

Article

Forecasting of GDP Growth in the South Caucasian Countries Using Hybrid Ensemble Models

Gaetano Perone ^{1,2,*}  and Manuel A. Zambrano-Monserrate ³ ¹ Department of Economics and Management, University of Pisa, Via Cosimo Ridolfi 10, 56124 Pisa, Italy² Kutaisi International University, Akhlagazrdoba Ave. Lane 5/7, 4600 Kutaisi, Georgia³ Universidad Espíritu Santo, Samborondón 0901952, Ecuador; manuelzambranom@uees.edu.ec

* Correspondence: gaetano.perone@ec.unipi.it

Abstract

This study aimed to forecast the gross domestic product (GDP) of the South Caucasian nations (Armenia, Azerbaijan, and Georgia) by scrutinizing the accuracy of various econometric methodologies. This topic is noteworthy considering the significant economic development exhibited by these countries in the context of recovery post COVID-19. The seasonal autoregressive integrated moving average (SARIMA), exponential smoothing state space (ETS) model, neural network autoregressive (NNAR) model, and trigonometric exponential smoothing state space model with Box–Cox transformation, ARMA errors, and trend and seasonal components (TBATS), together with their feasible hybrid combinations, were employed. The empirical investigation utilized quarterly GDP data at market prices from 1Q-2010 to 2Q-2024. According to the results, the hybrid models significantly outperformed the corresponding single models, handling the linear and nonlinear components of the GDP time series more effectively. Rolling-window cross-validation showed that hybrid ETS-NNAR-TBATS for Armenia, hybrid ETS-NNAR-SARIMA for Azerbaijan, and hybrid ETS-SARIMA for Georgia were the best-performing models. The forecasts also suggest that Georgia is likely to record the strongest GDP growth over the projection horizon, followed by Armenia and Azerbaijan. These findings confirm that hybrid models constitute a reliable technique for forecasting GDP in the South Caucasian countries. This region is not only economically dynamic but also strategically important, with direct implications for policy and regional planning.



Academic Editor: Julie Le Gallo

Received: 29 June 2025

Revised: 1 September 2025

Accepted: 2 September 2025

Published: 10 September 2025

Citation: Perone, G., & Zambrano-Monserrate, M. A. (2025). Forecasting of GDP Growth in the South Caucasian Countries Using Hybrid Ensemble Models. *Econometrics*, 13(3), 35. <https://doi.org/10.3390/econometrics13030035>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: GDP forecasting; hybrid ensemble models; machine learning; rolling-origin cross-validation; South Caucasus

JEL Codes: C22; C53; E01; O53

1. Introduction

Forecasting economic growth, particularly gross domestic product (GDP), represents a cornerstone activity in macroeconomic analysis, given its importance in shaping fiscal policy, investment decisions, and overall economic planning (Domit et al., 2019; Mariano & Ozmucur, 2021). Accurate GDP forecasting becomes even more critical during periods of significant uncertainty or rapid recovery phases, such as the post-pandemic era following COVID-19 (Sabri et al., 2023; Rostan et al., 2024). Such forecasting is especially vital for developing and transition economies, where timely policy interventions can significantly influence sustainable economic development (Cepni et al., 2019; Richardson et al., 2021).

The South Caucasian countries—Armenia, Azerbaijan, and Georgia—represent an intriguing regional cluster characterized by substantial economic recovery and structural transformations following the COVID-19 crisis (Darvas, 2011; Poghosyan & Poghosyan, 2020). Notably, Georgia recorded a remarkable average real GDP growth of 9.72% over the period 2021–2024, followed by Armenia with 8.15% and Azerbaijan with 3.95%. All these rates are significantly higher than those of the G7 economies, which registered an average growth of only 2.82% in the same period (International Monetary Fund, 2025). This highlights the stronger post-pandemic resilience of the South Caucasian economies.

This region, strategically located between Europe and Asia, presents unique economic dynamics driven by distinct geopolitical relationships, varying resource endowments, and diverse paths towards economic integration with global markets (Gerasimov et al., 2015; Jibuti, 2020). Given their economic heterogeneity and strategic importance, forecasting GDP in these countries holds critical implications for policymakers, investors, and international development organizations.

Previous research on GDP forecasting has extensively explored various econometric and statistical methodologies, highlighting the increasing application of hybrid models to capture complex economic dynamics. Gerasimov et al. (2015) pioneered forecasting scenarios—potential, optimistic, and pessimistic—in the North Caucasus, emphasizing the relevance of structural constraints in economic modeling. Similarly, Poghosyan and Poghosyan (2020) demonstrated the efficacy of dynamic factor models (DFM), specifically factor-augmented autoregressive regression (FAAR), factor-augmented vector autoregression (FAVAR), and Bayesian FAVAR, in enhancing forecast accuracy in transitional economies such as Armenia. Krkoska and Teksoz (2007) provided robust evidence that forecast accuracy significantly improves alongside institutional reforms and enhanced data availability, emphasizing the need for comprehensive modeling approaches. Additionally, Tsuchiya (2023) evaluated the accuracy of World Bank forecasts, finding significant improvements post the 2008 financial crisis while highlighting the rationality and efficiency of these forecasts within asymmetric loss contexts.

Furthermore, recent advances have highlighted hybrid and machine learning methodologies as increasingly potent in GDP prediction. Şen Doğan and Midiliç (2019), Chernis and Sekkel (2017), and Claudio et al. (2020) demonstrated that Mixed Data Sampling (MIDAS) models integrating monthly and quarterly indicators substantially enhance forecast precision, particularly when macroeconomic data are sparse or delayed. Machine learning techniques have also gained traction, with Fan (2024) introducing Least Squares Support Vector Machine (LSSVM) models, achieving superior accuracy rates in regional GDP forecasting. Ciccieri et al. (2020) further underlined the predictive superiority of nonlinear models such as support vector regression (SVR), k-nearest neighbors (KNN), and nonlinear autoregressive model with exogenous variables (NARX), particularly during economic downturns and recovery phases. Studies by Longo et al. (2022) and Shams et al. (2024) corroborate these findings, highlighting the efficacy of neural network architectures, such as recurrent neural networks (RNNs) and long short-term memory (LSTM), in capturing nonlinear economic patterns.

However, despite extensive research into econometric and machine learning methodologies, a significant theoretical gap persists regarding their application to the South Caucasian region specifically. Previous studies frequently concentrate on broader or distinctly different regional contexts, often neglecting the unique geopolitical and economic nuances characterizing Armenia, Azerbaijan, and Georgia (Darvas, 2011; Jibuti, 2020). Consequently, this leaves substantial room to explore and validate hybrid ensemble models explicitly tailored to the economic conditions and data characteristics of these South Caucasian economies.

Considering this theoretical gap, the main objective of this paper is to forecast GDP growth in Armenia, Azerbaijan, and Georgia from 1Q-2010 to 2Q-2024 using hybrid ensemble models. Specifically, this research evaluates several econometric methodologies, including seasonal autoregressive integrated moving average (SARIMA), exponential smoothing state space (ETS), neural network autoregressive (NNAR), and trigonometric exponential smoothing state space model with Box–Cox transformation, ARMA errors, and trend and seasonal components (TBATS), alongside their feasible hybrid combinations. The aim is to ascertain the most accurate forecasting methodology for each country, thereby enhancing the precision and reliability of economic projections pertinent to policy formulation and strategic decision-making. The original contribution of this paper is a first, country-comparable assessment of feasible ETS-NNAR-SARIMA-TBATS hybrid ensembles for South Caucasian GDP, implemented under a unified rolling-window cross-validation design and with reported prediction intervals.

The study's empirical findings notably reveal that hybrid models significantly outperform their individual counterparts by effectively managing both linear and nonlinear components inherent in GDP time series. Rolling-window cross-validation identified distinct optimal models for each country: hybrid ETS-NNAR-TBATS for Armenia, hybrid ETS-NNAR-SARIMA for Azerbaijan, and hybrid ETS-SARIMA for Georgia. Moreover, projections indicate significant GDP growth, with Georgia expected to achieve the highest rate, followed by Armenia and Azerbaijan, over the forthcoming two and a half years. These results highlight the robustness and reliability of hybrid ensemble models in predicting economic growth within the South Caucasian context. Practically, the resulting short-term GDP forecasts inform ministries of finance and international organizations in short-term fiscal planning, debt management, and the timing of development programs.

Despite the contributions of this article, we recognize three main limitations: (i) the relatively short historical spans for some series, especially for Armenia; (ii) the reliance on a univariate setup without external predictors, which may constrain long-term performance; and (iii) the use of equal-weight hybridization for parsimony rather than optimized weighting schemes. These limitations, however, do not affect the main results, as the evaluation is based on repeated out-of-sample validation and the findings are consistent across countries and model classes. Moreover, both the univariate design and the equal-weight combinations are consistent with established evidence in the forecasting literature, which highlights their robustness in small-sample environments (Timmermann, 2006; Claeskens et al., 2016).

While this study focuses solely on univariate GDP models, the forecasts can still be interpreted within the broader macroeconomic debate on economic growth and external factors. Rodrik (2008), for example, has shown that real exchange rate dynamics are closely linked to growth paths in developing and emerging economies (Rodrik, 2008). More recent evidence confirms that exchange rate misalignments and volatility can affect competitiveness, investment, and output stability (Habib et al., 2017; Sein & Sah, 2025). Recognizing these connections helps explain why accurate GDP forecasts matter for economic analysis, even when they rely exclusively on univariate models.

Overall, this paper addresses the issue of identifying the most suitable forecasting approach for predicting GDP dynamics in the South Caucasus. This is particularly relevant because reliable and timely short-term forecasts are critical for transition economies, while empirical evidence for this region remains scarce. To ensure robust findings, the analysis evaluates standard univariate time-series models (ETS, NNAR, SARIMA, and TBATS) and their feasible hybrid ensembles, assessed through multiple accuracy metrics and a rolling-origin cross-validation framework.

The remainder of this article is organized as follows: Section 2 presents a detailed theoretical framework encompassing relevant econometric and machine learning forecasting

methodologies. Section 3 outlines the methodological approach and validation techniques employed. Section 4 presents the data used. Section 5 provides an in-depth analysis of the empirical results derived from the forecasting models. Lastly, Section 6 discusses the broader implications of these findings, offering recommendations for policy and future research directions in the context of economic forecasting and planning.

2. Literature Review

Recent literature on GDP forecasting has increasingly explored diverse econometric and machine learning approaches, highlighting both methodological innovation and regional specificity. To structure this review, we group methods into three families: (i) classical econometric models (ARIMA/SARIMA, ETS, BVAR/DFM, bridge models); (ii) machine-learning approaches (NNAR, tree-based methods, SVM, shallow neural networks); and (iii) hybrid and mixed-frequency/nowcasting frameworks (model averaging, MIDAS, bridge, and ensemble combinations). We use this classification to frame the discussion and to motivate the empirical choices in Section 3. Among classical approaches, [Gerasimov et al. \(2015\)](#) proposed three forecasting scenarios—potential, optimistic, and pessimistic—for the socioeconomic development of Russia’s North Caucasus Federal District. Their study emphasizes structural constraints such as imbalances between income and consumption growth and weak innovation in the private sector, issues that could be relevant for post-Soviet economies more broadly. Similarly, [Poghosyan and Poghosyan \(2020\)](#) applied DFM, such as FAAR, FAVAR, and Bayesian FAVAR, to forecast Armenia’s GDP, underscoring the improved forecast accuracy of these models in transitional economies. [Jibuti \(2020\)](#) further examined regional economic divergence and convergence in Georgia, emphasizing the role of competitive growth centers to foster balanced regional development.

From a methodological standpoint, hybrid and high-frequency data models have gained prominence. [Şen Doğan and Midiliç \(2019\)](#) show that combining MIDAS with traditional autoregressive distributed lag (ADL) and factor autoregressive distributed lag (FADL) structures yields improved quarterly GDP forecasts in Turkey by integrating monthly indicators. Similarly, [Claudio et al. \(2020\)](#) utilize MIDAS to forecast East German GDP, confirming the usefulness of mixed-frequency data when timely macroeconomic indicators are scarce. These findings are echoed by [Chernis and Sekkel \(2017\)](#) and [Jiang et al. \(2017\)](#), who both employ DFM and MIDAS to nowcast GDP in Canada and China, respectively, demonstrating their effectiveness in data-rich environments. [Mariano and Ozmucur \(2021\)](#), using dynamic latent factor models (DLFM) and MIDAS for the Philippines, further support the robustness of mixed-frequency models in capturing short-term fluctuations. [Krkoska and Teksoz \(2007\)](#), evaluating GDP growth forecasts for 25 transition countries using efficiency and bias tests, demonstrate that forecast accuracy improves significantly with progress in institutional reforms and increased data availability.

Machine learning methods are also increasingly adopted in GDP prediction. [Cicceri et al. \(2020\)](#) assess various nonlinear models, including SVR, KNN, and NARX, to forecast recessions in Italy. Their findings suggest that these models perform better in turbulent economic periods, where linear models often fail to capture complex dynamics. [Fan \(2024\)](#) introduces the LSSVM for regional GDP prediction in China, achieving an accuracy rate of approximately 96.5% and demonstrating the model’s superiority compared to traditional techniques. In a cross-country context, [Alaminos et al. \(2022\)](#) apply quantum-inspired algorithms such as quantum neural networks (QNN) and support vector regression quantum bat algorithm (SVRQBA) across 70 countries. These models outperform traditional methods, particularly in volatile environments, reinforcing the need for adaptive and robust forecasting tools. Other studies, such as [Tümer and Akkuş \(2018\)](#), [Yoon \(2021\)](#), and [Ye-nilmez and Mugenzi \(2023\)](#), further corroborate the value of neural networks and ensemble

learning in capturing nonlinear patterns in GDP data. [Longo et al. \(2022\)](#) explore the use of recurrent neural network (RNN) and DFM with a generalized autoregressive score (GAS) to predict U.S. GDP, while [Shams et al. \(2024\)](#) introduce a PC-LSTM-RNN model for urban GDP prediction in India, highlighting the growing role of deep learning architectures.

In smaller or developing economies, data limitations often necessitate the use of factor-based or hybrid models. [Madhou et al. \(2020\)](#), using a FAVAR approach, show improved GDP forecasting for Mauritius, an economy characterized by limited high-frequency data. Likewise, [Maccarrone et al. \(2021\)](#) compare autoregressive model with exogenous variables (ARX), KNN, linear regression, and SARIMAX for the U.S. economy, concluding that hybrid models yield higher accuracy, especially when nonlinear relationships are present. [Mohamed \(2022\)](#) applies ARIMA to the Somali economy, and [Bragoli and Fosten \(2018\)](#) assess DFM and AR-based techniques for India, both illustrating the potential of simpler models in low-data contexts. [Domit et al. \(2019\)](#), using a medium-scale Bayesian VAR model, demonstrate strong predictive capacity in the UK context, while [Higgins et al. \(2016\)](#) apply a Bayesian VAR model to forecast China's economic growth, emphasizing the advantages of Bayesian techniques for dealing with uncertainty. [Abdić et al. \(2020\)](#) provide evidence for Bosnia and Herzegovina using ARIMA and factor models, and [Rusnák \(2016\)](#) applies DFM to forecast GDP in the Czech Republic. [Tsuchiya \(2023\)](#) further supports these findings through a comprehensive evaluation of the World Bank's GDP forecasts across 130 countries, highlighting improvements in forecast performance post-2008 financial crisis and underscoring the rationality of forecasts when accounting for asymmetric losses. [Darvas \(2011\)](#) also contributes to the literature by analyzing cross-country growth regressions for Central and Eastern Europe, the Caucasus, and Central Asia, emphasizing the significant impact of the 2008–2009 global financial crisis on regional growth forecasts.

[Richardson et al. \(2021\)](#) take a comprehensive approach to nowcasting GDP in New Zealand using a suite of machine learning models, including least absolute shrinkage and selection operator (LASSO), Ridge regression, gradient boosting (GB), and Random Forest. Their results suggest that model performance is highly sensitive to the nowcasting horizon and the data characteristics, a finding significant for countries with variable data quality. [Tsuchiya \(2024\)](#) provides a similar assessment of forecasts by the European Bank for Reconstruction and Development (EBRD), noting improved accuracy over time and reduced information rigidity across 38 countries, aligning well with broader trends in forecast methodology evolution. Finally, [Dritsaki and Dritsaki \(2023\)](#) employed a modified ETS framework to forecast Greek GDP, demonstrating that the ETS model with multiplicative error, damped trend, and additive seasonality outperformed alternative ETS configurations as well as standard ARIMA models. The effectiveness of ETS models in capturing the structural components of GDP has also been recently confirmed by [Elbatal et al. \(2025\)](#).

Recent studies have increasingly broadened the scope of macroeconomic forecasting by combining traditional econometric approaches with machine learning techniques. For instance, [Ballarin et al. \(2024\)](#) highlight the potential of a novel machine learning paradigm ("reservoir computing") for long-term forecasting with heterogeneous data, while [Schorfheide and Song \(2021\)](#) illustrate the effectiveness of mixed-frequency vector autoregression (MF-VAR) models for short-term nowcasting, both applied to the US. A recent survey by [Babii et al. \(2024\)](#) further underscores the growing importance of these methods in economic forecasting. Our contribution builds on this literature by extending the comparison of univariate and hybrid ensemble models to the South Caucasian countries, a region that is not only under-researched in forecasting studies but also geopolitically and economically strategic, given its role as an energy corridor and its vulnerability to external shocks.

The evidence above guides our empirical design. Quarterly GDP series in the South Caucasus are short and strongly seasonal; therefore, SARIMA and ETS provide transparent baselines. The presence of potential nonlinear patterns justifies NNAR as a complementary tool. Deep architectures and data-hungry learners are not appropriate given the sample length. Mixed-frequency and bridge models can improve short-term performance when timely monthly indicators are available; however, our main specification focuses on univariate quarterly GDP to ensure comparability across countries. This alignment between the literature and the data environment sets the scope for the model set evaluated in Section 3. Table 1 provides a summary of relevant studies on GDP forecasting across different countries.

Table 1. Summary of relevant studies on forecasting GDP across the world.

Authors	Data Used	Methodology	Investigated Area
Krkoska and Teksoz (2007)	1994–2004	Bias and efficiency evaluation	25 transition countries in Eastern Europe and former Soviet Union
Darvas (2011)	Up to 2010	Cross-country growth regressions	Central and Eastern Europe, Caucasus, and Central Asia
Gerasimov et al. (2015)	Not specified	Socioeconomic forecasting scenarios	North Caucasus Federal District, Russia
Higgins et al. (2016)	2000–2015	BVAR	China
Rusnák (2016)	2005–2012	DFM	Czech Republic
Chernis and Sekkel (2017)	1999–2016	DFM, MIDAS, bridge regressions	Canada
Jiang et al. (2017)	2000–2016	DFM, MIDAS	China
Bragoli and Fosten (2018)	1996–2015	AR, BR, DFM	India
Tümer and Akkuş (2018)	1996–2015	ANN with FFBP	13 countries
Cepni et al. (2019)	2003–2018	DFM, LASSO	Brazil, Indonesia, Mexico, South Africa, Turkey
Şen Doğan and Midiliç (2019)	2000–2016	ADL, FADL, MIDAS, and hybrid models	Turkey
Domit et al. (2019)	1987–2015	BVAR	UK
Abdić et al. (2020)	2006–2016	ARIMA, BM, FM	
Cicceri et al. (2020)	1995–2019	AR, BT, KNN, NAR, NARX, OLS, SVR	Italy
Claudio et al. (2020)	1991–2018	MIDAS	East German
Jibuti (2020)	Last 5–10 years	Logarithmic growth models	Georgia
Madhou et al. (2020)	2003–2016	BVAR, FVAR	Mauritius
Poghosyan and Poghosyan (2020)	1996–2019	Factor-augmented models (FAAR, FAVAR, Bayesian FAVAR)	Armenia
Maccarrone et al. (2021)	1976–2020	ARX, KNN, LR, SARIMAX	US
Mariano and Ozmucur (2021)	1999–2019	DLFM, MIDAS	Philippines
Richardson et al. (2021)	2009–2019	AR, DFM, EN, GB, LASSO, NN, Ridge, SVM,	New Zealand
Yoon (2021)	2001–2018	GB, RF	Japan
Alaminos et al. (2022)	1980–2018	DSVR, DNDT, DRCNN, SVRQBA, QBM, QNN	70 countries
Gu et al. (2022)	1992–2016	ARIMA, ARIMAX, SARIMA, LR	Chinese provinces
Longo et al. (2022)	1970–2021	RNN, DFM-GAS	US
Mohamed (2022)	1960–2022	ARIMA	Somalia
Dritsaki and Dritsaki (2023)	1995–2022	ARIMA, ETS	Greece
Puttanapong et al. (2023)	2000–2019	GLS, NN, RF, SVR	Thai provinces
Sabri et al. (2023)	1980–2022	ANN, ETS	Pakistan
Tsuchiya (2023)	1999–2019	World Bank growth forecast evaluation	130 countries and 6 global regions
Yenilmez and Mugenzi (2023)	1960–2021	ARIMA, GRNN, MLP, LSTM	Rwanda

Table 1. Cont.

Authors	Data Used	Methodology	Investigated Area
Almarashi et al. (2024)	1969–2021	ARIMA, ETS, NNAR, TBATS and hybrid models	Saudi Arabia
Fan (2024)	Not specified	LSSVM	Specific region in China
Rostan et al. (2024)	1994–2022	ARIMA, ETS, LR, MCS, PM, Wavelet	17 countries of the Eurozone
Shams et al. (2024)	1961–2021	PC-LSTM-RNN	India
Tsuchiya (2024)	1994–2019	Forecast evaluation and asymmetric loss	38 countries covered by EBRD
Kant et al. (2025)	1992–2018	DFM, LASSO, MIDAS, RF, RSR	Netherlands
Zhang et al. (2025)	2011–2022	DFM, EN, JMA, LASSO, MIDAS, Ridge	China

Notes: ADL, autoregressive distributed lag; ANN, artificial neural network; AR, autoregressive model; ARX, AR with exogenous variables; ARIMA, autoregressive integrated moving average; ARIMAX, ARIMA with exogenous variables; BM, bridge model; BR, bridge models; BT, Boosted Trees; BVAR, Bayesian vector autoregression; DBN, deep belief network; DFM, dynamic factor model; DFM-GAS, DFM with a generalized autoregressive score (GAS); DLFM, dynamic latent factor models; DNDT, deep neural decision trees; DRCNN, deep recurrent convolution neural network; DSVR, deep learning linear support vector machines; EN, elastic net; ETS, exponential smoothing; FADL, factor autoregressive distributed lag; FM, factor model; FFBP, feedforward backpropagation; GB, gradient boosting; GLS, generalized least squares; GRNN, generalized regression neural network; KNN, k-nearest neighbors; JMA, Jackknife model averaging; LASSO, least absolute shrinkage and selection operator; LR, linear regression; LSSVM, least squares support vector machine; LSTM, long short-term memory; MCS, Monte Carlo simulation; MIDAS, mixed data sampling; MLP, multilayer perceptron; NAR, nonlinear autoregressive model; NARX, NAR with exogenous variables; NN, neural network; OLS, ordinary least squares; PC, Pearson correlation; PM, polynomial model; QBM, quantum Boltzmann machines; QNN, quantum neural networks; RF, random forest; RNN, recurrent neural network; RSR, random subspace regression; SARIMA, seasonal autoregressive integrated moving average; SARIMAX, SARIMA with exogenous variables; SVM, support vector machine regression; SVR, support vector regression; SVRQBA, support vector regression quantum bat algorithm; TBATS, trigonometric seasonality, Box-Cox transformation, ARMA errors, trend, and seasonal components.

3. Methods

This section sets out the forecasting methods, explains why they are suitable for quarterly GDP in the South Caucasus, and describes the evaluation strategy and software. We focus on SARIMA, ETS, NNAR, and TBATS, as well as feasible hybrids, to balance interpretability and flexibility under short sample lengths.

3.1. Forecasting Models

This paper's empirical strategy includes the employment of the following forecasting methods: ETS, NNAR, SARIMA, TBATS, and their hybrid combinations. ETS is one of the most widely used approaches in time series forecasting and is based on exponential smoothing of components such as level, trend, and seasonality. We include ETS as a baseline because it captures structural components transparently and performs well in short quarterly samples. In the ETS framework, the observed time series y_t is decomposed additively or multiplicatively into these components. The general state space formulation can be expressed as [Equations (1) and (2)] (R. J. Hyndman & Athanasopoulos, 2021, Section 8.1):

$$(\text{forecasting equation}) \quad y_t = l_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t, \quad (1)$$

$$(\text{smoothing equation}) \quad l_t = l_{t-1} + b_{t-1} + \alpha \varepsilon_t, \quad (2)$$

$$\text{with } b_t = b_{t-1} + \beta \varepsilon_t,$$

$$\text{and } s_t = s_{t-1} + \gamma \varepsilon_t$$

where l_t , b_t , and s_t are the level, trend, and seasonal components at time t ; α , β , and γ are smoothing parameters; and ε_t is the random error term. Depending on the data structure, ETS models may employ additive (A), multiplicative (M), or damped trend specifications, leading to a taxonomy of 30 ETS variants (R. Hyndman et al., 2008).

To model potential nonlinearities, we also employ neural network autoregressive models (NNAR), which combine autoregressive structures with feedforward neural networks. Moreover, NNAR helps capture mild nonlinear patterns while controlling model complexity in limited samples. The $NNAR(p, P, k)_m$ model can be formalized as [Equation (3)] (R. J. Hyndman & Athanasopoulos, 2021, Section 12.4):

$$\hat{y}_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}, y_{t-m}, y_{t-2m}, \dots, y_{t-Pm}; \theta) + \varepsilon_t \quad (3)$$

where $f(\cdot)$ denotes a neural network function with parameters θ , p is the number of non-seasonal lags, P is the number of seasonal lags, m is the seasonal frequency (typically 4 for quarterly data), and k is the number of neurons in the hidden layer. The model captures both short-term dynamics and seasonal effects through nonlinear mappings.

For linear seasonal structures, we employ the Seasonal ARIMA (SARIMA) model, an extension of the ARIMA framework that incorporates seasonal components. Specifically, the addition of seasonal parameters to the autoregressive (AR), differencing (I), and moving average (MA) terms allows the model to effectively capture a wide range of seasonal patterns commonly observed in real-world economic data (Perone, 2021). The SARIMA model, denoted as $ARIMA(p, d, q)(P, D, Q)_m$ is represented as [Equation (4)] (Perone, 2022):

$$\Phi_P(B^m)\phi_P(B)(1-B)^d(1-B^m)^D y_t = \Theta_Q(B^m)\theta_Q(B)\varepsilon_t \quad (4)$$

where B is the backshift operator, $\Phi_P(B)$ and $\theta_Q(B)$ are the non-seasonal autoregressive and moving average polynomials, and $\Phi_P(B^m)$, $\Theta_Q(B^m)$ are the seasonal autoregressive and moving average polynomials with seasonal frequency m (equal to 4 for quarterly data). SARIMA complements ETS by accommodating integration and seasonal differencing with parsimonious, interpretable parameters. The differencing operators $(1-B)^d$ and $(1-B^m)^D$ remove trend and seasonality, respectively.

The TBATS model, introduced to handle complex seasonality and nonlinearity, combines Box–Cox transformation, ARMA errors, trend, and multiple seasonal components. Formally, it can be written as [Equations (5) and (6)] (De Livera et al., 2011):

$$y_t^{(\omega)} = l_{t-1} + \phi b_{t-1} + \sum_{j=1}^J s_{j,t-m_j} + d_t + \varepsilon_t \quad (5)$$

$$d_t = \sum_{k=1}^K \zeta_k \cos\left(\frac{2\pi kt}{m_j}\right) + \eta k \sin\left(\frac{2\pi kt}{m_j}\right) \quad (6)$$

where $y_t^{(\omega)}$ denotes the Box–Cox transformed series, l_t and b_t represent local level and trend, ϕ is the damping parameter, and d_t captures trigonometric seasonality. TBATS is particularly well suited for irregular seasonality and nonlinearity in short time series.

To capitalize on the individual strengths of these models, we implement hybrid ensemble models. These models linearly combine forecasts from selected base models. The general hybrid formulation is given by [Equation (7)]:

$$\hat{y}_t^{hybrid} = \sum_{i=1}^n w_i \hat{y}_t^{(i)}, \text{ where } \sum_{i=1}^n w_i = 1, w_i \geq 0 \quad (7)$$

with $\hat{y}_t^{(i)}$ being the forecast from model i , and w_i the weight assigned to that model. The minimization of forecasting error metrics indicates that equal weighting of the component models results in near-optimal performance. Therefore, the hybrid models use equal weights for each component model, reflecting the balanced contribution each makes to increasing overall forecast accuracy. All models are trained on quarterly GDP data from 1Q-2010 to 2Q-2024, with separate series for Armenia, Azerbaijan, and Georgia, adjusted for

seasonal effects and expressed in local currency at constant or current prices. Forecasting accuracy is assessed through rolling-origin cross-validation, using performance metrics including the mean absolute error (MAE), mean absolute percentage error (MAPE), root mean squared error (RMSE), and mean absolute scaled error (MASE). The formulas for each metric are provided below [Equations (8)–(11)]:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (8)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} * 100\% \quad (9)$$

$$\text{MASE} = \frac{1}{n} \sum_{i=1}^n \left(\frac{|y_i - \hat{y}_i|}{\frac{1}{n-1} \sum_{i=2}^n |y_i - \hat{y}_i - 1|} \right) \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (11)$$

where n indicates the number of observations, y_i displays the actual values, and \hat{y}_i identifies the predicted values (Perone, 2022).

3.2. Cross-Validation

A critical phase in the forecasting evaluation process involves assessing the predictive capability of the selected models for each country. We employ cross-validation, a widely used technique for evaluating forecasting model performance and robustness by partitioning the data into training and test sets (Allgaier & Pryss, 2024). This approach, known as out-of-sample evaluation in forecasting, can be implemented in two main ways: fixed-origin evaluation and rolling-origin evaluation (Hewamalage et al., 2023). We adopt a rolling-origin configuration, a prevalent resampling method that preserves the temporal order of observations and generally provides more reliable performance estimates in non-stationary, real-world contexts (Cerqueira et al., 2020). Moreover, this method is especially suited for relatively short time series, as is the case in our analysis (Tashman, 2000; Hewamalage et al., 2023).

In this framework, models are trained exclusively on historical data, and their forecast accuracy is evaluated on future, unseen data excluded from the training set. The training window gradually increases, and the forecast origin moves forward at each iteration (rolling origin). Specifically, the first training period runs from 1Q-2010 to 4Q-2019, and the test sets are made up of single quarters, simulating a realistic forecasting scenario. Specifically, the training set grows by one quarter per iteration, starting in 1Q-2020 and ending in 2Q-2024. This approach generates 18 one-observation test sets until 2Q-2024. This procedure emulates the iterative update process inherent to real-time forecasting systems, and assures a robust and reliable assessment of models' forecasting capabilities by considering a significant time window of more than four years.

Forecast accuracy is measured by averaging the error metrics across all test sets (R. J. Hyndman & Athanasopoulos, 2021, Section 5.10). We compute the average Mean Absolute Percentage Error (MAPE) between the observed values in each test set and the corresponding forecasts generated for each quarterly horizon.

4. Data

The data used to carry out the predictions for each country are the following:

- the quarterly GDP of Armenia at average prices of the previous year, expressed in million drams, from 1Q-2013 to 2Q-2024 (The Statistical Committee of the Republic of Armenia, 2025);

- the quarterly GDP of Azerbaijan at 2015 prices, expressed in million manats, from 1Q-2010 to 2Q-2024 ([The State Statistical Committee of the Republic of Azerbaijan, 2025](#));
- and the quarterly GDP of Georgia at current (market) prices, expressed in million Georgian laris (GEL), from 1Q-2010 to 2Q-2024 ([National Statistics Office of Georgia, 2025](#)).

Specifically, Figure 1 shows the historical GDP series for Armenia, Azerbaijan, and Georgia from 1Q-2010 to 2Q-2024. While Armenia and Georgia display a sinusoidal and consistent trend over time, Azerbaijan's GDP exhibits a more erratic and discontinuous pattern, alternating between periods of strong growth and significant slowdown. Notably, all three countries experienced a contraction in the first quarter of 2020, followed by a quarter of moderate GDP stagnation, showing a remarkable resilience to the COVID-19 outbreak.



Figure 1. The historical trend of GDP in the South Caucasian countries from 1Q-2010 to 2Q-2024.

5. Results

This section introduces the model estimates and presents the results in two steps: in-sample accuracy (Section 5.1) and out-of-sample cross-validation (Section 5.2). We first summarize model selection and fit, then report forecast accuracy and validation. Unless

noted, in-sample performance refers to models fitted on the full sample, while out-of-sample performance uses rolling-origin evaluation.

5.1. Forecast Accuracy

In this section, we present the GDP forecasting models for Armenia, Azerbaijan, and Georgia. Individual models were selected using the ‘forecast’ package in R software (R. J. Hyndman & Khandakar, 2008), while hybrid models were constructed via ensemble methods utilizing the ‘forecastHybrid’ package (Shaub & Ellis, 2020). The detailed configurations of the individual GDP forecasting models for the three countries are presented in Table 2. The optimal ETS model for Armenia was identified as (M,M,M), indicating multiplicative components for error, trend, and seasonality. Conversely, the best-fitting models for Azerbaijan and Georgia followed the (M,A,M) specification, implying an additive trend alongside multiplicative error and seasonal components. Regarding the NNAR models, all three countries shared the same architecture: one non-seasonal lagged input, one seasonal lag for the autoregressive term, and two nodes in the hidden layer, suggesting a common underlying temporal structure within the Caucasus region.

Table 2. The parameters for the individual GDP forecasting models for Armenia, Azerbaijan, and Georgia.

Models	Armenia	Azerbaijan	Georgia
ETS	(M,M,M)	(M,A,M)	(M,A,M)
NNAR	(1,1,2) ₄	(1,1,2) ₄	(1,1,2) ₄
SARIMA	(0,1,0)(0,1,1) ₄	(2,0,0)(3,1,0) ₄ with drift	(0,1,1)(2,1,0) ₄
TBATS	(0.502,{4,0},-,{<4,1>})	(0.023,{0,0},1,{<4,1>})	(0.077,{0,0},1,{<4,1>})

The SARIMA models varied across countries, reflecting tailored adjustments to country-specific data characteristics. Armenia’s best-fitting model was specified as (0,1,0)(0,1,1), Azerbaijan’s as (2,0,0)(3,1,0) with drift, and Georgia’s as (0,1,1)(2,1,0), each capturing distinct patterns in trend and seasonality. While the TBATS models exhibited minor differences in parameter estimates, they shared notable similarities. Armenia’s model employed an intermediate Box-Cox transformation of 0.5, whereas Azerbaijan’s and Georgia’s transformations approximated logarithmic transformations, with values of 0.02 and 0.08, respectively. None of the models incorporated a damping trend. Furthermore, all three TBATS models shared identical seasonal component structures, reflecting consistent seasonal frequency and modeling flexibility across countries.

Table 3 summarizes the performance of individual models and their feasible hybrid combinations on the whole series, from 1Q-2010 to 2Q-2024.¹ ACF1 reports the lag-1 autocorrelation of forecast errors. Values close to zero indicate residual independence; large positive or negative values suggest remaining structure in the errors and motivate model refinement. While Figures 2–4 depict visually the predictions for the period 3Q-2024 to 4Q-2026, which were estimated with 95% confidence intervals. Hybrid models are built with equal weights across component models, as suggested by forecasting error metrics minimization.

All fitted models achieved MASE values significantly below 1, indicating that these models outperform the naïve baseline forecast in accuracy (R. J. Hyndman & Koehler, 2006), thereby justifying the adoption of more complex forecasting techniques (Table 3). Specifically, for Armenia, the best-performing models in terms of forecast accuracy were the hybrid ETS-NNAR-TBATS and NNAR-TBATS models, with MAPEs of 3.24% and 3.25%, respectively. According to Makridakis and Wheelwright (1998), these values correspond to a high level of forecast accuracy. For Azerbaijan, the hybrids ETS-NNAR-SARIMA and ETS-NNAR demonstrated superior performance, with MAPEs of 5.27% and 5.40%, respectively, also indicative of high accuracy. For Georgia, the leading models were hybrid

ETS-SARIMA and ETS alone, with MAPEs of 2.66% and 2.78%, respectively. Consequently, the models for Georgia exhibit the highest forecast accuracy, with deviations from actual values as low as 2.66% in the best case. Specifically, the forecasts indicate that Georgia is expected to grow at a faster pace than Armenia and Azerbaijan, although the precise magnitude of this gap should be interpreted with caution, given data limitations.

Table 3. Performance of single and hybrid models for predicting GDP in Armenia, Azerbaijan, and Georgia.

Models	RMSE	MAE	MAPE	MASE	ACF1
Armenia					
ETS	76,053.37	52,243.16	3.3663	0.3661	0.0809
NNAR	95,504.71	66,465.49	4.145	0.4658	0.2469
TBATS	85,168.42	65,544.08	4.4943	0.1723	−0.0457
SARIMA	92,389.99	61,973.81	3.9364	0.4343	−0.224
Hybrid ES	78,038.76	53,243.5	3.3593	0.3712	−0.1
Hybrid EN	80,657.43	54,524.17	3.405	0.3821	0.1825
Hybrid ET	76,038.35	55,285.48	3.6549	0.3874	0.0313
Hybrid NS	86,336.48	59,009.18	3.6777	0.4135	−0.0254
Hybrid NT [2]	77,163.67	51,744.49	3.254	0.3626	0.1242
Hybrid ST	81,683.1	60,646.37	4.0117	0.425	−0.1218
Hybrid ENS	79,747.4	54,399.15	3.383	0.3812	0.0092
Hybrid ENT [1]	75,387.76	51,574.67	3.2356	0.3614	0.1128
Hybrid EST	75,478.1	54,219.91	3.5218	0.38	−0.0669
Hybrid NST	79,198.45	54,975.99	3.4615	0.3853	−0.0209
Hybrid SENT	76,772.39	53,192.27	3.3249	0.3728	0.0049
Azerbaijan					
ETS	1388.3	1005.48	5.4094	0.3925	0.1837
NNAR	1464.66	1192.99	6.6422	0.4657	0.1111
TBATS	1799.59	1410.51	7.6692	0.8366	0.1108
SARIMA	1410.98	1081.68	5.7723	0.4222	−0.0052
Hybrid ES	1346.01	997.26	5.3372	0.3893	0.1058
Hybrid ET	1496.58	1122.58	6.0807	0.4382	0.273
Hybrid EN [2]	1257.67	975.98	5.4047	0.381	0.2194
Hybrid NS	1290.13	1002.78	5.5279	0.3914	0.0772
Hybrid NT	1522.32	1190.72	6.4748	0.4648	0.0741
Hybrid ST	1472	1090.45	5.8955	0.4256	0.1485
Hybrid ENS [1]	1265.57	965.84	5.267	0.377	0.1469
Hybrid ENT	1394.78	1067.59	5.7895	0.4167	0.2105
Hybrid EST	1403.03	1023.8	5.5169	0.3996	0.2137
Hybrid NST	1393.46	1074.02	5.8408	0.4192	0.1128
Hybrid ENST	1353.01	1031.39	5.5819	0.4026	0.185
Georgia					
ETS [2]	885.64	520.89	2.7801	0.2441	0.0273
NNAR	1188.86	754.8	3.8325	0.3538	0.1521
TBATS	1399.75	1112.8	6.0662	0.4436	−0.2318
SARIMA	1058.86	611.5227	3.0789	0.2866	−0.0071
Hybrid ES [1]	930.03	510.45	2.6585	0.2393	0.0449
Hybrid ET	1005.25	696.45	3.8176	0.3264	0.0192
Hybrid EN	965.19	583.8	2.9778	0.2736	0.0786
Hybrid NS	1073.58	643.08	3.2586	0.3014	0.0358
Hybrid NT	1195.02	836.95	4.2162	0.3923	0.0087
Hybrid ENS	977.12	566.95	2.8921	0.2657	0.0458
Hybrid ENT	1020.98	655.52	3.3411	0.3072	0.072
Hybrid EST	983.59	636.95	3.3741	0.2985	0.0195
Hybrid NST	1107.87	722.97	3.6261	0.3389	−0.0047
Hybrid ENST	1011.97	633.12	3.2011	0.2967	0.0443

Notes: Best models for each country are made bold.

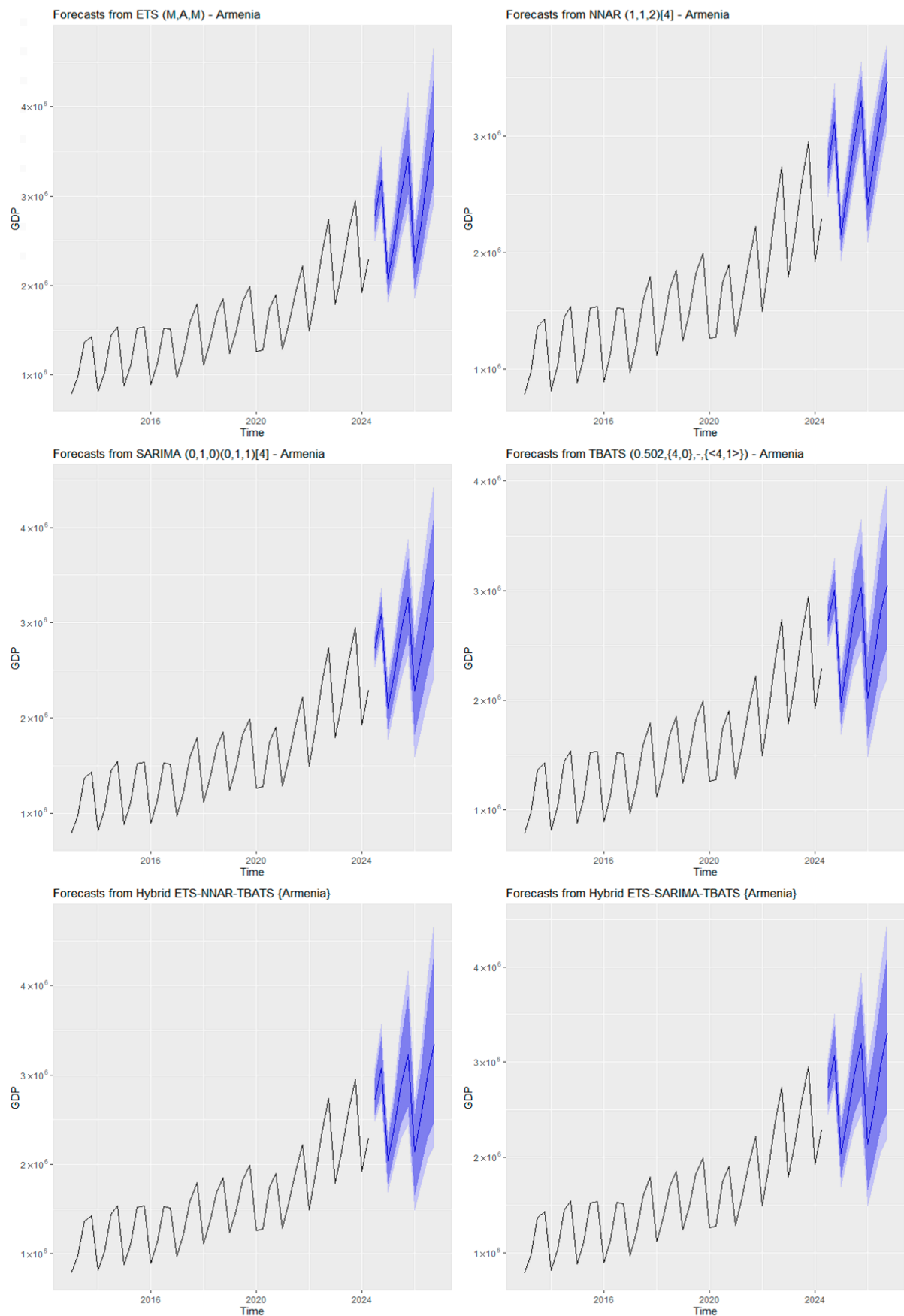


Figure 2. Forecasted GDP values from individual models and the two best-performing hybrid models for Armenia (from 3Q-2024 to 4Q-2026).

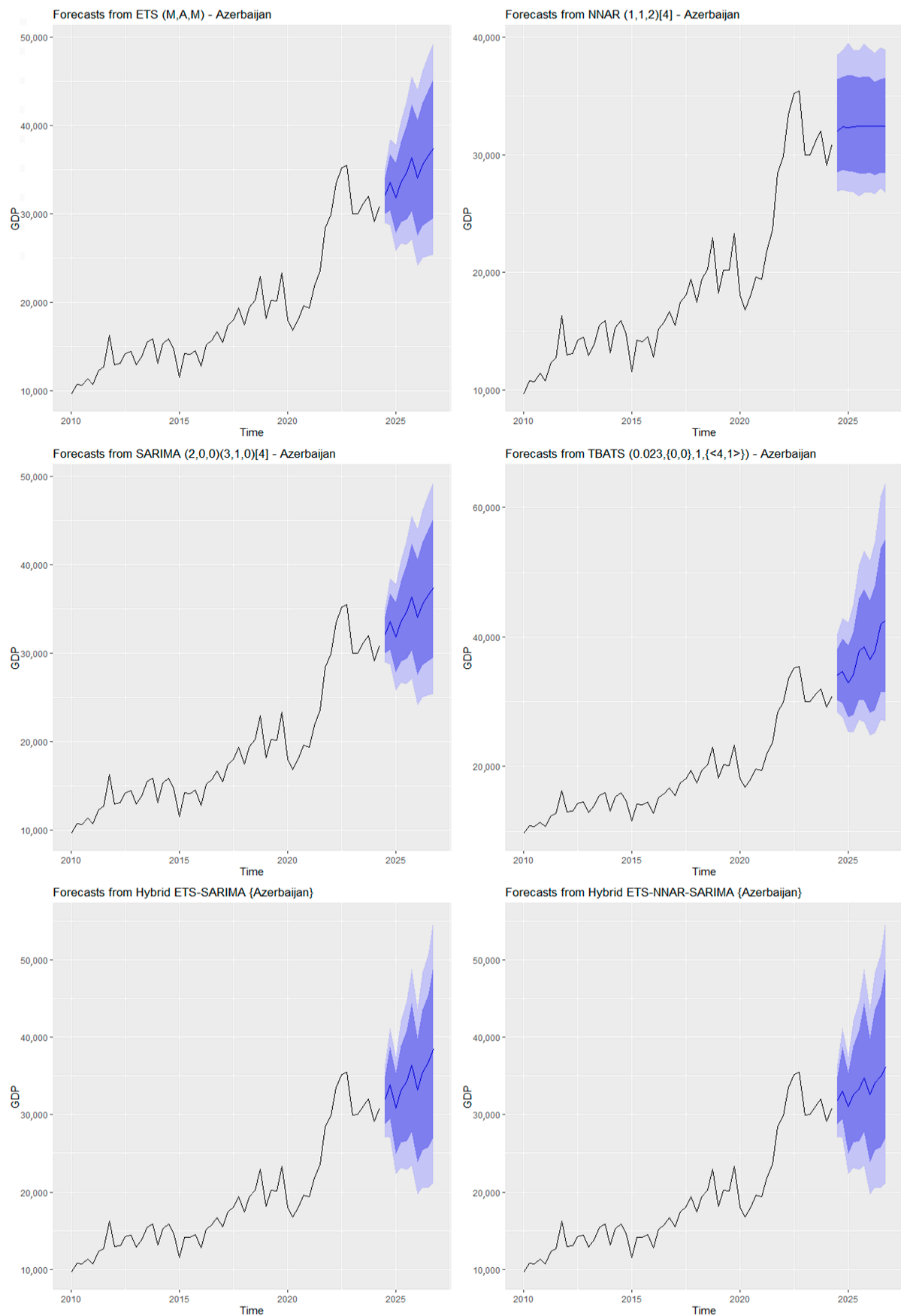


Figure 3. Forecasted GDP values from individual models and the two best-performing hybrid models for Azerbaijan (from 3Q-2024 to 4Q-2026).

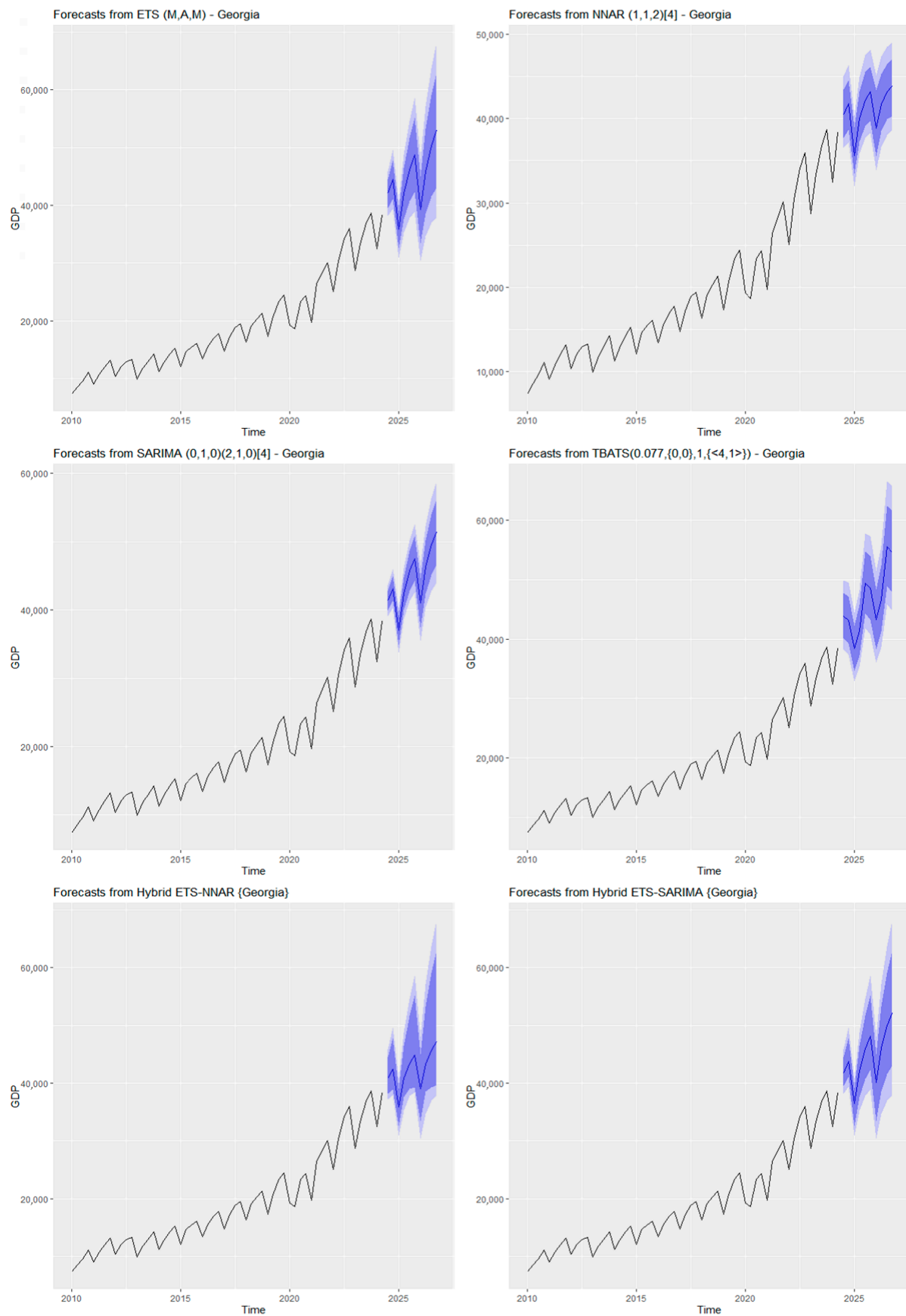


Figure 4. Forecasted GDP values from individual models and the two best-performing hybrid models for Georgia (from 3Q-2024 to 4Q-2026).

This paper aligns with recent literature indicating that hybrid models constitute an effective approach to capturing the linear, nonlinear, and seasonal dynamics inherent in GDP time series (Maccarrone et al., 2021; Richardson et al., 2021; Yang et al., 2022; Kumar & Yadav, 2023; Atif, 2024; Jallow et al., 2025). Notably, the consistent inclusion of ETS models within the highest-performing hybrid frameworks corroborates prior findings emphasizing the model's efficacy in capturing deterministic components across diverse economic contexts.

In the macroeconomic forecasting domain, Sabri et al. (2023) confirmed that ETS-based models offer a simple yet effective approach, particularly suitable for small datasets and sufficiently robust for noisy time series. This aligns with the context of our analysis, which is based on a relatively short series of GDP. As a macroeconomic variable, GDP is often subject to irregular fluctuations caused by exogenous shocks, such as unexpected policy decisions or, as in this case, the COVID-19 pandemic.

Overall, the findings confirm that hybrid ensembles consistently outperform individual models in the South Caucasian countries, though with notable differences across them. The improvements are most evident in Azerbaijan, possibly reflecting its higher historical GDP volatility, while Georgia attains the highest forecast accuracy, which is consistent with its more regular growth patterns. In Armenia, the performance gains are less pronounced, with accuracy levels positioned between those observed for Azerbaijan and Georgia.

These differences also explain why different hybrid combinations emerge as optimal in each country. The hybrid ETS-SARIMA better captures Georgia's stable trend and seasonal patterns, the hybrid ETS-NNAR is well suited to Azerbaijan's irregular dynamics, and the inclusion of TBATS in Armenia adds flexibility for shorter and more complex time series.

5.2. Out-of-Sample Cross-Validation

The evaluation of forecasting in Section 5.1 involved assessing model performance by fitting each model to the complete dataset and measuring accuracy on the corresponding observations. This in-sample method aids in identifying optimal model parameters; however, it does not safeguard against overfitting. Such issues may occur when the model identifies noise or transient fluctuations instead of the genuine underlying patterns, thereby compromising its capacity to generalize and produce reliable forecasts of future GDP values (Tashman, 2000; Staněk, 2023).

To address this limitation, we split the dataset into training and test sets and assessed the out-of-sample model's forecasting performance on unseen data that were excluded during the previous model fitting process. This ensures a more accurate evaluation of the predictive performance of models because the test data are not used for the forecasts (R. J. Hyndman & Athanasopoulos, 2021, Section 5.8). Within this framework, a cross-validation procedure is implemented utilizing a rolling forecasting origin. The test set comprises a single observation, while the training set includes all preceding observations utilized to generate a 1-step-ahead forecast (R. J. Hyndman & Athanasopoulos, 2021, Section 5.10). The initial training set specifically encompasses the period from 1Q-2020 to 4Q-2019, with a progressive increase of one observation at each step. As a result, we derive 18 test sets, each comprising a single observation from 1Q-2020 to 2Q-2024.

Table 4 reports the results obtained from the out-of-sample cross-validation procedure conducted for Armenia, Azerbaijan, and Georgia, highlighting the average forecasting accuracy of the selected models across multiple horizons to assess their sensitivity to specific shocks. Specifically, we computed the average MAPE between forecasted and actual values for four distinct periods: the COVID-19 recession (entire 2020), the energy shock caused by the Russia–Ukraine war (entire 2022), the most recent eight quarters (from 3Q-2022 to 2Q-2024), and the full evaluation period (from 1Q-2020 to 2Q-2024).

Table 4. Forecasted values derived from out-of-sample cross-validation for Armenia, Azerbaijan, and Georgia.

Models	Armenia			Azerbaijan			Georgia		
	ENT	NT	Actual	ENS	EN	Actual	ES	E	Actual
1Q-2020	1,365,750	1,382,919	1,263,058.8	19,377.7	19,254	18,043.6	20,405.7	20,012.1	19,305.5
2Q-2020	1,574,390	1,596,964	1,274,458.7	20,671.7	20,599	16,813.2	22,484.4	22,517.9	18,694.2
3Q-2020	1,771,272	1,773,823	1,743,630.8	17,581.8	17,544.1	18,103.1	19,938.7	20,174.4	23,428
4Q-2020	1,880,519	1,895,748	1,900,754.3	19,956.1	19,028.9	19,618.2	24,962.6	24,076.1	24,301.8
1Q-2021	1,278,545	1,295,823	1,282,290.1	17,490.1	18,640.9	19,393.4	19,551	19,329.4	19,739.9
2Q-2021	1,438,366	1,419,950	1,577,120.1	19,666.1	19,539.2	21,820.5	20,911.2	22,354.8	26,364.3
3Q-2021	1,890,049	1,872,546	1,909,558.9	21,370.7	21,127.4	23,571	27,429.8	27,264.1	28,413
4Q-2021	2,015,887	2,012,124	2,222,808.7	23,877.4	23,668	28,418.3	28,708.2	29,764.3	30,112.3
1Q-2022	1,441,316	1,407,310	1,493,436.3	23,861.5	23,411.6	29,881.5	24,652.6	24,078.1	25,089.6
2Q-2022	1,702,447	1,698,711	1,899,612.1	31,225.9	31,680.1	33,441.6	26,246.5	28,962.9	30,430.7
3Q-2022	2,139,738	2,105,522	2,373,227.4	34,523.4	34,644.9	35,212.3	33,251.2	33,100.8	34,120.3
4Q-2022	2,779,721	2,944,221	2,735,173.6	37,537.1	37,201.1	35,437.3	35,641.5	36,002.6	35,940.6
1Q-2023	2,032,655	2,043,999	1,789,887.9	34,249.3	33,197.2	29,977.3	30,004.6	29,029.7	28,720
2Q-2023	2,132,951	2,153,862	2,138,726.2	31,790.1	31,896.4	30,005.1	32,606.8	33,823.1	33,389.5
3Q-2023	2,581,360	2,572,687	2,575,328.7	28,262.9	27,395.1	31,152.3	36,846.9	36,862.3	36,807.9
4Q-2023	2,944,673	2,946,939	2,949,232.2	32,750.4	32,752	31,993.7	38,644.5	39,022.8	38,688.5
1Q-2024	2,016,150	2,075,621	1,922,772.8	29,041	29,955.8	29,096.8	31,853.3	30,768.7	32,438.8
2Q-2024	2,310,898	2,317,256	2,289,276.2	30,631.7	31,206.2	30,786.3	37,915	37,748.4	38,356.5
COVID-19	7.3%	7.71%	-	7.55%	7.74%	-	10.6%	9.39%	-
Russia-Ukraine	6.93%	9.44%	-	9.98%	9.89%	-	5.29%	3.13%	-
Last 8 quarters	3.84%	5.21%	-	4.86%	5.28%	-	1.67%	1.71%	-
Total	5.35%	6.39%	-	8.44%	8.54%	-	6.05%	4.77%	-

We specifically trained the models that exhibited superior performance in the in-sample evaluation described in Section 5.1. The models with the highest forecast accuracy were ETS and hybrid ETS-SARIMA for Georgia, which achieved average MAPE values of 4.77% and 6.05%, respectively, over the full period. During the COVID-19 shock, the MAPE rose to approximately 10%, while during the war-induced energy shock, it decreased to a value between 3.13% and 5.29%. For Azerbaijan, the hybrid models ETS-NNAR-SARIMA and ETS-NNAR yielded average MAPE values around 8% over the full period, slightly below 8% during the COVID-19 shock, and just under 10% during the war-induced energy shock. In the case of Armenia, the hybrid models ETS-NNAR-TBATS and NNAR-TBATS outperformed those used for Azerbaijan, with average MAPE values of 5.35% and 6.39%, rising to 7.30% and 7.71% during the COVID-19 shock, and to 6.93% and 9.44% during the war-related energy shock, respectively.

Table 5 further reports the gain or loss in forecasting performance during each shock period relative to the entire evaluation window. For all three countries, the lowest MAPE values were recorded during the most recent eight quarters, which is consistent with the short- to medium-term forecasting capability of the models under investigation. The COVID-19 shock resulted in an absolute accuracy loss of less than 2% for Armenia and approximately 4% for Georgia, whereas Azerbaijan showed a modest gain of just under 1%. Regarding the Russia-Ukraine conflict, the analysis revealed an accuracy gain ranging from 0.76% to 1.64% for Georgia, a slight accuracy loss (just over 1%) for Azerbaijan, and a loss between 1.58% and 3.05% for Armenia.

These findings confirm that our hybrid ensemble models not only outperform individual models under standard conditions but also maintain forecasting accuracy and stability during turbulent periods such as the COVID-19 recession and the Russia-Ukraine conflict. This reinforces their practical applicability for real-world forecasting under uncertainty.²

Table 5. Absolute difference in MAPE between each shock period and the full evaluation window (1Q-2020 to 2Q-2024).

Models	ENT	NT	ENS	EN	ES	E
COVID-19	1.95%	1.32%	−0.89%	−0.8%	4.55%	4.62%
Russia-Ukraine	1.58%	3.05%	1.54%	1.35%	−0.76%	−1.64%
Last 8 quarters	−1.51%	−1.18%	−3.58%	−3.26%	−4.38%	−3.06%

**Figure 5.** Out-of-sample cross-validation using a rolling forecasting origin for Armenia, Azerbaijan, and Georgia.

Finally, the superior out-of-sample forecast performance of hybrid ETS-based models for the South Caucasian countries is consistent with a growing body of literature that focuses on medium- and long-term forecasting of GDP in industrialized and developing countries (Botha et al., 2021; Dritsaki & Dritsaki, 2023; Almarashi et al., 2024; Perone, 2024; Elbatal et al., 2025).

6. Conclusions

This paper has investigated the forecasting performance of a range of univariate time series models for the quarterly GDP dynamics of Armenia, Azerbaijan, and Georgia over the period 2010–2024. Specifically, the analysis employed ETS, NNAR, SARIMA, and TBATS models, along with their feasible hybrid combinations. Forecast accuracy was evaluated through both in-sample performance and a more robust rolling-origin cross-validation procedure. Findings were interpreted using a comprehensive set of error metrics, including MAE, MASE, MAPE, and RMSE.

The results showed that Georgia is expected to experience the highest GDP growth over the next two and a half years, followed by Armenia and Azerbaijan. Moreover, the analysis indicated that hybrid models consistently outperform single-model alternatives, with ETS-based ensembles among the most accurate. In particular, the ETS-SARIMA hybrid provided the most accurate forecasts for Georgia, while the hybrid ETS-NNAR-SARIMA proved to be optimal for Azerbaijan. For Armenia, the most effective configurations combined ETS, NNAR, and TBATS. Thus, hybrid models proved to be useful for dealing with nonlinearities, structural fractures, and short or fragmented time series. These challenges are particularly common in developing and transitional countries, where data scarcity and economic volatility can reduce the effectiveness of traditional forecasting methodologies.

Furthermore, the rolling-origin cross-validation approach confirms the robustness of the selected models in real-world forecasting scenarios, reinforcing the importance of out-of-sample evaluation for macroeconomic policy and planning. These findings support the use of ETS-based hybrids as a practical and reliable solution for GDP forecasting in data-constrained environments. This aligns with Sabri et al. (2023) and is further reinforced by recent International Monetary Fund estimates (2025), which highlight the region's strong post-pandemic economic recovery. This study provides a country-comparable assessment of feasible ETS-NNAR-SARIMA-TBATS hybrids for South Caucasian GDP under a unified rolling-origin design with 18 test sets and reported prediction intervals. In practical terms, the short-term projections can inform budget setting, debt issuance timing, and contingency planning by ministries of finance and central banks; international organizations can use them to sequence program disbursements and monitor near-term growth risks; firms and investors may apply them to demand planning and inventory management; and statistical offices can treat the models as a baseline for routine nowcasting. At the same time, the evidence rests on relatively short histories for some series (notably Armenia), a univariate setup without external predictors, and equal-weight combinations; repeated out-of-sample checks across countries indicate that the main patterns remain stable under these conditions. Inevitably, the exclusive reliance on univariate GDP models constrains economic interpretability. However, this design is substantially consistent with the forecasting purpose of the study and is further supported by robustness checks that remain valid even under external shocks. Future research can add exogenous regressors (e.g., trade, exchange rates, inflation). It should also test the scalability of the hybrid framework to multi-country nowcasting and high-frequency settings. One avenue is mixed-frequency designs using timely monthly indicators. Another is data-driven combination weights via cross-validation or stacking. Further steps include evaluating density forecasts and calibration, such as the continuous ranked probability score (CRPS) and the probability

integral transform (PIT) and not only point errors. Models can allow structural breaks and time-varying parameters. Regularized learners suited to short samples should be tested with proper time-series cross-validation. Finally, assessments with real-time data vintages and revisions are needed.

Beyond methodological implications, these findings offer a reliable tool for national and international policy institutions in the South Caucasus, a region of growing geopolitical and energy strategic relevance (Pkhaldze, 2025). In such contexts, short-term GDP forecasts are essential to guide investment planning, labor market reforms, and preparedness strategies against external shocks.

Author Contributions: Conceptualization, G.P.; methodology, G.P. and M.A.Z.-M.; software, G.P.; formal analysis, G.P.; data curation, G.P.; writing—original draft preparation, G.P. and M.A.Z.-M., writing review and editing, G.P. and M.A.Z.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: All data used in this investigation are publicly available and appropriately acknowledged in the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Notes

- ¹ It should be noted that hybrid models are built using equal weights across components. This choice yields lower forecast errors compared to alternative weighting schemes and is consistent with the literature, which highlights the robustness of equal-weight combinations and their role as a natural benchmark in forecasting, particularly when sample sizes are limited, as in this case (Timmermann, 2006; Claeskens et al., 2016).
- ² In addition, Figure 5 offers a visual representation of the forecast versus actual values for the last eight quarters, allowing for an intuitive assessment of model performance and predictive alignment over the recent time.

References

- Abdić, A., Resić, E., Abdić, A., & Rovčanin, A. (2020). Nowcasting GDP of Bosnia and Herzegovina: A comparison of forecast accuracy models. *The South East European Journal of Economics and Business*, 15(2), 1–14. [\[CrossRef\]](#)
- Alaminos, D., Salas, M. B., & Fernández-Gámez, M. A. (2022). Quantum computing and deep learning methods for GDP growth forecasting. *Computer Economics*, 59, 803–829. [\[CrossRef\]](#)
- Allgaier, J., & Pryss, R. (2024). Cross-validation visualized: A narrative guide to advanced methods. *Machine Learning and Knowledge Extraction*, 6(2), 1378–1388. [\[CrossRef\]](#)
- Almarashi, A. M., Daniyal, M., & Jamal, F. (2024). Modelling the GDP of KSA using linear and non-linear NNAR and hybrid stochastic time series models. *PLoS ONE*, 19(2), e0297180. [\[CrossRef\]](#)
- Atif, D. (2024). Enhancing long-term GDP forecasting with advanced hybrid models: A comparative study of ARIMA-LSTM and ARIMA-TCN with dense regression. *Computational Economics*, 65, 3447–3473. [\[CrossRef\]](#)
- Babii, A., Ghysels, E., & Striaukas, J. (2024). Econometrics of machine learning methods in economic forecasting. In M. P. Clements, & A. B. Galvão (Eds.), *Handbook of research methods and applications in macroeconomic forecasting* (Chapter 10, pp. 246–273). Edward Elgar Publishing.
- Ballarin, G., Dellaportas, P., Grigoryeva, L., Hirt, M., Van Huellen, S., & Ortega, J. (2024). Reservoir computing for macroeconomic forecasting with mixed-frequency data. *International Journal of Forecasting*, 40(3), 1206–1237. [\[CrossRef\]](#)
- Botha, B., Reid, G., Olds, T., Steenkamp, D., & van Jaarsveld, R. (2021). Nowcasting South African gross domestic product using a suite of statistical models. *South African Journal of Economics*, 89(4), 526–554. [\[CrossRef\]](#)
- Bragoli, D., & Fosten, J. (2018). Nowcasting Indian GDP. *Oxford Bulletin of Economics and Statistics*, 80(2), 259–282. [\[CrossRef\]](#)
- Cepni, O., Guney, I. E., & Swanson, N. R. (2019). Forecasting and nowcasting emerging market GDP growth rates: The role of latent global economic policy uncertainty and macroeconomic data surprise factors. *Journal of Forecasting*, 39(1), 18–36. [\[CrossRef\]](#)
- Cerqueira, V., Torgo, L., & Mozetič, I. (2020). Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning*, 109(11), 1997–2028. [\[CrossRef\]](#)
- Chernis, T., & Sekkel, R. (2017). A dynamic factor model for nowcasting Canadian GDP growth. *Empirical Economics*, 53, 217–234. [\[CrossRef\]](#)

- Cicceri, G., Inserra, G., & Limosani, M. (2020). A machine learning approach to forecast economic recessions—An Italian case study. *Mathematics*, 8(2), 241. [CrossRef]
- Claeskens, G., Magnus, J. R., Vasnev, A. L., & Wang, W. (2016). The forecast combination puzzle: A simple theoretical explanation. *International Journal of Forecasting*, 32(3), 754–762. [CrossRef]
- Claudio, J. C., Heinisch, K., & Holtemöller, O. (2020). Nowcasting East German GDP growth: A MIDAS approach. *Empirical Economics*, 58(1), 29–54. [CrossRef]
- Darvas, Z. (2011). Beyond the crisis: Prospects for emerging Europe. *Comparative Economic Studies*, 53, 261–290. [CrossRef]
- De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496), 1513–1527. [CrossRef]
- Domit, S., Monti, F., & Sokol, A. (2019). Forecasting the UK economy with a medium-scale Bayesian VAR. *International Journal of Forecasting*, 35(4), 1669–1678. [CrossRef]
- Dritsaki, M., & Dritsaki, C. (2023). Modelling and forecasting GDP of Greece with a modified exponential smoothing state-space framework. In *Advances in empirical economic research* (pp. 89–110). Springer.
- Elbatal, I., Sarwar, M., Jamal, F., Daniyal, M., Hussain, Z., & Ben Ghorbal, A. (2025). Modelling on gross domestic product Annual growth rate data by using time series, machine learning, and probability models. *Journal of Radiation Research and Applied Sciences*, 18(2), 101481. [CrossRef]
- Fan, L. (2024, April 19–21). *Long-term forecast of regional economy based on least squares support vector machine*. 2024 International Academic Conference on Edge Computing, Parallel and Distributed Computing (ECPDC 2024), Xi'an, China. [CrossRef]
- Gerasimov, A. N., Gromov, Y. I., & Gulay, T. A. (2015). *Forecasting the indicators of socioeconomic development of the North Caucasus Federal District*. Stavropol State Agrarian University.
- Gu, Y., Shao, Z., Huang, X., & Cai, B. (2022). GDP forecasting model for China's provinces using nighttime light remote sensing data. *Remote Sensing*, 14(15), 3671. [CrossRef]
- Habib, M. M., Mileva, E., & Stracca, L. (2017). The real exchange rate and economic growth: Revisiting the case using external instruments. *Journal of International Money and Finance*, 73, 386–398. [CrossRef]
- Hewamalage, H., Ackermann, K., & Bergmeir, C. (2023). Forecast evaluation for data scientists: Common pitfalls and best practices. *Data Mining and Knowledge Discovery*, 37(2), 788–832. [CrossRef]
- Higgins, P., Zha, T., & Zhong, W. (2016). Forecasting China's economic growth and inflation. *China Economic Review*, 41, 46–61. [CrossRef]
- Hyndman, R., Koehler, A., Ord, K., & Snyder, R. (2008). *Forecasting with exponential smoothing: The state space approach*. Springer.
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts. Available online: <https://otexts.com/fpp3/> (accessed on 17 May 2025).
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast Package for R. *Journal of Statistical Software*, 27(3), 1–22. [CrossRef]
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. [CrossRef]
- International Monetary Fund. (2025). *World economic outlook database: April 2025*. Available online: <https://www.imf.org/en/Publications/WEO/weo-database/2025/April> (accessed on 28 June 2025).
- Jallow, H., Mwangi, R. W., Gibba, A., & Imboga, H. (2025). Transfer learning for predicting of gross domestic product growth based on remittance inflows using RNN-LSTM hybrid model: A case study of The Gambia. *Frontiers in Artificial Intelligence*, 8, 1510341. [CrossRef] [PubMed]
- Jiang, Y., Guo, Y., & Zhang, Y. (2017). Forecasting China's GDP growth using dynamic factors and mixed-frequency data. *Economic Modelling*, 66, 132–138. [CrossRef]
- Jibuti, M. (2020). Convergence and growth—Conflicting goals of economics policy—A case study of Georgia. *Environmental & Socio-Economic Studies*, 8(1), 1–8. [CrossRef]
- Kant, D., Pick, A., & Winter, J. d. (2025). Nowcasting GDP using machine learning methods. *ASTA Advances in Statistical Analysis*, 109, 1–24. [CrossRef]
- Krkoska, L., & Teksoz, U. (2007). Accuracy of GDP growth forecasts for transition countries: Ten years of forecasting assessed. *International Journal of Forecasting*, 23(1), 29–45. [CrossRef]
- Kumar, B., & Yadav, N. (2023). A novel hybrid model combining β SARMA and LSTM for time series forecasting. *Applied Soft Computing*, 134, 110019. [CrossRef]
- Longo, L., Riccaboni, M., & Rungi, A. (2022). A neural network ensemble approach for GDP forecasting. *Journal of Economic Dynamics and Control*, 134, 104278. [CrossRef]
- Maccarrone, G., Morelli, G., & Spadaccini, S. (2021). GDP Forecasting: Machine learning, linear or autoregression? *Frontiers in Artificial Intelligence*, 4, 757864. [CrossRef] [PubMed]

- Madhou, A., Sewak, T., Moosa, I., & Ramiah, V. (2020). Forecasting the GDP of a small open developing economy: An application of FAVAR models. *Applied Economics*, 52(17), 1845–1856. [CrossRef]
- Makridakis, S., & Wheelwright, S. (1998). *Forecasting: Methods and applications* (3rd ed.). John Wiley & Sons.
- Mariano, R. S., & Ozmucur, S. (2021). Predictive performance of mixed-frequency nowcasting and forecasting models (with application to Philippine inflation and GDP growth). *Journal of Quantitative Economics*, 19(Suppl. S1), 383–400. [CrossRef]
- Mohamed, A. O. (2022). Modeling and forecasting Somali economic growth using ARIMA models. *Forecasting*, 4(4), 1038–1050. [CrossRef]
- National Statistics Office of Georgia. (2025). *GEOSTAT*. Available online: <https://www.geostat.ge/en> (accessed on 10 February 2025).
- Perone, G. (2021). Comparison of ARIMA, ETS, NNAR, TBATS and hybrid models for forecasting COVID-19 cases. *Journal of Forecasting*, 40(6), 1009–1022. [CrossRef]
- Perone, G. (2022). Using the SARIMA model to forecast the fourth global wave of cumulative deaths from COVID-19: Evidence from 12 hard-hit big countries. *Econometrics*, 10(2), 18. [CrossRef]
- Perone, G. (2024). *A novel hybrid forecasting model for Georgian GDP*. International School of Economics at TSU, Policy Paper No. 2024/03. Available online: <https://iset-pi.ge/storage/media/other/2024-04-11/d630bf30-f80f-11ee-8485-056f0b7dde49.pdf> (accessed on 20 May 2025).
- Pkhaladze, T. (2025). *Navigating geopolitical realities: The EU's strategic positioning in the South Caucasus and Central Asia*. European Centre for International Political Economy (ECIPE). Available online: <https://ecipe.org/publications/eu-strategic-positioning-in-the-south-caucasus-central-asia/> (accessed on 30 August 2025).
- Poghosyan, K., & Poghosyan, R. (2020). On the applicability of dynamic factor models for forecasting real GDP growth in Armenia. *Applied Econometrics*, 61, 28–46. [CrossRef]
- Puttanapong, N., Prasertsoong, N., & Peechapat, W. (2023). Predicting provincial gross domestic product using satellite data and machine learning methods: A case study of Thailand. *Asian Development Review*, 40(2), 39–85. [CrossRef]
- Richardson, A., Van Florenstein Mulder, T., & Vehbi, T. (2021). Nowcasting GDP using machine-learning algorithms: A real-time assessment. *International Journal of Forecasting*, 37(2), 941–948. [CrossRef]
- Rodrik, D. (2008). The Real exchange rate and economic growth. *Brookings Papers on Economic Activity*, 2008(2), 365–412. [CrossRef]
- Rostan, P., Rostan, A., & Wall, J. (2024). Measuring the resilience to the COVID-19 pandemic of Eurozone economies with their 2050 forecasts. *Computational Economics*, 63(3), 1137–1157. [CrossRef]
- Rusnák, M. (2016). Nowcasting Czech GDP in real time. *Economic Modelling*, 54, 26–39. [CrossRef]
- Sabri, R., Tabash, M. I., Rahrouh, M., Alnaimat, B. H., Ayubi, S., & AsadUllah, M. (2023). Prediction of macroeconomic variables of Pakistan: Combining classic and artificial network smoothing methods. *Journal of Open Innovation: Technology, Market, and Complexity*, 9(2), 100079. [CrossRef]
- Schorfheide, F., & Song, D. (2021). *Real-time forecasting with a (standard) mixed-frequency VAR during a pandemic* (No. w29535). National Bureau of Economic Research. [CrossRef]
- Sein, P., & Sah, A. N. (2025). Export dynamics, exchange rate volatility, and economic stability: Evidence from Asia-Pacific economies. *Humanities and Social Sciences Communications*, 12, 808. [CrossRef]
- Şen Doğan, B., & Midiliç, M. (2019). Forecasting Turkish real GDP growth in a data-rich environment. *Empirical Economics*, 56, 367–395. [CrossRef]
- Shams, M. Y., Tarek, Z., El-kenawy, E. S. M., Eid, M. M., & Elshewey, A. M. (2024). Predicting Gross Domestic Product (GDP) using a PC-LSTM-RNN model in urban profiling areas. *Computational Urban Science*, 4, 3. [CrossRef]
- Shaub, D., & Ellis, P. (2020). *Package 'forecastHybrid'*. Available online: <https://cran.r-project.org/web/packages/forecastHybrid/forecastHybrid.pdf> (accessed on 10 January 2025).
- Staněk, F. (2023). Optimal out-of-sample forecast evaluation under stationarity. *Journal of Forecasting*, 42(8), 2249–2279. [CrossRef]
- Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting*, 16(4), 437–450. [CrossRef]
- The State Statistical Committee of the Republic of Azerbaijan. (2025). *System of national accounts and balance of payments*. Available online: https://www.stat.gov.az/source/system_nat_accounts/?lang=en (accessed on 2 March 2025).
- The Statistical Committee of the Republic of Armenia. (2025). *Economic and financial data for the Republic of Armenia*. Available online: <https://armstat.am/nsdp/> (accessed on 5 February 2025).
- Timmermann, A. (2006). Chapter 4: Forecast combinations. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting* (Vol. 1, pp. 135–196). Elsevier. [CrossRef]
- Tsuchiya, Y. (2023). Assessing the World Bank's growth forecasts. *Economic Analysis and Policy*, 77, 64–84. [CrossRef]
- Tsuchiya, Y. (2024). Conservatism and information rigidity of the European Bank for Reconstruction and Development's growth forecast: Quarter-century assessment. *Journal of Forecasting*, 43, 1399–1421. [CrossRef]
- Tümer, A. E., & Akkuş, A. (2018). Forecasting gross domestic product per capita using artificial neural networks with non-economical parameters. *Physica A: Statistical Mechanics and Its Applications*, 512, 468–473. [CrossRef]

- Yang, Y., Zhang, J., & Wang, L. (2022). A novel general-purpose hybrid model for time series forecasting. *Applied Intelligence*, 52(2), 2212–2223. [[CrossRef](#)] [[PubMed](#)]
- Yenilmez, İ., & Mugenzi, F. (2023). Estimation of conventional and innovative models for GDP per capita: A comparative analysis of artificial neural networks and Box–Jenkins methodologies. *Scientific African*, 22, e01902. [[CrossRef](#)]
- Yoon, J. (2021). Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach. *Computational Economics*, 57, 247–265. [[CrossRef](#)]
- Zhang, S., Li, Z., Jing, L., & Li, X. (2025). Nowcasting monthly Chinese GDP with mixed frequency data: A model averaging approach. *Computational Economics*, 1–19. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.