NATIONAL BANK OF KAZAKHSTAN

Forecasting Kazakhstan's GDP Based on a Dynamic Factor Model with Regularization

Monetary Policy Department
Economic Study №2026-01

Alisher Akhmet

The economic research and analytical reports of the National Bank of the Republic of Kazakhstan (hereinafter – NBRK) are intended to disseminate the results of research conducted by the NBRK, as well as other research studies by its staff. These economic studies are published to foster discussions. The opinions and judgements presented in this article reflect the views of the authors and should not be construed as representing the views of the NBRK or its management.

Forecasting Kazakhstan's GDP Based on a Dynamic Factor Model with Regularization

February 2026

NBRK – WP – 2026 – 01

ISSN: 2789-150X

Forecasting Kazakhstan's GDP Based on a Dynamic Factor Model with Regularization

Alisher Akhmet[1]

Abstract

This paper develops an approach to short-term forecasting of Kazakhstan's gross domestic product (GDP) based on a dynamic factor model (DFM) estimated using a broad panel of macroeconomic and sectoral indicators. The model enables the extraction of latent factors that represent the main sources of common variation in the economy and their use for forecasting GDP under conditions of incomplete and asynchronous information. The factors are estimated within a state-space framework using the Kalman filter, which allows for a consistent treatment of missing observations and mixed data frequencies. To enhance forecast robustness and accommodate time variation in economic relationships, a regularized regression specification with exponential decay of observation weights is applied. Forecast performance is assessed using an expanding-window evaluation scheme that replicates real-time forecasting conditions and precludes the use of future information. A comparison with a naïve benchmark and an autoregressive model of GDP indicates that the regularized factor-based specification substantially reduces forecast uncertainty and yields more informative short-term forecasts.

Keywords: dynamic factor models, GDP forecasting, nowcasting, regularization

JEL-classification: C32, C38, C51, C53, E32, O47

[1] Chief Specialist-Analyst, Macroeconomic Research and Forecasting Division, Monetary Policy Department, National Bank of the Republic of Kazakhstan. alisher.akhmet@nationalbank.kz

Table of Contents

Introduction

Effective implementation of monetary policy requires timely assessment of current economic conditions and an understanding of short-term movements in key macroeconomic indicators. It is precisely short-term dynamics that form the initial informational basis for medium-term forecasting and for the formulation of scenario-based policy decisions under uncertainty. Real-time assessment and short-term forecasting of GDP constitute one of the most challenging tasks in applied macroeconomics. This difficulty arises because official statistics are released with substantial delays, are subject to revisions, and because short-run output dynamics are driven by a combination of high-frequency volatility, structural shifts, and irregular shocks of both domestic and external origin. As a result, short-term GDP forecasting requires methods capable of operating under incomplete information, unstable relationships, and elevated uncertainty.

Univariate models rely exclusively on information contained in the past dynamics of GDP itself and ignore changes in the structure of economic processes taking place in the real sector, the financial system, fiscal policy, and the external environment. During periods of stable growth such models may exhibit acceptable performance; however, under conditions of business cycle turning points, sharp external shocks, or shifts in internal growth drivers, their forecasting ability deteriorates sharply.

Attempts to improve short-term forecasting performance by incorporating a broad set of leading indicators encounter another fundamental problem in the form of the high dimensionality of the data panel. Modern macroeconomic statistics provide dozens or even hundreds of potentially relevant indicators; however, directly including a large number of variables in regression models quickly leads to overfitting, parameter instability, and a loss of forecasting robustness, especially in the presence of relatively short time series. At the same time, the manual selection of a small subset of indicators is inherently subjective and creates a risk of omitting important information. An additional complication arises from the high degree of correlation among macroeconomic indicators, reflecting the presence of common latent factors driving the dynamics of most series. Consequently, there is a need for a method that can exploit a rich information set while avoiding model overload and parameter instability.

The central idea of the factor model is that the joint dynamics of a large number of macroeconomic indicators can be represented by a small number of unobserved common factors. This representation makes it possible to incorporate a broad set of indicators while substantially reducing dimensionality, thereby mitigating multicollinearity and improving the stability of parameter estimates. An important additional advantage is the ability to work with data of different frequencies and with incomplete samples, which makes factor models particularly well suited for short-term forecasting tasks. Nevertheless, the practical application of factor models is subject to several important limitations. In particular, estimates are sensitive to the “ragged edge” problem, whereby in the current period only a subset of intra-period observations is available, leading to increased uncertainty and reduced accuracy. Moreover, the relationship between the factors and GDP is not structurally stable and may change substantially over time under the influence of crises, shifts in policy regimes, and external shocks. Taken together, this implies that even with correctly estimated factors, the classical specification does not always provide sufficient flexibility and forecasting robustness. In addition, an important limitation concerns the economic interpretability of the factors themselves. Because each factor is a linear combination of a large number of indicators that may move in different directions, its economic meaning can become blurred, especially when different groups of variables exhibit opposing dynamics.

In the presence of unstable and time-varying relationships between the factors and GDP dynamics, methods of regularization and adaptive estimation become particularly relevant, as they belong to a class of regularized and weighted forecasting models that reduce the sensitivity of estimates to noise and allow the model to adapt to changes in economic relationships over time. Ridge regression stabilizes coefficient estimates by penalizing their magnitude, which is especially important in the presence of correlated regressors, lagged factors, and limited sample size. Exponential decay assigns greater weight to more recent observations and reduces the influence of more distant data, enabling the model to adjust more rapidly to changes in economic regimes. The joint use of these two approaches yields a flexible and robust forecasting framework that is oriented less toward an exact reproduction of the entire historical trajectory and more toward improving short-term forecast accuracy (up to three quarters ahead) under conditions of elevated uncertainty.

The objective of this paper is to develop and empirically evaluate a short-term GDP forecasting model based on a dynamic factor structure with the use of regularization and an adaptive mechanism for weighting observations. The predictive performance of the models is assessed using standard accuracy metrics, and additional analysis of residuals and parameter stability is conducted. The contribution of the paper lies in demonstrating the practical effectiveness of combining a dynamic factor model, regularization, and exponential decay of observations for short-term GDP forecasting. In contrast to standard factor specifications, the emphasis is placed on improving forecast robustness in environments where the relationships among macroeconomic variables evolve over time, as is typical in economies experiencing significant structural transformation.

Literature review

The application of dynamic factor models to GDP forecasting in the context of the Kazakhstan's economy is examined in Orlov (2019). In that study, a dynamic factor model in state-space form is used to estimate and short-term forecast quarterly GDP on the basis of a broad panel of monthly macroeconomic indicators. The factors are estimated using a combination of principal components and the Kalman filter, which makes it possible to handle mixed data frequencies and the problem of missing observations in the current period. It should be noted that the empirical analysis is conducted primarily in terms of year-on-year growth rates, which smooth short-term fluctuations in the variables and reduce the sensitivity of the estimates to intra-year volatility. In this setting, the resulting forecasts are therefore mainly oriented toward aggregate output dynamics, while their use in applied settings that require quarter-on-quarter growth rates (for example, in structural models) entails additional transformations and assumptions. Overall, the results confirm that the factor-based approach provides stable and economically interpretable estimates of output dynamics and can serve as an effective tool for aggregating heterogeneous macroeconomic information for short-term forecasting purposes.

The theoretical and applied foundations of dynamic factor models in macroeconomic forecasting were established in the work of Stock and Watson (2002) on the concept of diffusion indexes. The authors showed that a small number of latent factors extracted from a large panel of monthly indicators is able to capture most of the information relevant for forecasting the U.S. economy. Rather than selecting a narrow set of leading indicators, they proposed estimating a few common factors using principal component analysis (PCA) and then employing these factors to forecast key aggregates such as industrial production or inflation. Their analysis demonstrated that forecasts based on only one or two factors outperform traditional benchmark models, ranging from univariate autoregressions to small VARs and indicator systems, particularly at longer horizons. An important finding of their work is that factors obtained via principal components remain consistent even in the presence of data heterogeneity, missing observations, and moderate cross-sectional dependence (multicollinearity), making this approach a practical tool for aggregating high-dimensional economic information for forecasting purposes.

Subsequent research sought to exploit the time-series dynamics of the factors and to handle missing observations more effectively in real time. Doz, Giannone, and Reichlin (2006) proposed a quasi-maximum likelihood approach that formulates the factor model in state-space form, making it suitable for estimation with the Kalman filter. Starting from factor estimates obtained by principal components, they apply an expectation–maximization procedure to refine the factors through likelihood maximization. This two-step procedure remains consistent and asymptotically normal as both the cross-sectional and time dimensions grow, while delivering efficiency gains relative to the simple principal components estimator. A key advantage of the parametric state-space representation is its ability to accommodate missing observations and asynchronous data releases, which makes the model directly applicable to real-time forecasting.

These methodological advances were brought together in Giannone, Reichlin, and Small (2006), who showed how a large dynamic factor model can be used to produce real-time estimates of current-quarter GDP and inflation as data is released within the quarter. Their U.S. application demonstrates that survey data can provide informative early signals about current quarterly GDP and inflation growth, often outperforming traditional statistical indicators. The paper illustrates that factor models make it possible to formalize the process of sequential forecast updating, which is used in practice by central banks, and it serves as a benchmark for assessing the marginal informational content of individual data releases in real time.

Mariano and Murasawa (2003) developed a dynamic factor model that explicitly accommodates variables observed at different frequencies, allowing monthly data to be used consistently for the estimation of quarterly GDP. The model assumes the existence of a single latent monthly factor of economic activity that jointly drives the dynamics of a panel of monthly indicators, such as industrial production and employment, and generates quarterly real GDP as the three-month aggregate of the same factor. By formulating the model in state-space form and estimating it using the Kalman filter and maximum likelihood, the authors obtain a smoothed monthly estimate of aggregate economic activity that is consistent with published quarterly national accounts. The mixed-frequency structure enables the model to exploit the timeliness of monthly statistical releases without loss of information and to track the business cycle in real time.

Against this background, changes in economic relationships over time require the use of methods that can adapt to structural shifts and reweight the importance of past observations. West and Harrison (1997) provided a foundational treatment of exponential decay of observations in time series forecasting. In their Bayesian dynamic linear model, past observations receive geometrically declining weights, implying that parameter estimates rely more heavily on recent data and adjust more rapidly to structural change. Although the method was developed in a Bayesian framework, the principle of exponential decay is more general and is widely applied in forecasting as an effective way to improve predictive accuracy when relationships evolve gradually over time.

At the same time, De Mol, Giannone, and Reichlin (2008) consider Bayesian shrinkage as a practical alternative to principal components in forecasting with large macroeconomic datasets. The authors note that while factor methods are primarily designed to reduce dimensionality, Bayesian shrinkage addresses the forecasting problem directly by stabilizing coefficient estimates in the presence of a large number of predictors. It is worth noting that the form of Bayesian coefficient shrinkage they employ has a direct interpretation as ridge regression. Their empirical results show that forecasts based on Bayesian shrinkage often perform at least as well as, and sometimes better than, those obtained using factor models, highlighting its value as a robust and effective tool for high-dimensional macroeconomic data. In addition, Hastie, Tibshirani, and Friedman (2009) provide a standard statistical foundation for ridge regression within the framework of supervised learning methods. They show that introducing a penalty on the magnitude of the coefficients reduces the variance of the estimates and improves forecasting accuracy through a controlled trade-off between bias and variance. In this sense, ridge regression is a useful tool for stabilizing estimates and is widely applied in forecasting tasks, including settings with a limited number of observations and explanatory variables characterized by a high stochastic component.

Data and Methodology

A broad panel of monthly and quarterly indicators capturing key aspects of economic activity in the country is used for the modeling. The sample includes measures of real economic activity, price dynamics, monetary conditions, fiscal policy, corporate financial reporting, and the external environment, covering the period from January 2010 to July 2025. The inclusion of multiple blocks of indicators provides a comprehensive representation of the interaction between demand and supply, as well as between domestic and external drivers of business activity (the full list of variables is reported in the Appendix). It should be noted that some of the original indicators were excluded during data preparation and preliminary modeling due to low variability or a negligible contribution to factor dynamics.

The time series were transformed into a comparable form using standard preprocessing procedures. Seasonal adjustment was performed using the X-13 ARIMA-SEATS method, logarithmic transformation and differencing were applied to achieve stationarity, and standardization[2] was used to eliminate differences in scale. For selected indicators exhibiting strong autocorrelation (construction, financial and insurance activity, and business credit issuance), a preliminary procedure was applied to remove the autocorrelated component, and the resulting residuals were used in the analysis. The final data panel consists of 25 monthly and 16 quarterly variables, with a total of 184 observations.

In a broad panel of monthly macroeconomic indicators $x_{i,t}$, $i = 1, \dots, N$ indexes the variables and $t = 1, \dots, T$ denotes time. The central assumption of the dynamic factor model is that the joint movement of these series can be captured by a small number of unobserved factors $f_t \in R^r$. Formally, the measurement equation takes the form:

$$x_t = \Lambda f_t + \varepsilon_t,$$

where x_t is the $N \times 1$ vector of observed variables, Λ is the $N \times r$ matrix of factor loadings, and ε_t is the vector of idiosyncratic errors. In principle, the errors may be weakly correlated; however, in this paper a diagonal covariance structure is assumed.

The dynamics of the latent factors f_t can in general be described by a vector autoregression (VAR), allowing for interactions among the factors and their lags. In practice, however, such flexibility is often accompanied by a substantial increase in the

[2] Standardization refers to transforming variables to a common scale by subtracting their mean and dividing by their standard deviation, yielding variables with zero mean and unit variance.

number of parameters and by instability of the estimates, especially when working with relatively short samples. To avoid overparameterization, the factor dynamics in this paper are restricted to a set of independent first-order autoregressive processes. Formally, for each factor $j = 1, \dots, r$:

$$f_{j,t} = \rho_j f_{j,t-1} + \eta_{j,t}, \;\; \eta_{j,t} \sim N(0, \sigma_j^2),$$

where ρ_j characterizes the persistence of the factor and $\eta_{j,t}$ denotes the innovation with variance σ_j^2. This specification preserves persistence in the common component while keeping the transition equation sufficiently simple.

In addition to the dynamics of the common factors, the idiosyncratic components $\varepsilon_{i,t}$ associated with each observed series are also modeled as independent AR(1) processes:

$$\varepsilon_{i,t} = \phi_i \varepsilon_{i,t-1} + \xi_{i,t}, \;\; \xi_{i,t} \sim N(0, \psi_i^2),$$

where ϕ_i captures series-specific persistence and $\xi_{i,t}$ denotes white noise. By imposing a diagonal covariance structure on the idiosyncratic errors, all systematic co-movement in the data is attributed to the common factors rather than to correlated measurement noise.

The combination of AR(1) dynamics for the factors and the idiosyncratic components yields a state-space system that is computationally convenient and stable in settings with limited sample sizes. Estimation is carried out by maximum likelihood using the Kalman filter, which allows both the model parameters and the trajectories of the latent factors to be inferred from the observed data panel.

In the actual implementation[3] of the model, a standard gradient-based optimizer (a numerical optimization method that relies on gradient information from the likelihood function) did not deliver stable convergence, which is common in problems with a large number of parameters and potentially flat regions of the likelihood surface. The objective of the optimization procedure in this context is to find the set of model parameters that maximizes the probability of the observed data. To refine the initial estimates, a derivative-free Powell method was employed, which does not require the computation of gradients and proceeds by iteratively refining the parameters based on previous estimates.

After obtaining the factor time series, the forecasting equation for quarterly GDP is estimated using weighted least squares with exponentially decaying weights. The objective is to reduce the influence of older observations that reflect economic patterns which may no longer be representative of current macroeconomic dynamics. Each time point t is assigned a weight

$$w_t = \gamma^{T-t}, \;\; 0 < \gamma < 1,$$

where T denotes the last available observation in the sample. Under this scheme, the most recent observations retain their full influence, while the contribution of earlier

[3] The optimization procedures were implemented using standard numerical algorithms in the Python environment, relying on implementations provided by the *statsmodels* library and the numerical optimization methods of the *SciPy* package.

data declines exponentially. An intuitive characterization of the sensitivity of the weighting function is given by its half-life, defined as the number of periods after which the weight falls by one half:

$$h = \frac{\ln(0.5)}{\ln(\gamma)}.$$

In this study, a half-life of four quarters is used, allowing the forecasting equation to respond more flexibly to current changes in the macroeconomic environment. Exponential decay of observations is applied directly at the stage of estimating the GDP-factor regression through the weighting of observations in the objective function:

$$y = \alpha + \beta' f_t + u_t,$$

where the contribution of each residual u_t is scaled by the weight w_t. This approach allows for gradual structural change and ensures a controlled influence of recent observations while still incorporating information from the full historical sample.

Ridge regression is employed as a regularization device to stabilize the forecasting equation. This approach introduces a penalty on the magnitude of the coefficients, which stabilizes the estimates in the presence of strong multicollinearity or an expanded set of regressors arising from the inclusion of additional variables. The model minimizes a modified quadratic error function:

$$\beta, \alpha = \arg\min_{\alpha,\beta} \left[\frac{1}{T} \sum_{t=1}^{T} (y_t - \alpha - \beta' f_t)^2 + \lambda_R \beta' \beta \right],$$

where $\lambda_R \geq 0$ denotes the regularization parameter. When $\lambda_R = 0$, the model reduces to ordinary least squares, whereas for positive values of the penalty the coefficient estimates become smoother and less sensitive to random fluctuations in the data. This is particularly important in applied macroeconomic panels, where time series often exhibit substantial mutual correlation and the inclusion of additional variables can lead to unstable or excessively variable coefficients. Ridge regression therefore makes it possible to extend the forecasting specification by including additional factors, lags, or forward-looking variables, while preserving interpretability and stability even when the effective information content of the sample is relatively limited.

The two regularization methods address distinct sources of instability in the forecasting equation. Their joint use allows the model to accommodate both time variation in economic relationships and redundancy or collinearity among predictors.

Results and Discussion

In selecting the specification of the dynamic factor model (DFM), information criteria did not meaningfully differentiate among candidate models. Consequently, the final choice of parameters was guided by in-sample fit and forecasting performance over the evaluation sample. In total, 24 specifications were tested, combining different values for the number of factors, the order of factor dynamics, and the order of the idiosyncratic components. The results indicate that the most suitable specification is a model with four latent factors, first-order factor dynamics, and a first-order autoregressive structure for the idiosyncratic errors. The inclusion of additional factors increases the explanatory content of the factor structure by capturing incremental shares of the common variation across the panel of variables. In the chosen specification, the common factor component explains approximately 26.1% of the total variance of the observed variables, implying that more than one quarter of the fluctuations in the broad data panel are systematic and can be represented by a limited number of latent processes. The remaining variation reflects idiosyncratic dynamics of individual series, sector-specific effects, and irregular shocks that do not generate stable co-movement. Although increasing the number of factors further raises the share of explained variance, this comes at the cost of reduced economic interpretability of the factors, higher model complexity, and weaker forecasting robustness.

Interpretation of the Factors and Decomposition of Real GDP Growth.

The interpretation of the extracted factors is based on an analysis of the factor loadings (Table 1), which indicate the strength of the association between each variable and the corresponding latent component. Since dynamic factor models extract common sources of variation from high-dimensional data panels, the loading structure provides a way to form an approximate understanding of the economic content of each factor. It should be kept in mind that the factors primarily reflect statistical patterns of co-movement across many indicators and are not designed to achieve a strict separation of individual sectors of the economy. The method captures common fluctuations that account for a substantial share of the latent variance in the data. In this context, the contribution of the factors to GDP should not be interpreted as a literal decomposition of output, but rather as an assessment of how common components influence deviations of actual GDP from its average historical growth rate.

The first factor reflects a broad procyclical component of economic dynamics. Its structure is shaped by loadings distributed across all major groups of indicators, with the largest contribution coming from real sector variables and related measures of the business environment. Indicators capturing funding sources and liquidity conditions, as well as fiscal policy variables, also play a significant role, reinforcing the interpretation of this factor as an index of underlying economic activity.

The second factor characterizes the dynamics of funding sources and liquidity and is closely linked to corporate financial flows. The largest loadings are associated with monetary aggregates, credit, and deposit indicators, which capture the volume of financial resources available to firms and households. Corporate financial variables also rank among the key components of this factor, since firms' financial positions and their ability to raise and deploy resources are tightly connected to overall financing conditions.

The third and fourth factors reflect a common external component associated with the influence of international conditions on the Kazakhstan's economy. These factors capture elements of the external environment that generate key impulses for domestic economic dynamics, including changes in export revenues, exchange rate movements, and price shocks transmitted through income, imported inflation, and terms of trade. The separation of the external factor into two distinct components arises because the block of external variables exhibits a high degree of common variation and constitutes one of the strongest sources of co-movement in the panel. As the number of factors increases, the model identifies two independent subcomponents within the external block. In a more aggregated specification, these components would be combined into a single external factor; however, under the current structure they remain separate due to their partial orthogonality. The fourth factor has a more pronounced transmission character and is linked to the propagation of external shocks into the domestic economy through trade, transport, and household financial behavior, making its contribution to deviations in GDP dynamics more noticeable. The third factor plays a more auxiliary role, reflecting the general external background, whose influence is weaker and more indirect.

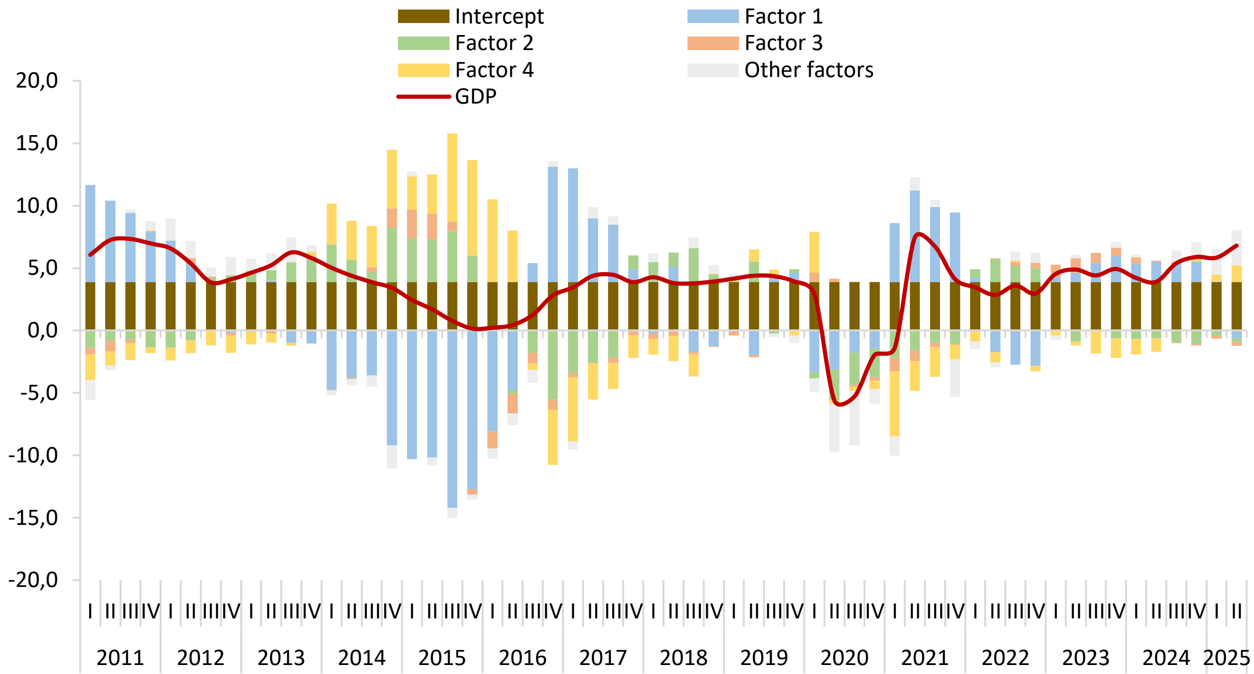
Table 1. Structure of Latent Factor Loadings, Weights in the Normalized Factor Space

	Factor 1	Factor 2	Factor 3	Factor 4
Real sector	0.193161	0.137172	0.051180	0.104307
Price dynamics	0.080784	0.110968	0.130264	0.087360
External conditions	0.132057	0.088652	0.379565	0.321411
Funding and liquidity	0.225112	0.498673	0.126714	0.139813
Corporate finance	0.125378	0.150806	0.127445	0.111759
Fiscal policy	0.222670	0.067194	0.123863	0.055936

The model constructs latent factors without using information on GDP, and the resulting factors reflect structural patterns of co-movement in the economy rather than indicators designed specifically to track output. Their impact on GDP is assessed using a regression model in which the coefficients measure the contribution of the factors to deviations of actual GDP dynamics from its average historical level.

The intercept in the model captures the average rate of GDP change, corresponding to a persistent component of output that remains even when the factors take zero values. In analytical terms, it can be interpreted as a background component associated with long-run trends in productivity and the expansion of physical output, which do not appear as common fluctuations in the panel data and are therefore not extracted by the factor model. This is natural for a developing economy, where demographic dynamics, capital accumulation, and gradual infrastructure expansion generate a baseline rate of growth. In this sense, the constant term can be viewed as reflecting a catching-up effect, whereby sustained growth is driven by long-term structural processes that do not necessarily manifest themselves as synchronous movements in macroeconomic indicators. At the same time, the constant should not be interpreted as a direct measure of potential growth in the economic sense. It reflects the average historical rate of GDP change implied by the linear specification and does not explicitly account for structural determinants of long-run growth such as productivity, demographics, capital accumulation, and technological change.

Figure 1. Decomposition of GDP Growth Rates, OLS, % y/y



For reference: Factor 1 – real sector and fiscal policy; Factor 2 – monetary aggregates, credit and deposits; Factors 3 and 4 – exchange rates and global prices; Other factors – idiosyncratic noise.

In the period up to mid-2013, the main source of positive deviations of GDP from its average historical growth rate was the factor capturing domestic business activity and real sector dynamics (Figure 1). This corresponds to the phase of post-crisis economic recovery following the global financial crisis, when growth was driven primarily by the expansion of production, trade, and related activities. At the same time, the factor associated with funding and liquidity exerted a restraining effect on output dynamics, reflecting an incomplete recovery of financial conditions and a weakened transmission of financial flows into economic growth after the crisis. In addition, external conditions during this period remained uncertain, limiting the emergence of a stable procyclical external impulse.

The period from late 2013 to mid-2016 is characterized by a deterioration in domestic economic conditions and a slowdown in business activity against the backdrop of a combination of external and internal shocks. The decline in global oil prices in 2014 and the weakening of economic conditions in Kazakhstan's major trading partner, Russia , heightened uncertainty and constrained real sector dynamics, which is reflected in the negative contribution of the first factor and indicates a weakening of the procyclical impulse of domestic economic activity relative to its average historical growth rate. At the same time, the policy of maintaining the exchange rate and the use of foreign exchange reserves acted as a nominal buffer, smoothing the sharp deterioration in external conditions and limiting the volatility of domestic prices and expectations. The drawdown of reserves mitigated the scale of the correction in domestic activity, although it was accompanied by increased import demand and an outflow of foreign currency liquidity. Institutional and regime changes also played an important role during this period, in particular the transition to an inflation targeting regime and a floating exchange rate. The depreciation of the previously overvalued real exchange rate and the stabilization of expectations constrained excessive imports and reduced internal imbalances, thereby exerting a stabilizing influence on economic dynamics. The funding and liquidity factor also made a positive contribution, not by generating an independent source of growth but by attenuating the depth of the deviation of GDP from its average historical growth rate. Taken together, during this period the economy was operating in a regime of adaptation and shock absorption rather than sustained procyclical growth.

During 2017-2019, GDP dynamics evolved close to its average historical growth rate. The absence of persistently positive or negative impulses from individual components indicates a phase of macroeconomic stabilization in which growth was broadly balanced. The contributions of individual factors displayed offsetting movements over time, but their aggregate effect largely cancelled out, pointing to the absence of dominant sources of overheating or contraction. Economic growth during this period therefore reflected normalization and stable functioning of the economy following earlier structural shocks.

In 2020, under the conditions of the coronavirus pandemic and the introduction of quarantine restrictions, the decline in domestic consumption, reduced mobility, disruptions in logistics chains, and the contraction of external trade were reflected in negative contributions from all factors. In 2022, the negative contribution of the first factor reflects increased uncertainty and a weakening of procyclical real sector dynamics in the context of the disruption of trade links and heightened geopolitical tensions. At the same time, the positive contribution of the second factor is associated with the dynamics of funding and liquidity indicators, which proved to be more pronounced and synchronized than on average over the sample period. A significant role was played by the reallocation of corporate and private financial flows from Russia to Kazakhstan, including the placement of funds in deposits and the execution of cross-border settlements and transactions. This process does not reflect an expansion of productive investment but rather an enlargement of the deposit base and an increase in financial system turnover. In this context, growth in liquidity aggregates is not equivalent to an increase in real financing of the economy; however, within the factor model it strengthens the contribution of the financial component, which in this period serves primarily a stabilizing function.

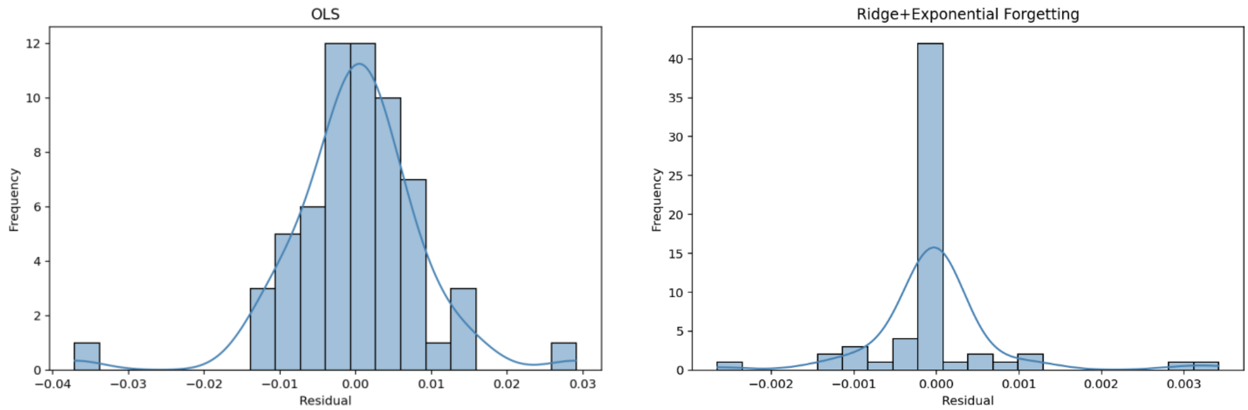
In the most recent period, the role of the real sector in shaping deviations of GDP from its average historical level has changed. In 2025, the influence of the first factor is barely detectable despite the continued growth of real GDP. This suggests that current growth is predominantly quantitative in nature and is driven mainly by an expansion of output through temporary or extensive factors, while qualitative internal drivers of the real sector exert only a limited influence.

The divergence between the observed trajectory and the model-implied estimate may be explained by the fact that each factor aggregates a broad set of indicators, and in periods when subclusters within a factor move in different directions, the resulting index becomes smoothed and loses amplitude. It should also be noted that the reported decomposition is based on a simple linear regression, which reflects the average linear relationship between the factors and GDP over the entire historical sample and relies on standard assumptions regarding correct model specification, parameter stability, and the absence of systematic violations.

Forecasting Methods and Comparative Assessment of Forecast Accuracy.

Despite the simplicity of the linear regression (OLS) specification, the model exhibits a reasonable degree of explanatory power. In an environment of high volatility and frequent irregular shocks characteristic of macroeconomic data, an R^2 of 0.566 indicates that the factor structure indeed captures a substantial portion of the variation in quarterly output. At the same time, the root mean squared error ($rmse = 0.0083$) amounts to approximately 65% of the typical scale of GDP fluctuations ($\sigma = 0.0127$). This implies that while the model explains a non-negligible share of the variance, its ability to reproduce individual quarterly movements remains limited. Residual diagnostics further point to deviations from the classical assumptions of linear regression. The error distribution is characterized by negative skewness and the presence of outliers in the tails (Figure 2). The Jarque-Bera test strongly rejects the null hypothesis of normality of the residuals. The Breusch-Pagan test for heteroscedasticity likewise indicates a violation of the constant variance assumption. These features suggest that the model does not fully adapt to large shocks, such as the sharp depreciation of the tenge in 2015 associated with the transition to a floating exchange rate regime, the 2020 pandemic, or the geopolitical events of 2022.

Figure 2. Comparison of Residual Distributions



The analysis of the regularized approach (Ridge with exponential forgetting) indicates an improvement in model fit within the estimation sample. The weighted coefficient of determination increases to 0.578, while the weighted root mean squared error declines to 0.0053, indicating a substantial improvement in the model's ability to capture quarterly fluctuations. The structure of residual weighting plays a key role in shaping the distribution of the residuals. Most residuals are concentrated in a narrow range around zero, reflected in a sharp increase in the height of the central bin of the histogram, while occasional large deviations remain in the tails of the distribution (Figure 2). The skewness of the distribution shifts to the right, indicating episodes in which the model underpredicts positive GDP growth. Exponential downweighting of older observations allows the model to adapt more rapidly to changes in macroeconomic regimes and reduces the influence of structurally outdated information. At the same time, the ridge penalty stabilizes the coefficient estimates and protects the model from overfitting as the number of explanatory variables increases. As a result, the variance of the errors becomes more stable over time, effectively eliminating heteroscedasticity. Diagnostic tests for residual autocorrelation, however, indicate the presence of statistically significant serial dependence in the model errors. This suggests that, in this specification, part of GDP dynamics is not fully captured by the contemporaneous factors and remains in the residuals as an inertial component.

Forecast accuracy is evaluated in a pseudo real-time setting using an expanding window. The initial estimation sample covers the period from the second quarter of 2010 to the first quarter of 2022. At each subsequent iteration, the estimation window is extended by one quarter, and only information available at that point in time is used, thereby replicating the conditions of a real forecasting round. In each iteration, forecasts are generated for three horizons (nowcast, t+1, and t+2), and it is assumed that for the current quarter only one month of actual data is available. The factors in the dynamic factor model are re-estimated at each iteration, preventing the use of future information. The evaluation sample spans the period from the fourth quarter of 2022 to the second quarter of 2025. Forecast performance is assessed using RMSE, MAE, and Theil's U, which are computed separately for each forecast horizon rather than for each individual iteration.

Benchmark comparisons are conducted against a naïve forecast (as measured by Theil's U) and a simple autoregressive AR(1) model of GDP. The OLS specification with four latent factors as regressors delivers a reasonable level of forecast accuracy given the intrinsic volatility of GDP itself ($\sigma = 0.0127$). Forecast accuracy at longer horizons is higher than for the current quarter (Table 2). Under conditions of incomplete information, the dynamic factor model relies more heavily on smoothed prior estimates of the latent states produced by the Kalman filter, which reduces the precision with which the current phase of the business cycle is identified in the linear specification. As the forecast horizon increases, the importance of within-quarter uncertainty diminishes and factor dynamics are driven increasingly by their own inertial structure, leading to more stable forecasts. Although the OLS model substantially outperforms the naïve benchmark according to Theil's U, it is outperformed by the simple AR(1) model for GDP. This indicates that short-run GDP dynamics remain strongly influenced by output inertia, while factor information contributes more to improving forecast performance at longer horizons.

Table 2. Comparison of Forecast Accuracy for the OLS and AR(1) Models

GDP, OLS

	RMSE	MAE	Theil U1
nowcast	0.007428	0.006301	0.298309
t+1	0.006165	0.004603	0.267683
t+2	0.006232	0.004911	0.269399

GDP, AR(1)

	RMSE	MAE	Theil U1
nowcast	0.006173	0.004694	0.263819
t+1	0.006319	0.004962	0.275454
t+2	0.006372	0.004983	0.279015

In the regularized specification, the first lag of GDP is included in addition to the latent factors, reflecting the presence of residual autocorrelation. It should be noted, however, that the autocorrelation structure of the errors may arise both from the regularization itself, which partially shrinks the coefficients, and from the use of exponential weighting, which increases the influence of recent observations; as a result, output inertia may be concentrated primarily in the most recent years. The value of the regularization parameter (λ, shrinkage coefficient) is selected via cross-validation, ensuring an appropriate balance between bias and variance. The results of the regularized model (Table 3) show higher forecast accuracy than both the OLS specification and the GDP AR(1) benchmark. A key source of this improvement is the model's focus on current economic relationships, together with the partial incorporation of output inertia through the autoregressive component.

Table 3. Comparison of Forecast Accuracy for the Regularized Model and AR(1)

GDP, Ridge+Exp.Forg.

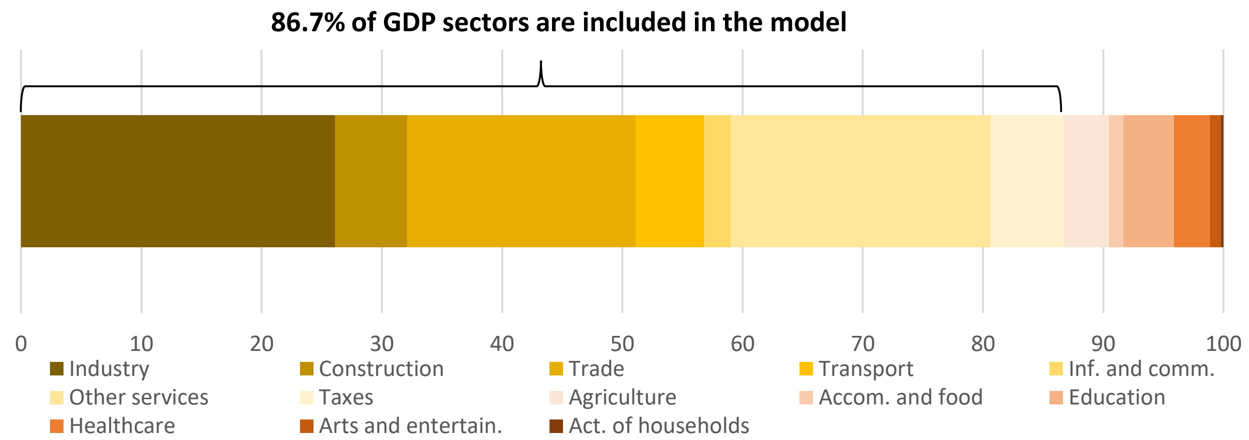
	RMSE	MAE	Theil U1
nowcast	0.005388	0.004306	0.221441
t+1	0.005836	0.004531	0.248103
t+2	0.006063	0.004487	0.261612

GDP, AR(1)

	RMSE	MAE	Theil U1
nowcast	0.006173	0.004694	0.263819
t+1	0.006319	0.004962	0.275454
t+2	0.006372	0.004983	0.279015

In addition to GDP forecasting, the estimated latent factors are also used to forecast individual macroeconomic and sectoral indicators. The regression parameters are estimated using the regularized method, which improves stability and places greater weight on current relationships. Forecast accuracy is highest for GDP components, particularly in the services sector, which largely reflects the nature of the data, their level of aggregation, and their stronger synchronization with common macroeconomic factors (Appendix). Relatively high forecast accuracy is also observed for variables related to deposits and credit, as well as for monetary aggregates reflecting domestic financial conditions. By contrast, exchange rate forecasts are less accurate, reflecting the high sensitivity of exchange rates to external shocks, news, and short-term expectations, which are only weakly captured by a model focused on common macroeconomic components. Results are reported only for those variables for which the root mean squared error is below their standard deviation, indicating a substantial reduction in forecast uncertainty relative to the natural variability of the corresponding series.

Figure 3. Structure of Kazakhstan's GDP in 2024



According to data from the Bureau of National Statistics of the Republic of Kazakhstan, the sectors included in the panel of forecasted variables account in total for approximately 86.7% of the economy's gross value added (Figure 3). This implies that the proposed factor model covers the overwhelming majority of the output structure and allows for coherent forecasts of the main components of GDP to be formed on the basis of a common macroeconomic factor framework. The model therefore serves not only as a tool for aggregate GDP forecasting but also as an econometric foundation for analyzing and forecasting the dynamics of the main sectors of the economy, whose evolution is driven by systemic macroeconomic relationships.

Conclusion

This paper proposes and empirically evaluates a short-term forecasting model for Kazakhstan's GDP and for individual sectors of the economy based on a dynamic factor structure with regularization and adaptive weighting of observations. The use of a broad panel of macroeconomic indicators makes it possible to extract stable latent components capturing the main sources of co-movement in the economy, while the application of the Kalman filter ensures that the model operates properly under conditions of incomplete and asynchronous information.

The empirical results show that the baseline OLS model with factors provides acceptable predictive performance and substantially outperforms naive benchmarks, but remains inferior to a simple GDP AR(1) model at short horizons. This highlights the important role of output inertia in GDP dynamics, especially in periods of heightened uncertainty. The regularized specification with an exponential decay mechanism substantially improves forecast accuracy, particularly for the current period, by emphasizing recent relationships and partially capturing inertial dynamics.

A further improvement in forecast performance may be achieved by expanding the set of explanatory variables to include forward-looking indicators, without compromising the stability of the parameter estimates.

The results confirm that the factor structure is most effective for forecasting variables whose dynamics are driven by stable macroeconomic relationships, including output in the services sector, monetary aggregates, deposits, and credit. By contrast, variables characterized by high volatility and strong exposure to external shocks are less amenable to accurate forecasting within the present specification. Overall, the use of latent factors makes it possible to generate forecasts for a substantial share of economic sectors, providing broad coverage of the dynamics of the main components of GDP.

It should be emphasized that the proposed model does not aim to provide a structural interpretation of the drivers of economic growth and does not constitute an estimate of potential output in a strict economic sense; rather, it allows for a quantitative assessment of the contribution of individual factors to deviations of actual GDP dynamics from their average historical growth rate. Its primary purpose is to improve the quality of short-term forecasts. Future extensions may include the incorporation of nonlinear elements, the explicit treatment of possible regime shifts, and the integration of the factor-based approach with more structural macroeconomic models.

References

De Mol, C., Giannone, D., & Reichlin, L. (2008). Forecasting using a large number of predictors: Is Bayesian regression a valid alternative to principal components? Journal of Econometrics, 146(2), 318–328. https://doi.org/10.1016/j.jeconom.2008.08.011

Doz, C., Giannone, D., & Reichlin, L. (2006). A quasi-maximum likelihood approach for large approximate dynamic factor models. Review of Economics and Statistics, 94(4), 1014–1024. https://doi.org/10.1162/REST_a_00225

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. Journal of Monetary Economics, 55(4), 665–676. https://doi.org/10.1016/j.jmoneco.2008.05.010

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Mariano, R. S., & Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. Journal of Applied Econometrics, 18(4), 427–443. https://doi.org/10.1002/jae.695

Orlov, K. (2019). Assessment and analysis of the effectiveness of dynamic factor models for GDP estimation and forecasting: Evidence from Kazakhstan. Working Paper. https://www.nationalbank.kz/file/download/8979

Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. Journal of Business & Economic Statistics, 20(2), 147–162. https://doi.org/10.1198/073500102317351921

West, M., & Harrison, J. (1997). Bayesian Forecasting and Dynamic Models (2nd ed.). Springer. https://doi.org/10.1007/978-1-4612-0844-2

Appendix

Table 1. List of Variables

	Variable	Frequency	SA	Type
Real activity	Mining	monthly	+	index
	Manufacturing	monthly	+	index
	Construction	monthly	+	index
	Residential area	monthly	+	index
	Retail trade	monthly	+	index
	Wholesale trade	monthly	+	index
	Freight turnover	monthly	+	index
	Passenger turnover	monthly	+	index
	Communications	monthly	+	index
	Investments	monthly	+	index
	Financial and insurance	quarterly	+	real level
	Real estate	quarterly	+	real level
	Professional, scientific and technical	quarterly	+	real level
	Administrative and support	quarterly	+	real level
	Public administration	quarterly	+	real level
	Others	quarterly	+	real level
	Taxes	quarterly	+	real level
	Wages	quarterly	+	level, deflated
Prices	CPI	monthly	+	index
	PPI manufacturing	monthly	-	index
Financing and liquidity	Credits to business (issued)	monthly	+	level, deflated
	Credits to households (issued)	monthly	+	level, deflated
	Deposits of business (stock)	monthly	+	level, deflated
	Deposits of households (stock) (p 0.055)	monthly	+	level, deflated
	M0	monthly	+	level, deflated
	M2	monthly	+	level, deflated
Fiscal	Current budget expenditures	quarterly	+	level, deflated
	Capital budget expenditures	quarterly	+	level, deflated
Corporate	Cash	quarterly	-	level, deflated
	Accounts receivable	quarterly	-	level, deflated
	PPE	quarterly	+	level, deflated
	Unfinished construction	quarterly	-	level, deflated
	Authorised capital	quarterly	-	level, deflated
	EBTM	quarterly	-	ratio
External	REER	monthly	-	index
	Brent	monthly	-	nominal price
	Copper	monthly	-	nominal price
	Iron	monthly	-	nominal price
	FAO	monthly	-	index
	USD/KZT	monthly	-	nominal level
	RUB/KZT	monthly	-	nominal level

Table 2. Forecast Accuracy of Variables

Manufacturing σ **0.040320**

	RMSE	MAE	Theil U1
nowcast	0.036422	0.028657	0.655663
t+1	0.035807	0.028713	0.667553
t+2	0.037675	0.031127	0.726535

Wholesale trade σ **0.073435**

	RMSE	MAE	Theil U1
nowcast	0.041020	0.025909	0.687732
t+1	0.038705	0.025994	0.656044
t+2	0.044526	0.027751	0.767205

Communication σ **0.037672**

	RMSE	MAE	Theil U1
nowcast	0.033818	0.028117	0.662156
t+1	0.032744	0.026614	0.630331
t+2	0.031096	0.026000	0.592300

Investments σ **0.044727**

	RMSE	MAE	Theil U1
nowcast	0.038056	0.032993	0.596136
t+1	0.039047	0.034613	0.630325
t+2	0.042272	0.036350	0.702937

FAO σ **0.056514**

	RMSE	MAE	Theil U1
nowcast	0.032325	0.024222	0.749844
t+1	0.032168	0.025356	0.852111
t+2	0.036386	0.026368	0.876298

Brent σ **0.177268**

	RMSE	MAE	Theil U1
nowcast	0.115695	0.106612	0.643760
t+1	0.113947	0.089982	0.854572
t+2	0.117725	0.097330	0.877179

Residential area σ **0.126425**

	RMSE	MAE	Theil U1
nowcast	0.094782	0.072842	0.894429
t+1	0.085810	0.066772	0.837883
t+2	0.088833	0.069127	0.869364

Passenger turn. σ **0.185458**

	RMSE	MAE	Theil U1
nowcast	0.079711	0.063207	0.589334
t+1	0.053957	0.026619	0.690507
t+2	0.040792	0.026896	0.566065

PPI manufacturing σ **0.063600**

	RMSE	MAE	Theil U1
nowcast	0.042279	0.037022	0.606878
t+1	0.039515	0.033892	0.647226
t+2	0.041845	0.036387	0.656951

CPI σ **0.013432**

	RMSE	MAE	Theil U1
nowcast	0.007483	0.006373	0.135638
t+1	0.008615	0.007287	0.157442
t+2	0.009600	0.008652	0.176909

REER σ **0.056128**

	RMSE	MAE	Theil U1
nowcast	0.034858	0.029906	0.564315
t+1	0.041347	0.034420	0.794994
t+2	0.040692	0.029855	0.781175

Copper σ **0.088394**

	RMSE	MAE	Theil U1
nowcast	0.064619	0.058275	0.759430
t+1	0.060617	0.052529	0.892078
t+2	0.050948	0.041251	0.747097

Table 2. Forecast Accuracy of Variables (continued)

Iron σ **0.181309**

	RMSE	MAE	Theil U1
nowcast	0.143335	0.120523	0.919725
t+1	0.123661	0.107421	0.865747
t+2	0.122268	0.102889	0.853553

USD/KZT σ **0.063891**

	RMSE	MAE	Theil U1
nowcast	0.042465	0.036911	0.552077
t+1	0.042040	0.037817	0.666196
t+2	0.040229	0.036830	0.645800

Credits to business σ **0.169506**

	RMSE	MAE	Theil U1
nowcast	0.118836	0.094183	0.889670
t+1	0.114749	0.090289	0.887496
t+2	0.114882	0.091596	0.886453

Credits to households σ **0.118595**

	RMSE	MAE	Theil U1
nowcast	0.045932	0.035811	0.422669
t+1	0.036815	0.027554	0.467651
t+2	0.037030	0.030149	0.452726

Deposits of business σ **0.056842**

	RMSE	MAE	Theil U1
nowcast	0.027614	0.021457	0.649714
t+1	0.031046	0.024123	0.777288
t+2	0.029979	0.023209	0.741716

Deposits of households σ **0.069506**

	RMSE	MAE	Theil U1
nowcast	0.021030	0.018063	0.301724
t+1	0.021996	0.018521	0.324672
t+2	0.022370	0.018043	0.333963

M0 σ **0.050063**

	RMSE	MAE	Theil U1
nowcast	0.027606	0.022173	0.645454
t+1	0.027347	0.021836	0.875386
t+2	0.027939	0.022507	0.909372

M2 σ **0.042281**

	RMSE	MAE	Theil U1
nowcast	0.019633	0.015486	0.460124
t+1	0.023706	0.018960	0.601292
t+2	0.023162	0.018251	0.583879

Financial and insur. σ **0.047498**

	RMSE	MAE	Theil U1
nowcast	0.009727	0.007910	0.666891
t+1	0.010531	0.008377	0.739551
t+2	0.009343	0.007617	0.663219

Real estate σ **0.019521**

	RMSE	MAE	Theil U1
nowcast	0.003561	0.002820	0.389904
t+1	0.002590	0.002165	0.284257
t+2	0.002755	0.002311	0.302083

Prof. scientific σ **0.028894**

	RMSE	MAE	Theil U1
nowcast	0.010081	0.008428	0.565508
t+1	0.009747	0.008375	0.597354
t+2	0.009707	0.008308	0.632832

Administrative σ **0.032072**

	RMSE	MAE	Theil U1
nowcast	0.006702	0.005754	0.702645
t+1	0.006731	0.005637	0.783085
t+2	0.007359	0.006027	0.773800

Table 2. Forecast Accuracy of Variables (continued)

Public administr. σ **0.038547**

	RMSE	MAE	Theil U1
nowcast	0.024404	0.021081	0.744338
t+1	0.024169	0.021691	0.792367
t+2	0.021412	0.018538	0.707553

Other serivces σ **0.273217**

	RMSE	MAE	Theil U1
nowcast	0.022465	0.016779	0.882107
t+1	0.022930	0.017509	0.883350
t+2	0.020351	0.015507	0.784469

Taxes σ **0.035644**

	RMSE	MAE	Theil U1
nowcast	0.014804	0.010646	0.506116
t+1	0.013988	0.009662	0.472067
t+2	0.014830	0.010289	0.499374