей основана на концептуальной обоснованности, системном анализе выходных данных и проверке бизнес-смысла, что соответствует специфике генерации долгосрочных сценариев. Представлена методика калибровки гиперпараметров алгоритма PCST и описано моделирование экстремальных сценариев для оценки хвостовых рисков. Система способна генерировать вероятностные сценарии с горизонтом до 30 лет, что позволяет оценивать различные аспекты рисков, включая экстремальные сценарии. Модульная архитектура системы обеспечивает гибкость и адаптивность к различным экономическим условиям. Результаты исследования имеют практическую значимость для управления рисками в финансовых институтах и стратегического планирования в проектном финансировании.

Ключевые слова: разреженность графов, отбор признаков, дерево Штейнера с призами, регрессия, метод Монте-Карло, модель Кокса–Ингерсолла–Росса, проектное финансирование

Для цитирования: *Podgornyi A. S.* Stochastic framework for macroeconomic scenario forecasting using sparse graph optimization and CIR++ simulations [*Подгорный А. С.* Стохастическая модель прогнозирования макроэкономических сценариев с использованием оптимизации разреженных графов и симуляций CIR++] // Известия Саратовского университета. Новая серия. Серия: Математика. Механика. Информатика. 2026. Т. 26, вып. 2. С. 296–301. DOI: https://doi.org/10.18500/1816-9791-2026-26-2-296-301, EDN: RYRBIP

Introduction

Project finance necessitates decision-making based on long-term macroeconomic forecasts. Existing models frequently fail to account for the complex interdependencies between diverse economic indicators, leading to significant forecasting errors over extended horizons [1]. This research aims to overcome these limitations by developing an integrated approach that synthesizes machine learning techniques with economic modeling. Classical regression models and ARIMA frameworks exhibit inherent constraints when applied to long-term forecasting, as they inadequately accommodate unexpected economic shifts and are insufficient to model the uncertainty and volatility characteristic of macroeconomic factors [2, 3].

1. Research objective

The primary research objective is the development of a flexible and scalable forecasting system for macroeconomic indicators capable of incorporating structural relationships between variables and generating probabilistic scenarios for long-term risk analysis.

The following specific tasks were defined to achieve this objective:

1) developing a methodology for selecting significant predictors utilizing sparse graphs and the Steiner tree algorithm;

2) integrating regression models into a stochastic simulation system employing the CIR++ (Cox – Ingersoll – Ross) model and Monte Carlo methods;

3) evaluating the efficacy of the proposed approach using data from the Russian economy, including simulations for GDP, inflation, and interest rates.

1.1. Materials and research methods

This research presents an integrated approach to constructing a forecasting system. This approach encompasses a methodology for selecting macroeconomic predictors via sparse graph analysis, followed by their utilization within a stochastic simulation system designed to generate probabilistic economic scenarios. Scenario generation is based on the CIR++ model [4–6] and Monte Carlo methods applied to macroeconomic indicators such as interest rates, housing price indices, mortgage rates, and other variables, explicitly incorporating their correlational dependencies. The methodology combines model calibration to market data, stochastic simulation with monthly discretization [7, 8], and regression relationships between variables. The coefficients of these regression relationships are optimally determined using the Steiner tree algorithm on sparse graphs.

Utilizing an implementation of the minimum Steiner tree algorithm (henceforth also referred to as PCST — Prize-Collecting Steiner Tree) [9] for feature selection [10], we identify a subgraph that maximizes the aggregate prize of selected vertices while minimizing the total cost of selected edges. The algorithm identifies a subgraph minimizing the cost function:

$$C(T) = \sum_{v_i \in V \setminus V_T} p_i + \sum_{e_{ij} \in E_T} w_{ij}, \tag{1}$$

where V_T denotes the set of selected vertices, and E_T denotes the set of selected edges.

As a practical implementation of the proposed method, regressions were employed for forecasting macroeconomic indicators.

1.2. Research results and discussion

In all the aforementioned models, the PCST algorithm was successfully applied for feature selection, ensuring the incorporation of both the individual significance of variables and their interdependencies. This enabled the creation of stable and accurate models with high explanatory power, suitable for scenario modeling and forecasting macroeconomic indicators. The application of the PCST algorithm enhanced model quality by integrating feature interdependencies, thereby improving the reliability and interpretability of the results.

Visualization of the regression of the key rate on the 1-month zero-coupon yield spread is shown in Fig. 1.

Based on the pre-constructed regression models, a comprehensive system for stochastic macroeconomic simulation was developed (experimental code: https://github.com/andreypodgorny10/macro_sims). The architecture of the system consists of four interconnected modules.

1. *Interest rate modeling module (CIR++).* Interest rates are simulated using the stochastic CIR++ model, modified to align with the observed market yield curve. The latent short rate $x(t)$ follows the stochastic differential equation

$$dx_t = a(\theta - x_t)\, dt + \sigma \sqrt{x_t}\, dW_t, \tag{2}$$

and the observable short rate is computed as

$$r_t = x_t + \varphi(t), \tag{3}$$

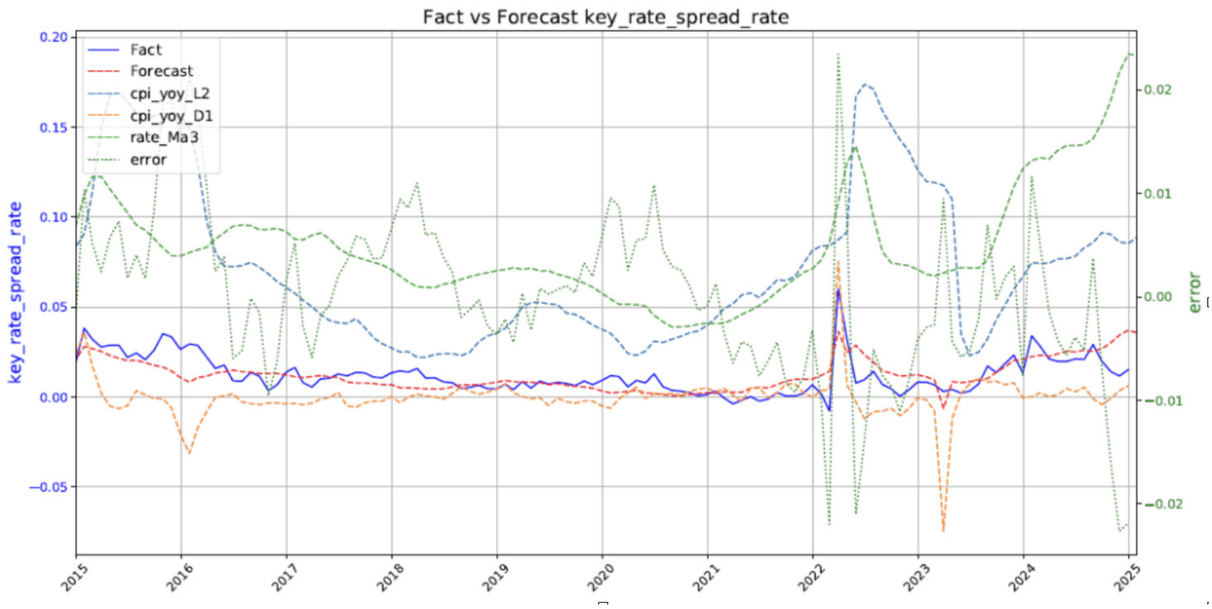

Fig. 1. Example implementation of PCST for key rate to 1-month zero-coupon yield spread forecasting. Compiled by the article author (color online)

where $\varphi(t)$ is a deterministic shift function calibrated to match the market forward curve. The model is discretized using the Euler – Maruyama scheme with a monthly step size $\Delta t = 1/12$:

$$x_{t+1} = x_t + \left[a\,\theta_0 - (a + B_{T-t}\,\sigma^2\,\mathbf{1}_{\text{T-measure}})\,x_t\right]\Delta t + \sigma\sqrt{x_t}\,\varepsilon_t. \tag{4}$$

The functions B_{T-t} and $\varphi(t)$ are computed analytically during the calibration stage based on the forward rate curve.

2. *Module for generating correlated stochastic shocks* [11]. A correlation matrix R between macroeconomic variables is constructed from historical data. Cholesky decomposition is used to simulate correlated shocks: $L = \text{chol}(R)$. Independent standard normal vectors $Z \sim \mathcal{N}(0, I)$ are transformed into correlated vectors $\varepsilon = LZ$. The resulting shocks are simultaneously applied to all variables within the regression models, ensuring internally consistent scenario paths.

3. *Regression modeling module for macroeconomic variables.* Each macroeconomic variable is modeled using a stochastic regression that includes lag terms, differences, moving averages, and interaction effects. The general structure of the regression model for a variable Y_t is

$$Y_t = \beta_0 + \sum_{i=1}^{k} \beta_i X_{i,t-\ell_i} + \sum_{j=1}^{m} \gamma_j (Z_{j,t} - Z_{j,t-d_j}) + \sum_{p=1}^{n} \delta_p Y_{t-p} + \sigma\,\varepsilon_t, \tag{5}$$

where $X_{i,t-\ell_i}$ are lagged explanatory variables; $(Z_{j,t} - Z_{j,t-d_j})$ are first- or higher-order differences; Y_{t-p} are autoregressive terms; and $\varepsilon_t \sim \mathcal{N}(0, 1)$ represents correlated shocks per Module 2.

4. *Simulation orchestration.* Summary statistics include Value-at-Risk (VaR), Expected Shortfall (ES), median-based scenario aggregates, and confidence intervals (e.g., 5th and 95th percentiles). The Monte Carlo method [12] is employed to generate the distribution of possible future states for each macroeconomic variable. Correlations are incorporated via Cholesky decomposition. The final implementation element is the visualization module. Examples of the visualization module output are presented in Fig. 2.

1.3. Numerical validation and performance assessment

Model validation focuses on conceptual validity, systematic output analysis, and business logic verification rather than traditional point forecast metrics, which is appropriate for long-term scenario generation. Conceptual validity confirms economic interpretability (e.g., inflation and monetary policy relationships) and the mean-reversion and non-negativity properties of the CIR++ model. Output analysis checks preservation of correlation structure, absence of implausible discontinuities or explosive behavior over 30-year horizons, and convergence of long-term mean values to economically reasonable equilibria. Business logic verification confirms that HPI, mortgage spreads, and GDP dynamics are consistent with market behavior and development stages. A hyperparameter calibration strategy for PCST (threshold $\tau = 0.3$, sparsity bounds, cluster count, and adaptive step sizes) was validated via sensitivity tests; ±20% changes produced $< 5\%$ variation in performance metrics. Extreme scenarios are captured via CIR++ dynamics, correlation propagation, and $N = 10{,}000$ Monte Carlo paths, enabling VaR/ES estimation and coverage of historical stress events within the probabilistic envelope.

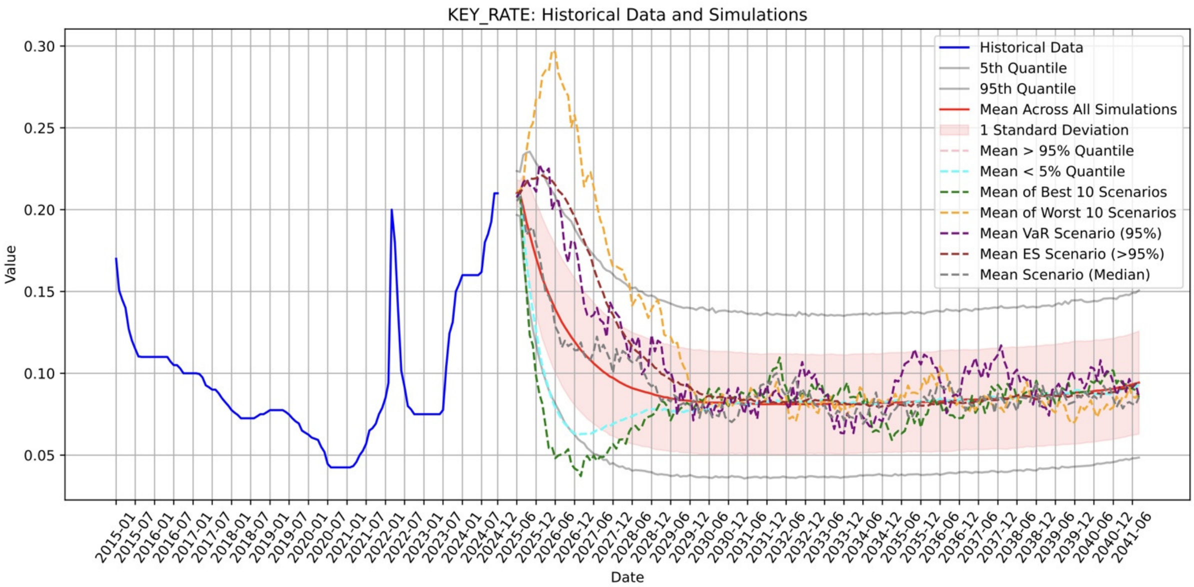

Fig. 2. Example implementation of the visualization module for the key rate. Compiled by the article author (color online)

Conclusion

The results of this research demonstrate the effectiveness of the proposed approach in forecasting macroeconomic indicators while accounting for their interdependencies and stochastic nature. The key advantages are flexibility, scalability, and adaptability to diverse economic conditions. The approach enables better risk assessment for long-term residential real estate projects and improves financial planning accuracy.

References

1. Bekaert G., Engstrom E., Xu N. R. The time variation in risk appetite and uncertainty. *Management Science*, 2022, vol. 68, iss. 6, pp. 3975–4004. DOI: https://doi.org/10.1287/mnsc.2021.4068
2. Antolín-Díaz J., Petrella I., Rubio-Ramirez J. F. Structural scenario analysis with SVARs. *Journal of Monetary Economics*, 2021, vol. 117, pp. 798–815. DOI: https://doi.org/10.1016/j.jmoneco.2020.06.001
3. Kaushik M., Giri A. *Forecasting foreign exchange rate: A multivariate comparative analysis between traditional econometric, contemporary machine learning & deep learning techniques*. 2020. DOI: https://doi.org/10.48550/arXiv.2002.10247
4. Akhtari B., Li H. The Cox – Ingersoll – Ross process under volatility uncertainty. *Journal of Mathematical Analysis and Applications*, 2024, vol. 531, iss. 1, pt. 1, art. 127867. DOI: https://doi.org/10.1016/j.jmaa.2023.127867
5. Bayraktar E., Clément E. Estimation of a pure-jump stable Cox – Ingersoll – Ross process. *Bernoulli*, 2025, vol. 31, iss. 1, pp. 484–508. DOI: https://doi.org/10.48550/arXiv.2304.02386

6. Chen R.-R., Scott L. Multi-Factor Cox – Ingersoll – Ross models of the term structure: Estimates and tests from a Kalman filter model. *The Journal of Real Estate Finance and Economics*, 2003, vol. 27, pp. 143–172. DOI: https://doi.org/10.1023/A:1024736903090
7. Poignard B., Asai M. High-dimensional sparse multivariate stochastic volatility models. *Journal of Time Series Analysis*, 2022, vol. 44, iss. 1, pp. 4–22. DOI: https://doi.org/10.1111/jtsa.12647
8. van Rhijn J., Oosterlee C. W., Grzelak L. A., Liu S. Monte Carlo simulation of SDEs using GANs. *Japan Journal of Industrial and Applied Mathematics*, 2023, vol. 40, pp. 1359–1390. DOI: https://doi.org/10.1007/s13160-022-00534-x
9. Hegde C., Indyk P., Schmidt L. A nearly-linear time framework for graph-structured sparsity. *Proceedings of the 37th International Conference on Machine Learning*, 2015, vol. 37, pp. 928–937.
10. Yu G., Liu Y. Sparse regression incorporating graphical structure among predictors. *Journal of the American Statistical Association*, 2016, vol. 111, iss. 514, pp. 707–720. DOI: https://doi.org/10.1080/01621459.2015.1034319
11. Fernández-Villaverde J., Guerrón-Quintana P. A. Uncertainty shocks and business cycle research. *Review of Economic Dynamics*, 2020, vol. 37, pp. 118–146. DOI: https://doi.org/10.1016/j.red.2020.06.005
12. Hull J. C. *Options, futures, and other derivatives*. Moscow, Williams Publ. House LLC, 2014. 1072 p. (in Russian).

Поступила в редакцию / Received 29.10.2025
Принята к публикации / Accepted 25.11.2025
Опубликована / Published 01.06.2026