NATIONAL BANK OF KAZAKHSTAN

Selective-Combined Inflation Forecasting System (SSCIF)

Monetary Policy Department
Economic Study No.2024-13
Working Paper

Adilkhanova Z.
Yerzhan I.

Economic studies and research notes of the National Bank of the Republic of Kazakhstan (“the NBK”) are meant to disseminate the outcomes of the NBK’s studies as well as of other academic research works of the NBK’s staff. Economic studies are disseminated in order to encourage discussions.

Working papers of the NBK’s staff reflect in-process studies as at the date of publication and are meant for discussion and receiving comments and remarks. Opinions and judgments presented in the paper are only personal opinions and judgments of authors and should not be regarded as those reflecting the view of the NBK and its top management.

Selective-Combined Inflation Forecasting System (SSCIF)

December 2024

NBRK – WP – 2024 – 13

ISSN: 2789-150X

Selective-Combined Inflation Forecasting System (SSCIF)

Adilkhanova Zarina[1]
Yerzhan I.[2]

Abstract

In an environment of macroeconomic instability, improving the accuracy of inflation forecasting is a priority for central banks, especially those operating under inflation targeting regimes. Traditional econometric models face limitations in accounting for volatility, external shocks, and nonlinear relationships. This study aims to enhance inflation forecasting by integrating machine learning methods into the existing Selective-Combined Inflation Forecasting System (SSCIF). The inclusion of algorithms such as Ridge Regression, Lasso Regression, and Elastic Net enables the identification of complex patterns in macroeconomic data, thereby improving forecast accuracy.

A comparative analysis of forecasts generated using traditional econometric models (OLS, LTAR, BVAR, RW) and machine learning algorithms demonstrates that the hybrid approach significantly reduces forecasting errors and enhances the reliability of short-term forecasts. The results contribute to the advancement of macroeconomic forecasting tools and the development of more effective monetary policy, supporting better decision-making by central banks.

Keywords*: inflation, forecasting, consumer price index, model, machine learning, econometric models, forecast accuracy.*

JEL Classification: *E31, E37, C52, C61.*

[1] Zarina Adilkhanova – Chief Specialist-Analyst of the Macroeconomic Research and Forecasting Division, Monetary Policy Department, National Bank of the Republic of Kazakhstan.
Email: zarina.adilkhanova@nationalbank.kz

[2] Yerzhan Islam – Deputy Head of the Macroeconomic Research and Forecasting Division, Monetary Policy Department, National Bank of the Republic of Kazakhstan
E-mail: islam.yerzhan@nationalbank.kz

1 CONTENT

1. Introduction

Amid global economic instability and rising uncertainty, the ability to accurately forecast inflation has become a cornerstone of macroeconomic resilience. Reliable inflation forecasts are essential not only for effective monetary policy management but also for enhancing the predictability of economic dynamics. This is particularly relevant for developing countries like Kazakhstan, where precise inflation forecasts play a critical role in mitigating negative impacts on household purchasing power, investment activity, and fiscal decisions. Consequently, analyzing inflationary processes is important for economists and policymakers striving to maintain a stable and predictable economic environment that fosters sustainable growth.

Traditional inflation forecasting methods, such as Autoregressive (AR) models, Vector Autoregressive (VAR) models, and Error Correction Models (ECM), have demonstrated high effectiveness under stable macroeconomic conditions (Stock and Watson, 2003). However, as economic volatility and uncertainty intensify, these approaches face significant limitations. They may inadequately account for nonlinear relationships and fail to adapt swiftly to rapid economic changes, diminishing forecast accuracy in unstable conditions.

In response to these challenges, modern research has explored the integration of machine learning (ML) models. ML methods offer advantages in modeling complex economic processes and improve forecast accuracy by uncovering hidden patterns in macroeconomic data. Empirical studies indicate that ML algorithms can substantially reduce forecast errors compared to traditional econometric models, particularly under high volatility conditions (Medeiros et al., 2021).

Machine learning techniques have become a valuable complement to traditional econometric models, enhancing the ability to analyze hidden relationships within data. As Mullainathan and Spiess (2017) note, ML algorithms flexibly account for nonlinear dependencies and data variability, improving the adaptability of forecasts to changing economic conditions. Research by Kohlscheen (2021) and Medeiros et al. (2021) confirms that ML is effective in handling large datasets and identifying relationships inaccessible to traditional approaches, enabling a more comprehensive analysis of factors influencing inflation.

This study aims to enhance the accuracy of inflation forecasts by integrating ML methods into the existing System of Selective Combined Inflation Forecasting (SSCIF), used by the National Bank of Kazakhstan (NBK) for short-term inflation forecasting. Previously, SSCIF demonstrated strong short-term forecasting accuracy using classical econometric models. However, contemporary economic conditions demand greater flexibility and adaptability, justifying the adoption of ML methods such as Ridge, Lasso, and Elastic Net. These algorithms effectively capture complex relationships and hidden patterns that are difficult to identify with traditional approaches (Mullainathan and Spiess, 2017; Zou and Hastie, 2005).

The primary objective of this research is to improve inflation forecast accuracy by integrating ML methods with traditional econometric approaches. This hybridization within the SSCIF framework enhances adaptability to evolving economic conditions and improves the quality of short-term forecasts.

The study is organized into several sections. The second section reviews the literature, and the third elaborates on the methodological aspects of the research. The fourth section, focused on results, comprises two key components: the first updates coefficients and equations of traditional models to reflect current economic conditions, and the second integrates four ML algorithms adapted to economic uncertainty. The final section provides conclusions. The results demonstrate the high potential of ML methods for macroeconomic modeling, significantly improving inflation forecast accuracy in Kazakhstan.

2. Literature review

In the context of global economic uncertainty and accelerating structural transformations, forecasting has evolved from being a purely analytical task to a vital tool for economic policy. The precision of forecasts plays a critical role in maintaining macroeconomic stability and ensuring the effectiveness of policy decisions.

Inflation forecasting is particularly important for all economic agents, as it enables informed planning of investments, expenditures, and savings under conditions of uncertainty. This is especially true for households and businesses, which adjust their decisions based on expectations of future price levels. For central banks pursuing inflation-targeting policies, accurate inflation forecasts are foundational for determining interest rates and other monetary y policy measures. These forecasts help stabilize inflation expectations, increase trust in economic policy, and contribute to macroeconomic resilience and sustained growth.

Classical econometric models have long been central to inflation forecasting (Stock and Watson, 2003). These methods have proven highly effective in relatively stable economic environments due to their capacity to quantitatively assess both long-term and short-term trends. However, the contemporary economic landscape—characterized by burgeoning data volumes, heightened volatility, and global uncertainty—places new demands on the flexibility and adaptability of models. Traditional approaches increasingly encounter limitations in responding promptly to external shocks and shifts in inflation expectations, necessitating innovative solutions to enhance forecast accuracy.

In the face of these challenges, machine learning methods are gaining increasing importance in economic research. Due to their flexibility, ML techniques have opened new avenues for macroeconomic forecasting. Mullainathan and Spiess (2017) emphasize that ML approaches excel in uncovering complex nonlinear relationships, thereby significantly extending the capabilities of traditional econometric methods. Unlike classical models, ML algorithms adapt to intricate and rapidly changing data, improving forecast accuracy. Kohlscheen (2021) and Medeiros et al. (2021) further highlight that modern algorithms can process large datasets and identify dependencies that remain hidden to traditional methods, thereby offering deeper insights into the determinants of inflation.

Comparative analyses reveal several key advantages of ML methods over classical econometric approaches. First, ML models effectively manage large datasets with numerous variables, addressing dimensionality challenges—a critical factor in inflation analysis, which depends on a wide array of determinants (Kohlscheen, 2021). Second, these models prioritize optimizing forecast accuracy over parameter estimation, making them particularly effective for short-term forecasting, which is essential for timely responses to macroeconomic changes (Mullainathan and Spiess, 2017). Third, the ability of ML models to automatically update with incoming data enables them to adapt swiftly to changes in the economic environment, providing real-time, relevant forecasts.

In recent years, the integration of ML methods with traditional econometric models has become increasingly prevalent in inflation forecasting. These hybrid approaches combine the interpretability and historical trend analysis of econometric models with the capacity of ML algorithms to detect nonlinear relationships and adapt to rapidly evolving conditions. For example, Medeiros et al. (2021) demonstrate that combining Bayesian Vector Autoregressive (BVAR) models with ML methods such as Ridge and Elastic Net regressions significantly reduces forecast errors, particularly during periods of high volatility.

The strength of hybrid models lies in their ability to deliver more accurate forecasts by accounting for hidden dependencies that are challenging to detect using classical methods. Kohlscheen (2021) emphasizes that incorporating algorithms like Adaptive LASSO into econometric models enhances their adaptability to changing macroeconomic conditions, thereby improving forecast precision.

Further research confirms that applying ML methods and factor models substantially enhances inflation forecasting accuracy, especially when working with large datasets. Medeiros and Vasconcelos (2016) demonstrate that linear models employing the adaLASSO method improve forecast precision through adaptive variable selection. Garcia et al. (2017) illustrate that combining various approaches, such as shrinkage techniques and subset regressions, can be particularly effective in data-rich economies, as evidenced in the Brazilian context. Similarly, Maehashi and Shintani (2020) note that ML methods and factor models exhibit superior predictive performance for medium- and long-term forecasts, underscoring their versatility.

Despite their advantages, the application of ML methods presents certain methodological challenges. One of the primary concerns is overfitting, especially when dealing with high-dimensional datasets characteristic of macroeconomic forecasting. To mitigate this risk, penalized regression techniques such as Ridge, Lasso, and Elastic Net are widely used. These approaches play a crucial role in structuring and interpreting data effectively, facilitating variable selection and enhancing model robustness—particularly important in volatile and uncertain environments (Bai and Ng, 2008; Kim and Swanson, 2018; Smeekes and Wijler, 2018).

Thus, hybrid models that combine ML techniques with traditional econometric approaches have already established themselves as essential tools in the arsenal of macroeconomic forecasting. They contribute to the generation of more reliable forecasts and provide central banks with robust decision-making support.

Contemporary research underscores that the synergy between ML and econometric methods can significantly improve inflation forecasting accuracy, offering flexible and adaptive models. However, unresolved issues remain, including the interpretability of complex ML models and their applicability under structural economic changes. Future research should focus on optimizing ML methods for nonstationary data and enhancing the resilience of forecasts to abrupt economic shocks.

3. Description of data and methodology

3.1. Description of the selective inflation forecasting system and data

The Selective-Combined Inflation Forecasting System (SSCIF) represents a practical implementation of the selective combination forecasting approach developed and employed by the National Bank of Kazakhstan (NBK) for short-term inflation forecasting within the framework of the Forecasting and Policy Analysis System (FPAS). SSCIF enables the modeling of Consumer Price Index (CPI) dynamics using disaggregated data, thereby improving forecast accuracy. The system incorporates 66 subindices, which represent the main categories of goods and services included in the CPI (Table 1 in the Appendix).

Exogenous factors that partially explain the dynamics of these subindices include exchange rates of the Kazakhstani tenge against foreign currencies, inflation in Russia, industrial producer prices in Kazakhstan, money supply, real household incomes, and global prices for oil and wheat. This detailed approach allows for a more accurate assessment of inflation's sensitivity to changes in macroeconomic conditions.

All variables in SSCIF are represented as seasonally adjusted monthly indicators, and to ensure the stationarity of time series, they are transformed into log differences. The forecasting process uses data starting from January 2011. SSCIF combines four types of models, categorized as unconditional and conditional: unconditional models include Linear Trend Autoregression (LTAR) and Random Walk (RW), while conditional models include Ordinary Least Squares (OLS) regression and Bayesian Vector Autoregression (BVAR). Forecast accuracy is evaluated using the Root Mean Squared Error (RMSE), enabling models to be weighted according to their predictive performance.

The SSCIF model employs a selective approach, allowing for the evaluation of each model's accuracy across various time horizons and the calculation of combinatorial weights for individual time periods. These weights, referred to by Tuleuov (2017) as "recursive combinatorial weights," enhance forecast accuracy for shorter horizons and provide greater flexibility in adapting to changing conditions. Based on the forecast obtained, each subindex is weighted according to its share in

Kazakhstan's consumer basket. The final weighted forecast represents the overall CPI projection, reflecting the expected inflation dynamics for the given time period.

Thus, SSCIF serves as a flexible tool for inflation forecasting, combining models and accounting for different time intervals, which enables the generation of more reliable CPI forecasts.

In the updated version of the model, as in the previous one, the CPI remains the primary inflation indicator. Within the framework of using machine learning methods, the list of variables has been expanded. Unlike traditional models, machine learning models use disaggregated components of industrial producer prices across individual sectors and commodities instead of aggregated indicators. This increases the number of explanatory variables and allows for a more precise assessment of each sector's influence on inflation. The model will also incorporate global prices not only for wheat but for other important commodities included in the FAO Food Price Index, which will improve forecast accuracy.

The newly added variables, like the existing ones, are represented as monthly seasonally adjusted data, transformed into log differences to ensure the stationarity of time series at a 5% significance level.

3.2. Enhancements to existing econometric models

As part of this study, significant improvements were made to the SSCIF system to enhance forecasting accuracy through the use of econometric models. Detailed information on the econometric models used in SSCIF is provided in the research by Tuleuov (2017). The updates to the models include the following:

1. Improvement of equations and forecasts for exogenous variables

The equations for exogenous variables have been revised to improve the accuracy of their forecasts. This enhancement ensures more reliable modeling of external factors influencing inflation in Kazakhstan.

2. Increased accuracy in seasonal factor forecasting

The model now incorporates more precise equations for forecasting seasonal factors associated with exogenous variables. These updates have enhanced the forecasting of seasonal fluctuations, which play a crucial role in shaping inflation, particularly in sectors highly susceptible to seasonal variations.

3. Exogenous forecasts for inflation in russia and global oil prices

Forecasts for inflation in Russia and global oil prices are now exogenously specified in the model, based on projections from satellite models. This approach allows the integration of external economic conditions using more robust and up-to-date data.

4. Expansion and optimization of BVAR equations

The Bayesian Vector Autoregression (BVAR) model has been expanded by increasing the number of equations from 15 to 18, based on expert assessments and identified correlations among groups of goods and services within the CPI. Additionally, the number of lags in the model was updated using econometric tests, such as the Akaike Information Criterion (AIC) and residual autocorrelation tests, ensuring a more comprehensive incorporation of current input data.

These enhancements aim to improve the reliability of SSCIF forecasts, enabling more effective modeling of inflationary processes in Kazakhstan.

3.3. Application of machine learning models

This methodological section outlines the machine learning models employed for short-term inflation forecasting in Kazakhstan. The application of methods such as Ridge, Lasso, and Elastic Net is based on their proven ability to enhance forecast accuracy while addressing challenges such as multicollinearity and overfitting, which are common in macroeconomic data.

These approaches enable the simultaneous consideration of multiple economic variables, ensuring forecast stability and adapting the model to changing economic conditions. By leveraging the strengths of machine learning, these models provide a robust framework for capturing complex relationships and improving the precision of inflation forecasts.

3.3.1. Ridge Regression

Ridge Regression is a regularization technique widely used to address the problem of multicollinearity in multiple linear regression tasks. Traditional regression methods assume the independence of explanatory variables; however, in practice, these variables are often interrelated. This interdependence reduces the accuracy of estimates and makes the model unstable. Ridge Regression mitigates the impact of multicollinearity by incorporating regularization, which imposes a penalty on the magnitude of regression coefficients, thereby stabilizing the model.

Ridge Regression is built upon the standard multiple linear regression model:

$$Y = X\beta + \epsilon$$

where Y is the dependent variable (in this case, inflation), X is the matrix of explanatory variables (e.g., global commodity prices, exchange rates, money supply), β represents the model coefficients to be estimated, and ϵ is the random error term.

In standard linear regression, the goal is to minimize the Residual Sum of Squares (RSS) to determine the coefficients:

$$RSS = \sum_{i=1}^{n} (Y_i - X_i\beta)^2$$

Ridge regression adds a penalty to this function based on the magnitude of the coefficients to control their size and prevent overfitting. The function minimized by Ridge regression is as follows:

$$RSS + \lambda \sum_{j=1}^{p} \beta_j^2$$

Where RSS represents the residual sum of squares, and λ is the regularization parameter that controls the degree of "penalty" for large coefficient values.

This additional parameter, known as the ridge parameter, mitigates the effects of multicollinearity, making the model's estimates more robust and accurate. When λ is small, ridge regression approximates ordinary linear regression, where coefficients are estimated using the least squares method. However, as λ increases, the coefficients shrink, stabilizing their values under multicollinearity and reducing model variability. This is a key feature of Ridge Regression: it reduces the influence of variables that may introduce excessive variability into the model, thereby improving forecast accuracy, particularly when dealing with a large number of correlated variables.

Ridge Regression was introduced by Hoerl and Kennard (1970) and has since become a fundamental method for addressing multicollinearity. The ridge parameter λ determines the extent to which the model penalizes large coefficient values, balancing accuracy and simplicity. To select the optimal λ, cross-validation is often employed, testing the model on various data subsets to identify the value that maximizes forecast accuracy.

In the context of inflation forecasting, Ridge Regression is particularly useful for analyzing a wide range of macroeconomic variables, such as global oil prices, exchange rates, money supply, household incomes, and other indicators that are often correlated. Multicollinearity arising from these interdependent variables can significantly reduce forecast accuracy in traditional models. Ridge Regression stabilizes the model, delivering precise forecasts in the short- and medium-term horizons.

Thus, Ridge Regression is an effective tool for building inflation forecasting models, capable of reducing the impact of multicollinearity and overfitting. This method is especially relevant for macroeconomic data characterized by high complexity and interdependencies among variables.

3.3.2. Lasso Regression

Lasso Regression (Least Absolute Shrinkage and Selection Operator), introduced by Tibshirani (1996), is a linear regression method that combines regularization with variable selection. Unlike traditional methods, Lasso Regression not only estimates model coefficients but also automatically excludes insignificant variables, making it a valuable tool for building econometric models, particularly for inflation forecasting.

The core principle of Lasso lies in applying regularization using the L1-norm, which adds a penalty for the absolute values of the coefficients. This modifies the standard minimization function as follows:

$$RSS + \lambda \sum_{j=1}^{p} |\beta_j|$$

Where RSS represents the residual sum of squares, and λ is the regularization parameter.

Unlike Ridge Regression, which penalizes the square of the coefficients, Lasso applies a penalty to their absolute values, allowing some coefficients to shrink to zero. This means Lasso not only reduces the influence of individual variables but also eliminates those that contribute minimally to the forecast, leaving only significant factors.

The selection of the λ value is a critical step in Lasso, as it determines the number of variables retained in the model. If λ=0, the model reduces to ordinary linear regression. As λ increases, more variables are excluded from the model. The optimal λ value is determined through cross-validation to achieve a balance between accuracy and model complexity.

Lasso is a powerful tool for inflation forecasting as it helps identify key factors influencing inflationary processes. Its advantage lies in its ability to automatically exclude irrelevant variables, retaining only the most significant ones. This makes Lasso particularly useful when working with high-dimensional data where many variables may be correlated. Unlike Ridge Regression, which reduces all coefficients, Lasso can completely eliminate variables, simplifying model interpretation and reducing the risk of overfitting.

3.3.3. Elastic net

The Elastic Net method is employed in regression analysis to address challenges associated with a large number of explanatory variables, among which strong correlations may exist. Developed by Zou and Hastie (2005), Elastic Net combines elements of two regularization techniques - Lasso and Ridge - allowing it to simultaneously handle multicollinearity and select the most significant variables.

Mathematically, Elastic Net extends linear regression by incorporating two penalty functions. The primary objective of this method is to minimize the following function:

$$RSS + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2$$

In this formula, λ_1 represents L1-regularization (as in Lasso), which "zeroes out" irrelevant coefficients, while λ_2 accounts for L2-regularization (as in Ridge), which shrinks all coefficients and stabilizes the model in the presence of correlated variables.

This combination of L1 and L2 regularization makes Elastic Net particularly useful in scenarios where the data contains numerous interrelated variables. L1-regularization aids in variable selection by excluding those that have minimal influence on the forecast, while L2-regularization reduces the variance of coefficient estimates, ensuring the model's robustness against multicollinearity. Thus, Elastic Net effectively combines the strengths of Ridge and Lasso, delivering more reliable results when dealing with complex and high-dimensional datasets.

A critical step in Elastic Net is determining the optimal values for λ_1 and λ_2, which is typically done through cross-validation. This process helps achieve a balance between variable selection and coefficient stabilization, thereby improving forecast accuracy.

4. Results

4.1. Enhancing econometric models

The modifications made to the econometric SSCIF model have significantly enhanced its predictive performance, reducing forecast errors and improving the accuracy of key indicator estimates. These improvements have allowed for a more effective consideration of interrelationships among economic variables, thereby increasing the reliability of the results.

Table 1 presents a quantitative assessment of the accuracy of pseudo out-of-sample inflation forecasts for Kazakhstan using the original and updated versions of the SSCIF model[3]. Forecast accuracy was measured using the root mean squared error (RMSE), where lower values indicate higher forecast quality. According to the data provided, the updated model demonstrates improved forecast accuracy for most periods. However, during the period from March to November 2016, the previous model exhibited lower RMSE in forecasting non-food and service inflation. This is attributed to its initial design, which accounted for the characteristics of this time interval. Conversely, the updated model shows higher forecast errors for this period due to its adaptation to a broader time horizon and the inclusion of new data obtained in subsequent years.

For the period from January to September 2018, the updated model achieved a significant reduction in RMSE across all inflation categories. The most substantial improvement was observed in overall inflation, where the error decreased by 18.6%. These results underscore the updated model's ability to incorporate new economic trends that were not captured by the earlier version.

Between July 2020 and March 2021, the updated model also demonstrated significant improvements, particularly for overall inflation, where the forecast error decreased by 15.8%. This indicates that the updates enabled the model to account for more long-term relationships between economic variables. For the most recent period, from May 2023 to January 2024, the updated model reduced the RMSE for overall inflation by 15.3%. This highlights the model's effective adaptation to new macroeconomic conditions, including changes in price dynamics and exchange rate factors.

Thus, the previous model demonstrates better accuracy for earlier periods, such as March–November 2016, due to its initial focus on these time intervals.

[3] The forecasts presented in this study are purely model-based estimates derived from the internal algorithms of econometric models. They do not account for additional factors such as external assumptions, expert judgments, geopolitical developments, changes in regulated prices, or other administrative measures. Consequently, these results may differ from actual forecasts, which incorporate a broader range of external factors and expert assessments that influence the economic situation and price dynamics.

However, for later periods, the updated model significantly outperforms the previous one, as it incorporates a larger dataset and is tailored to address new economic conditions.

Table 1. Quantitative Assessment of Pseudo-Out-of-Sample Inflation Forecast Accuracy (CPI in % m/m) in Kazakhstan Across Different Test Periods for the Previous and Updated Model

Period		Food Inflation	Non-Food Inflation	Service Inflation	Overall Inflation
March 2016 - November 2016	Previous Model (RMSE)	0.519	0.210	0.281	**0.198**
	Updated Model (RMSE)	0.494	0.249	0.285	0.206
January 2018 - September 2018	Previous Model (RMSE)	0.331	0.627	0.277	0.242
	Updated Model (RMSE)	0.312	0.543	0.247	**0.197**
July 2020 - March 2021	Previous Model (RMSE)	0.124	0.323	0.293	0.076
	Updated Model (RMSE)	0.134	0.263	0.290	**0.064**
May 2023 - January 2024	Previous Model (RMSE)	0.484	0.347	0.184	0.203
	Updated Model (RMSE)	0.464	0.293	0.184	**0.172**

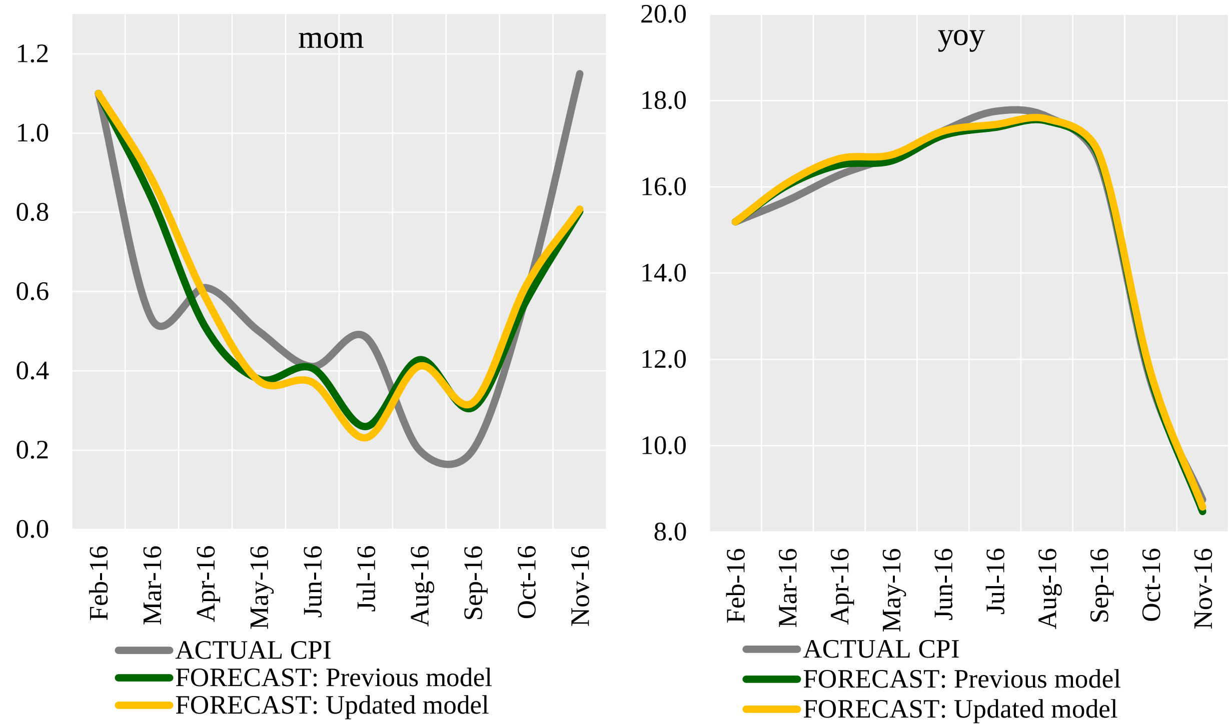
Source: BNS ASPIR RK, authors' calculations

Accuracy Color Scale *Less Accurate* *More Accurate*

Figures 1–4 illustrate a comparison of actual data and pseudo out-of-sample inflation forecasts for both versions of the SSCIF model. The updated model reflects monthly and annual changes with slightly greater accuracy, confirming its enhanced ability to account for economic dynamics and forecast CPI components.

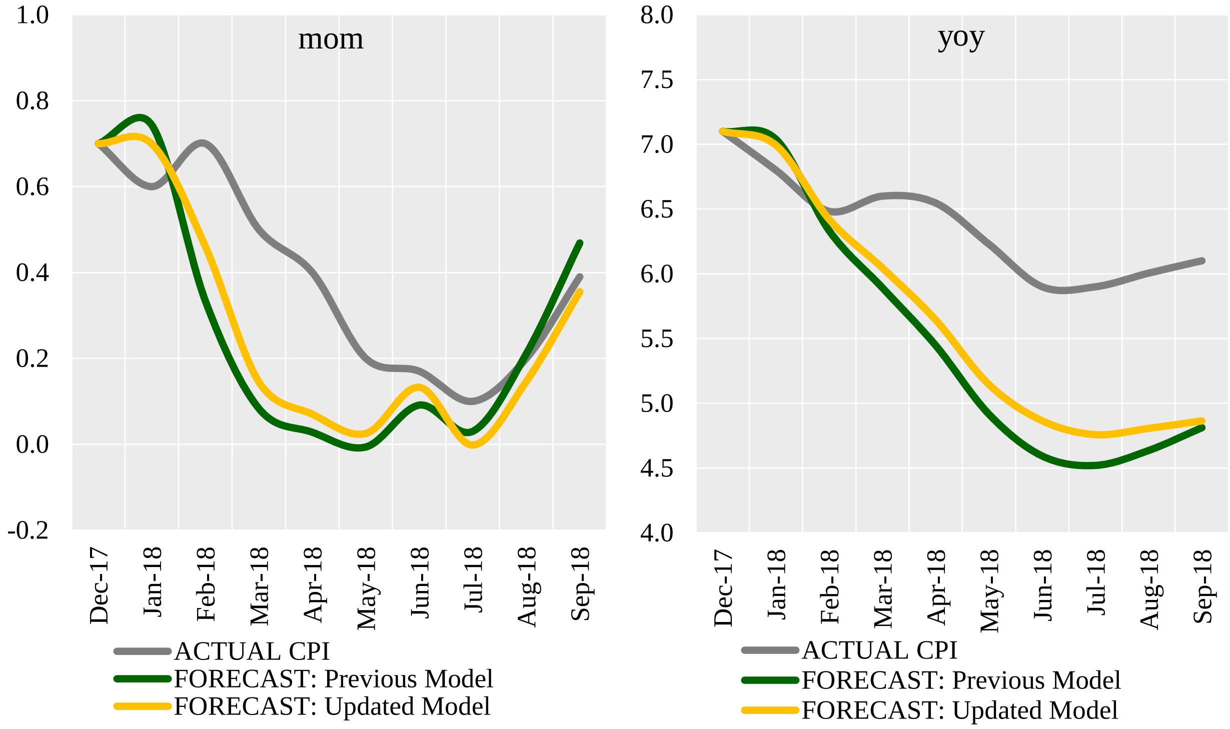
The updated version of the original SSCIF provides a higher level of inflation forecast accuracy, making it a more reliable tool for inflation forecasting.

Figure 1. Comparison of Actual Data and Real Out-of-Sample Inflation Forecast in Kazakhstan Using SSCIF from March to November 2016, %



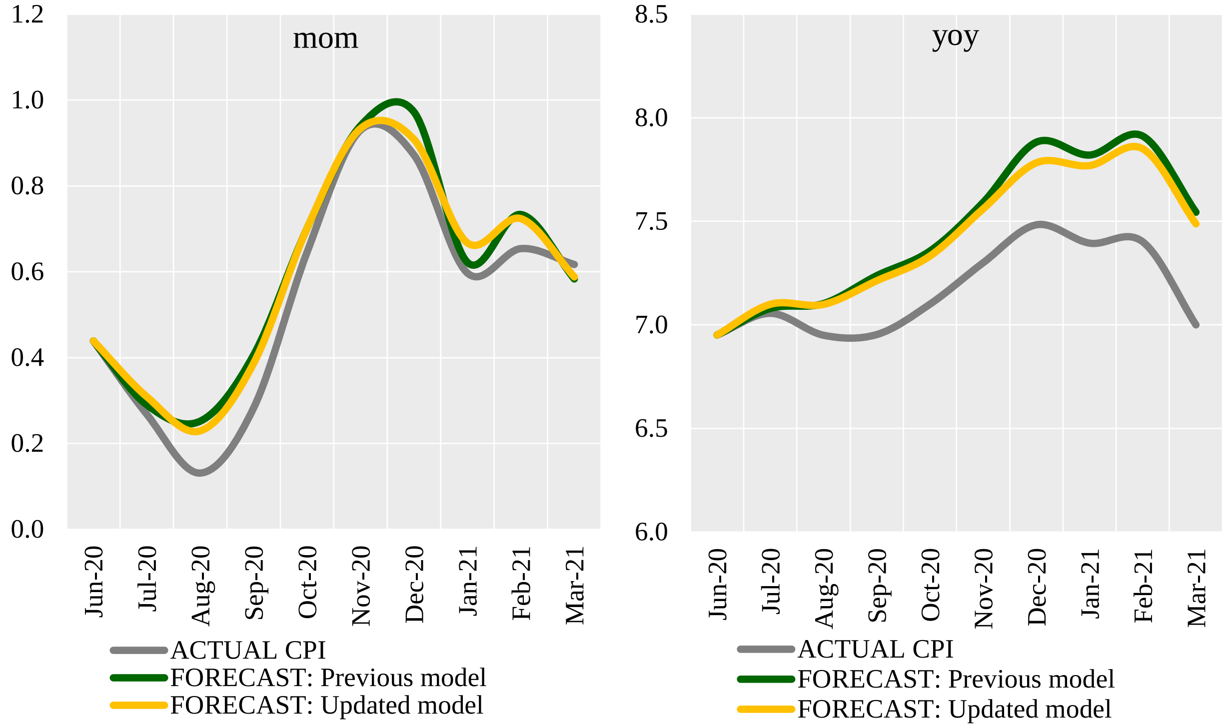
Source: BNS ASPIR RK, authors' calculations

Figure 2. Comparison of Actual Data and Real Out-of-Sample Inflation Forecast in Kazakhstan Using SSCIF from January to September 2018, %



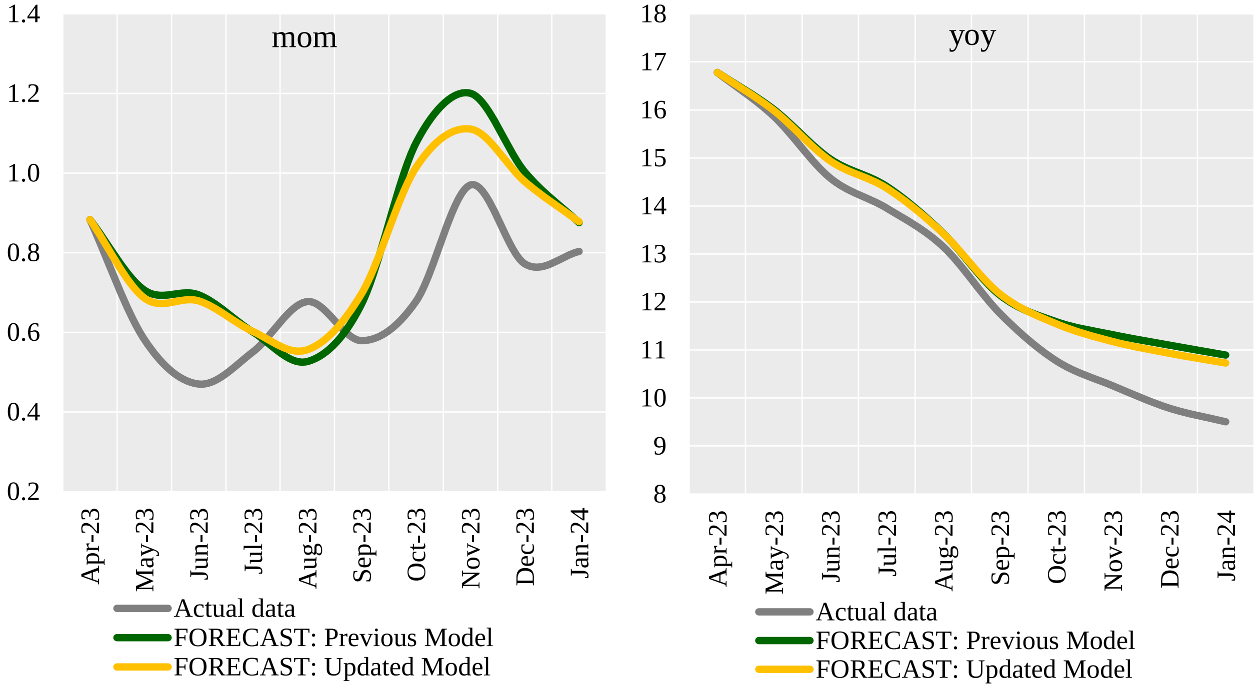
Source: BNS ASPIR RK, authors' calculations

Figure 3. Comparison of Actual Data and Real Out-of-Sample Inflation Forecast in Kazakhstan Using SSCIF from July 2020 to March 2021, %



Source: BNS ASPIR RK, authors' calculations

Figure 4. Comparison of Actual Data and Real Out-of-Sample Inflation Forecast in Kazakhstan Using SSCIF from May 2023 to January 2024, %



Source: BNS ASPIR RK, authors' calculations

4.2. Application of machine learning models

The results of the quantitative assessment of the accuracy of pseudo out-of-sample inflation forecasts, presented in Table 2, reveal differences in the effectiveness of traditional models and machine learning (ML) models across various time periods.

During the period from March to November 2016, the ML model was less effective in this interval, particularly in forecasting non-food and overall inflation. This can be attributed to its optimization for later periods and reliance on a broader dataset.

For the period from January to September 2018, the updated model demonstrated higher accuracy in forecasting food and service inflation, as evidenced by a lower root mean squared error (RMSE). However, the extended model employing ML methods outperformed the updated model in forecasting non-food and overall inflation, showcasing greater adaptability to changes in the economic environment. The RMSE for overall inflation decreased by 12.2% during this period.

Between July 2020 and March 2021, the ML model showed significant improvement in forecasting non-food and overall inflation, reducing RMSE by 32.3% and 6.3%, respectively, compared to the updated model. Nonetheless, for food and service inflation, the updated model maintained better results, demonstrating stability in forecasting these inflation components.

Table 2. Quantitative Assessment of Pseudo-Out-of-Sample Inflation Forecast Accuracy (CPI in % m/m) in Kazakhstan Across Different Test Periods for the Updated and Extended Models

Period		Food Inflation	Non-Food Inflation	Service Inflation	Overall Inflation
March 2016 - November 2016	Updated Model (RMSE))	0.494	0.249	0.285	**0.206**
	Extended Model with ML (RMSE)	0.771	0.662	0.336	0.473
January 2018 - September 2018	Updated Model (RMSE)	0.312	0.543	0.247	0.197
	Extended Model with ML (RMSE)	0.340	0.351	0.265	**0.173**
July 2020 - March 2021	Updated Model (RMSE)	0.134	0.263	0.290	0.064
	Extended Model with ML (RMSE)	0.221	0.178	0.258	**0.060**
May 2023 - January 2024	Updated Model (RMSE)	0.464	0.293	0.184	0.172
	Extended Model with ML (RMSE)	0.276	0.167	0.234	**0.131**

Source: BNS ASPIR RK, authors' calculations

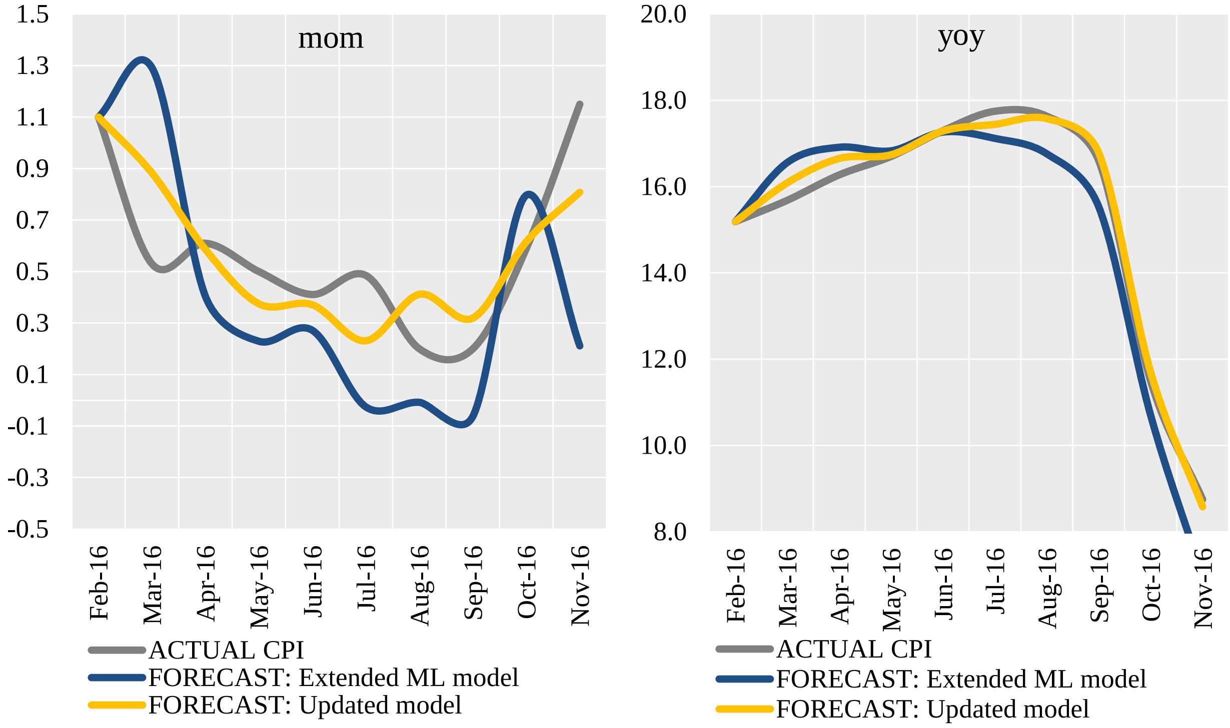
Accuracy Color Scale *Less Accurate* *More Accurate*

In the more recent period from May 2023 to January 2024, the ML model displayed a clear advantage in forecasting almost all inflation categories. The most notable improvement was in forecasting non-food inflation, underscoring the high accuracy and flexibility of the ML approach under modern economic challenges and volatility. However, in the service inflation segment, the updated model retained its leadership by delivering a lower forecast error. This confirms that traditional approaches remain relevant for predicting more stable and slow-moving inflation components. For overall inflation, the ML model achieved a 23.8% reduction in RMSE, illustrating its adaptability to new economic challenges.

The use of machine learning methods significantly enhances inflation forecasting accuracy (Figures 5–8). However, traditional approaches embedded in the updated model remain effective for certain segments, such as service inflation, and can complement ML models within a comprehensive forecasting framework.

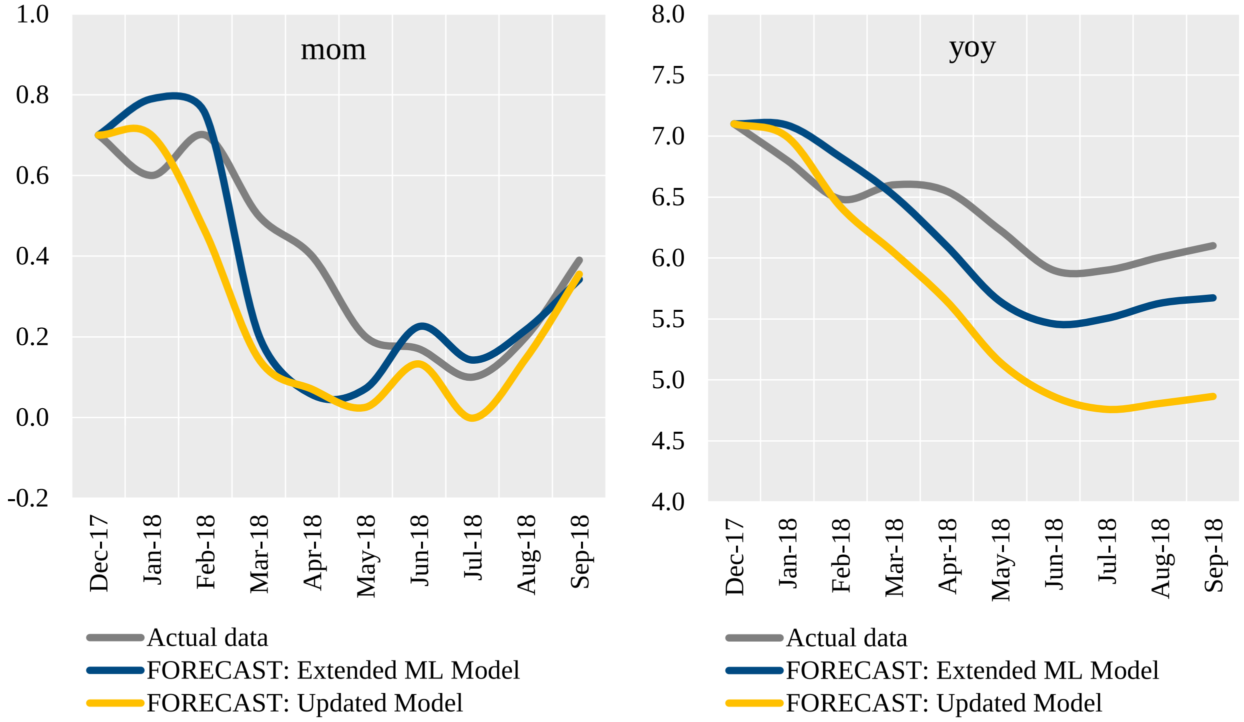
In conclusion, traditional models retain their advantages in forecasting stable inflation components, particularly during earlier time periods. At the same time, ML models demonstrate significant improvements in forecast accuracy in later periods, driven by their ability to account for a wide range of economic factors, adapt to changing conditions, and leverage larger datasets for training.

Figure 5. Comparison of Actual Data and Real Out-of-Sample Inflation Forecast in Kazakhstan Using SSCIF from March to November 2016, %



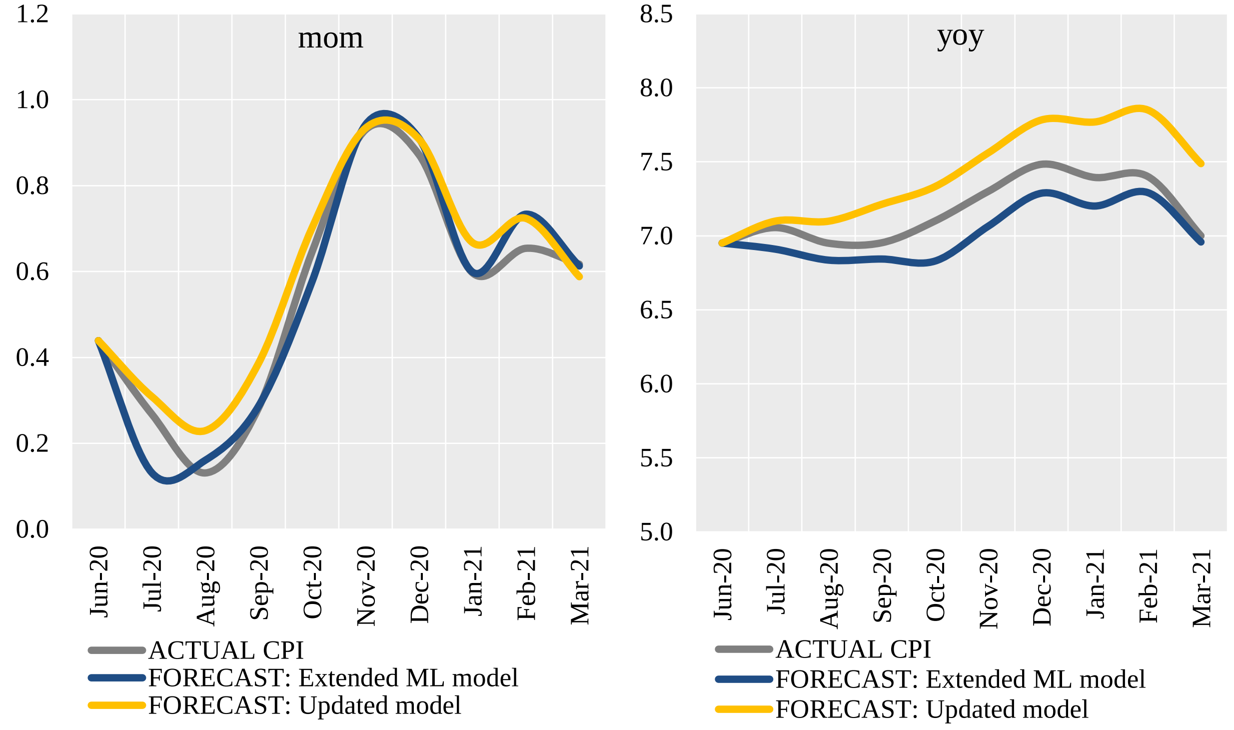
Source: BNS ASPIR RK, authors' calculations

Figure 6. Comparison of Actual Data and Real Out-of-Sample Inflation Forecast in Kazakhstan Using SSCIF from January to September 2018, %



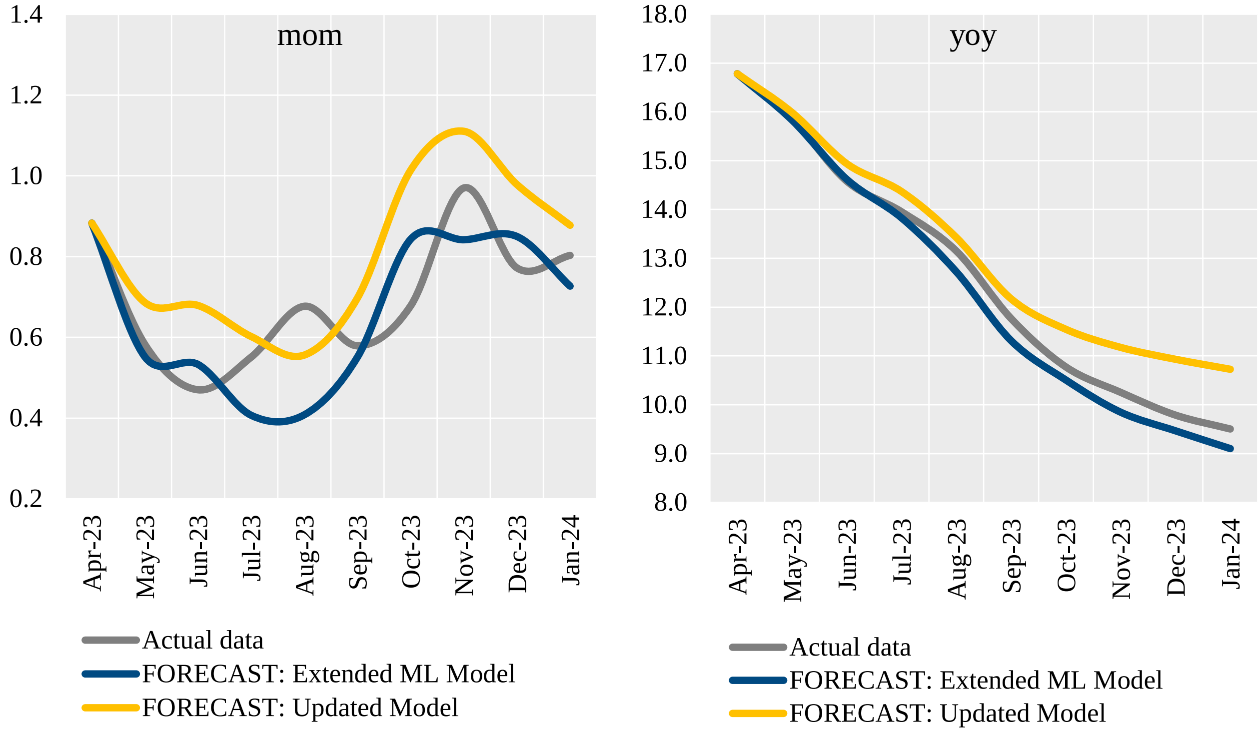
Source: BNS ASPIR RK, authors' calculations

Figure 7. Comparison of Actual Data and Real Out-of-Sample Inflation Forecast in Kazakhstan Using SSCIF from July 2020 to March 2021, %



Source: BNS ASPIR RK, authors' calculations

Figure 8. Comparison of Actual Data and Real Out-of-Sample Inflation Forecast in Kazakhstan Using SSCIF from May 2023 to January 2024, %



Source: BNS ASPIR RK, authors' calculations

5. Conclusion

This study has demonstrated the significant potential of integrating machine learning (ML) methods into the Selective-Combined Inflation Forecasting System (SSCIF) to improve the accuracy of short-term inflation forecasts in Kazakhstan. The modifications to the SSCIF model and the incorporation of ML algorithms have substantially enhanced its forecasting performance, reducing the Root Mean Squared Error (RMSE) and increasing the reliability of key indicator estimates by more effectively accounting for interrelations among economic variables.

The analysis showed that hybrid models combining traditional econometric approaches with modern ML algorithms significantly reduce RMSE, particularly under conditions of macroeconomic instability and high volatility. In particular, the integration of ML methods such as Ridge Regression, Lasso Regression, and Elastic Net into the SSCIF model demonstrated superiority over the updated econometric model in forecasting food, non-food, and overall inflation on later test intervals. These algorithms exhibit high adaptability and an ability to uncover hidden nonlinear relationships in macroeconomic data.

At the same time, traditional econometric approaches remain effective for forecasting service inflation, highlighting their value as tools for analyzing stable and slowly changing components of inflation.

The integration of ML algorithms has enabled the consideration of a broader range of macroeconomic factors, including disaggregated industrial prices and global commodity indices, significantly expanding the model's analytical capabilities. However, retaining elements of traditional methods, such as Bayesian

Vector Autoregression (BVAR), remains critical for ensuring the interpretability and robustness of models in data-constrained environments.

The results of this study have important practical implications for the development of macroeconomic forecasting tools and the enhancement of monetary policy effectiveness. Hybrid approaches can improve decision-making quality, which is particularly relevant for economies characterized by high uncertainty, such as Kazakhstan. The comprehensive use of ML methods and traditional econometric models, such as the updated SSCIF, offers the most balanced and effective forecasting tool, leveraging the advantages of both approaches: the adaptability and flexibility of ML models and the robustness of traditional methods.

The findings underscore the necessity for further development and integration of ML methods into economic analysis and forecasting. Future research should focus on optimizing the integration of ML methods into macroeconomic forecasting, emphasizing improved model interpretability, adaptation to nonstationary data, and minimization of overfitting risks. This will enhance the quality of economic forecasts and decision-making effectiveness, which is especially crucial in the context of global economic uncertainty and instability.

6. References

1. Bai, J., & Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics,* 146, 304–317.
2. Garcia, M.G.P., Medeiros, M.C., & Vasconcelos, G.F.R. (2017). Real-time inflation forecasting with high-dimensional models: The case of Brazil. *International Journal of Forecasting*, 33(3), 679-693.
3. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.
4. Kim, H. H., & Swanson, N. R. (2018). Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *International Journal of Forecasting*, 34(2), 339-354.
5. Kohlscheen, E. (2021). What does machine learning say about the drivers of inflation? BIS Working Papers 980.
6. Maehashi, K., & Shintani, M. (2020). Macroeconomic forecasting using factor models and machine learning: an application to Japan. *Journal of the Japanese and International Economies,* 58, 101104.
7. Medeiros, M. C., Vasconcelos, G. F., Veiga, A., & Zilberman, E. (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics*, 39(1), 98-119.
8. Medeiros, M. C., & Vasconcelos, G. F. R. (2016). Forecasting macroeconomic variables in data-rich environments. *Economics Letters,* 138, 50-52.
9. Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87-106.
10. Smeekes, S., & Wijler, E. (2018). Macroeconomic forecasting using penalized regression methods. *International Journal of Forecasting*, 34(3), 408-430.
11. Stock, J. H., & Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, 41(3), 788-829.
12. Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.
13. Tuleuov, O. (2017). System of selective-combined inflation forecast (SSCIF): Choosing the optimal technique for forecasting the dynamics of consumer prices in the context of a structural shock (case of Kazakhstan) //
    Тулеуов, О. (2017). Система селективно-комбинированного прогноза инфляции (SSCIF): выбор оптимальной техники прогнозирования динамики потребительских цен в условиях структурного шока (на примере Казахстана) (No. 2017-9).
14. Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2), 301-320.

7. Appendix

Table 1. SSCIF Input Data

Food products (13 price sub-indices)	Non-food products (23 price sub-indices)	Services (30 price sub-indices))	Exogenous variables
Bakery products and cereals	Materials for making clothes	Cleaning , repair and rental of clothes	The price of crude oil Brent
Meat	Outerwear	Shoe repair and rental	Nominal exchange rate USD/KZT
Fish and seafood	Other clothing items and clothing accessories	Housing rent	Nominal exchange rate RUB/KZT
Dairy products, cheese and eggs	Boots , shoes and other footwear	Residential maintenance and repair	Nominal exchange rate USD/RUB
Oils and fats	Materials for maintenance and repair of residential premises	Water supply	Food inflation in Russia
Fruits	Solid fuel	Waste collection	Non-food inflation in Russia
Vegetables	Furniture , household items, carpets and other floor coverings, their repair	Sewerage	Service inflation in Russia
Sugar, jam, honey, chocolate and confectionery	Textiles used in the household	Other services related to the maintenance of residential premises, not classified elsewhere	Production price index in Kazakhstan
Food products not classified in other categories	Appliances	Electricity	Manufacturing price index in Kazakhstan
Coffee, tea and cocoa	Glassware , cutlery and housewares	Gas	Money supply (in Tenge) in Kazakhstan
Mineral water, soft drinks, fruit and vegetable juices	Tools and devices used in everyday life and gardening	Heating energy	Real income in Kazakhstan
Alcoholic beverages	Goods and services used for housekeeping	Outpatient services	FAO Cereals Index
Tobacco products	Medicines , medical equipment and equipment	Hospital	
	Purchase of vehicles	Maintenance and repair of personal vehicles	
	Spare parts and accessories for personal vehicles	Other services related to personal vehicles	
	Fuels and lubricants for personal vehicles	Transport	
	Audiovisual and photographic equipment, information processing equipment	Communication services	
	Other large consumer durables for leisure and cultural events	Recreation, entertainment and culture	
	Other goods and equipment for recreation, sports, gardening and pets	Organization of comprehensive recreation	
	Newspapers , books and stationery	Preschool and primary education	
	Electrical appliances for personal use	Secondary education	
	Other items, devices and personal goods	Continued secondary education	
	items not classified elsewhere	Higher education	
		Education not divided into levels	
		Catering	
		Hotel services	
		Hairdressing and personal service establishments	
		Insurance	
		Financial services not classified elsewhere	
		Other services not classified in other categories	

Source: Compiled by Tuleuov (2017) and the authors based on data from the BNS ASPIR RK, National Bank of Kazakhstan, and Federal State Statistics Service of Russia.