

International Journal of Economics and Financial Issues

ISSN: 2146-4138

available at http: www.econjournals.com

International Journal of Economics and Financial Issues, 2026, 16(6), 32-43.

Public Expenditure Forecasting in Cambodia: A Hybrid Econometric and Machine Learning Approach

Tepwinuth Chhim, Siphat Lim*

Department of Accounting and Finance, CamEd Business School, Phnom Penh, Cambodia. *Email: lsiphat@cam-ed.com

Received: 20 July 2026 **Accepted:** 19 September 2026 **DOI:** https://doi.org/10.32479/ijefi.24858

ABSTRACT

This study compared traditional time-series and machine-learning methods for the monthly forecasting of government expenditure of Cambodia with dataset. The study employed a natural logarithm of the monthly flows of government expenditure from January 2009 to March 2016. The dataset has 207 observations and was divided into train set, where the general model training took place (80%) and test set out of sample (20%). Mean absolute error (MAE), root mean squared error (RMSE) and mean absolute percentage error (MAPE) are the metrics that are used to evaluate forecasting performance. The study compares the benchmark against the autoregressive integrated moving average (ARIMA) and four machine-learning approaches: Hybrid ARIMA-random forest, random forest (RF), extreme gradient boosting (XGBoost) and support vector regression (SVR). The ARIMA model detects a statistical significant seasonal autoregressive behaviour and a drift, indicating that government expenditure is characterized by long-run growth and seasonal dynamics. Nevertheless, the forecasting accuracy using the traditional method is outmatched by results of machine-learning models in general. In particular, XGBoost outperforms all the other models with the lowest MAE, RMSE and MAPE metrics indicating that it is the best forecasting technique. The random forest is in the second position while hybrid ARIMA-random forest model significantly outperforms standalone ARIMA model, though it remains below the tree-based ensemble methods. SVR also beats ARIMA, but fails to beat XGBoost and random forest. Overall, the results illustrate that machine-learning approaches, particularly XGBoost and random forest, consistently outperform the traditional ARIMA model in terms of fitting government expenditure´s nonlinear, seasonal and trend pattern, which would enhance forecasts accuracy for fiscal planning and policy decisions.

Keywords: Government Expenditure Forecasting, Autoregressive Integrated Moving Average, Machine Learning, XGBoost
JEL Classification: C22, C45, C53, H50

1. INTRODUCTION

Forecasting government expenditure has increasingly become an integral part of fiscal policy and public financial management. Expenditure forecasting is an essential role of governments to establish realistic budgets; ensuring fiscal sustainability; maximising the social and economic return on public resources invested; and anticipating macroeconomic shocks. Governments have encountered extraordinary exchequer unpredictability after the worldwide monetary emergency, COVID-19 pandemic and resulting geopolitical interruptions prompting more prominent need for accurate use conjecture than at any other time. Not just seen as a budgeting tool anymore but is regarded as an essential policy weapon to support macroeconomic stability and evidence based decision making (Carabotta and Claeys, 2024; Ghysels and Ozkan, 2015).

Statistical and econometric time-series techniques were the main ones used in early studies on forecasting state budgets. Of these methods, the autoregressive integrated moving average (ARIMA) model is still one of the most commonly used forecasting techniques due to a well-established theoretical background, transparency and logically sound modeling framework for linear temporal dependencies. In particular, assessments of government

This Journal is licensed under a Creative Commons Attribution 4.0 International License

expenditure forecasting studies for Indonesia and Nigeria in comparison to a number of studies from other emerging economies seem to indicate that ARIMA is relatively resilient for short-term fiscal forecasts, especially where past expenditures align themselves with stable historical properties (Iskandar et al., 2018; Salisu and Omotor, 2024; Marsudiyanto, 2025).

However, later research has more challenged the idea that standard statistical models are enough for modern fiscal forecasting. Structural breaks, policy interventions, political decisions, external shocks and complex nonlinear interactions violate many of the assumptions that underlie classical time-series models for fiscal variables. For instance, Ghysels and Ozkan (2015) showed that using mixed-frequency macroeconomic indicators massively enhances federal budget forecasts compared to standard annual models. Building on this concept, Asimakopoulos et al. (2020) revealed that combining flow real-time fiscal data along with mixed data sampling (MIDAS) techniques strongly converses multiple forecast errors, especially when forecasting government revenues and expenditures by individual components rather than aggregate sums. Similarly, Paredes et al. (2014) showed that richer quarterly fiscal databases yield much greater analytical power for the evaluation of fiscal policy than annual aggregates alone can provide.

More recent advancements then moved towards machine learning methods, which can learn highly complicated nonlinear interactions without explicit functional assumptions. The algorithms such as random forest (RF), XGBoost, and support vector regression (SVR) observe high dimensional relationships emerged as the most used with great predictive performance. Recent comparison studies keep pointing out that ensemble tree-based methods outperform many classical statistical models across different predictive tasks. As an example, Ghannam (2026), by comparing a pool of thirteen models for national CO_2 emissions in Australia, found random forest as the selected model that was more accurate than ARIMA or deep learning approaches based on the out-of-sample error but with better interpretability by employing SHAP analysis. Likewise, Wang et al. (2025) even noted that ARIMA remains the most competitive model for long-range prediction of health expenditure trends, machine-learning approaches discovered richer dynamics in health expenditure than more traditional statistical methods could uncover.

Nevertheless, the rise of machine learning has also affected fiscal forecasting. More recently, Random Forest has been employed successfully for forecasting Italian public debt, outperforming traditional econometric approaches without the need of explicitly identifying the underlying macroeconomic structure (Sica et al., 2024). Recognizing that macroeconomic datasets are often relatively small, combining between dynamic stochastic general equilibrium (DSGE) and machine learning for synthetic training data to make a far better job forecasting public debt. These researches reflect a wider trend in public finance research away from classic structural econometric modelling towards data informed predictive analytics. More recently, in addition to or instead of primarily theory-driven models that depend on strong assumptions about causation and economic behaviour, studies increasingly apply machine learning methods to identify complex relations, increase predictive performance, and capitalise on large administrative and financial datasets. This shift is propelled by the increasingly rich high-dimensional databases and new computational methods, which unlock better predictions and can inform evidence-based public policy (Sofianos et al., 2026).

Several studies warn against the complete replacement of statistical methods, notwithstanding the increasing use of machine learning. Traditional econometric models are very interpretable and often good at capturing linear temporal structure, machine-learning models are generally better at modelling the nonlinear residual behaviour. As a result, more studies have embraced hybrid-forecasting frameworks that combine both approaches. Hassanpouri Baesmat et al. (2025) showed that ARIMA captures linear pattern while random forest model nonlinear residuals, so a success using both models together larger than any one alone (weighted average between ARIMA and RF). Moreover, the ARIMA contributed in electric load forecasting substantially. Similarly, Iskandar et al. (2018) concluded that neither ARIMA nor neural networks consistently improved over the other for all categories of government expenditure, but that hybrid approaches provided better forecasting performance depending on data characteristics. There is an emerging literature that sees statistical and machine learning models not as competing alternatives to one another but as largely complementary forecasting instruments.

A critical key development relates to the realization that no forecasting model has a universal upper hand across all fiscal situations. Carabotta and Claeys (2024) showed that drawing on forecasts from a combination of models, as well as institutions, allowed for significantly reduced budget forecast uncertainty due to the complementary strengths of individual forecasting methods. Williams and Kavanagh (2016) arrived at analogous conclusions as these two studies found that forecasting performance is influenced significantly by data frequency, revenue characteristics, and forecasting methodology with differences in forecast ability being largely explained by institutional context or changes while Reitano (2018) argued that institutional, political, and economic circumstances jointly determine the accuracy of short-term forecasts within local governments. Taken together, these studies indicate that model comparison has recently taken on equal importance in the modern fiscal forecasting literature with respect to model development.

Although the literature on fiscal forecasting has expanded considerably over the last decade, several research gaps continue to characterize this field. Most of the previous studies mostly target developed economies like United States and European Union, are mainly focused on government revenues, budget balances or public debt while little attention has been paid to forecasting of government expenditure in developing economies. Second, there are few studies directly comparing traditional econometric techniques with standalone machine-learning algorithms and with hybrid statistical–machine learning approaches in one analytical design. To the best of our understanding, so far, there is no published study extensively comparing four forecasting methods for Cambodia's government expenditure including

ARIMA, Random Forest, XGBoost and support vector regression in addition to Hybrid ARIMA–random forest models.

In order to overcome these shortcomings, this study provides a rigorous comparative assessment of five forecasting methods including: ARIMA, Hybrid ARIMA–Random Forest, Random Forest, XGBoost and support vector regression on Cambodian government expenditure. Using multiple forecasting accuracy measures, we evaluate the performance of several candidate approaches to identify the optimal one for public expenditure typical in the Cambodian fiscal environment. This study provides contributions to the literature on fiscal forecasting and the role of evidence-based public financial management as means for developing countries, using traditional econometric methods with current machine learning techniques.

2. LITERATURE REVIEW

Public expenditure forecasting is both intangible and significant sphere of public financial management as expenditure forecasts impacts the formation of budget ceilings, fiscal discipline, delivery of public services and credibility and predictability of government fiscal policy. Accurate revenues forecast allow governments to plan their spending more reliably, reduce fiscal risks and better react to shocks, whilst challenges in forecast accuracy may lead to budgets lacking credibility, reduced ability for expenditure control and burden on public finances. This problem is further exacerbated in developing economies, where fiscal buffers might be more limited and forecasting errors can have more severe implications for budget stability and policy execution. The public expenditure forecasting documentation builds upon information from classical statistical methods to machine learning models. It is well known widespread that ARIMA and related time-series methods continue to be used because they provide transparent and interpretable benchmarks based on past behaviours of autocorrelation, trends patterns. In contrast, government expenditure data can also include nonlinear shifts resulting from policy reforms, demographic change, inflationary pressure and investment cycles as well as emergency spending and institutional behaviour. These features cause a need to assess the possibility of improving forecasting proficiency beyond specifically straightforward statistical techniques using machine-learning models.

2.1. Public Expenditure Forecasting and Fiscal Planning

As forecasts affect budget ceilings, expenditure priorities and fiscal policy credibility, public expenditure forecasting has grown in importance within the field of public financial management. Budget forecasts are not just a technical exercise; they determine fiscal space, the provision of public goods and services as well as fiscal responsiveness during economic shocks. This issue is of special relevance for developing economies with low fiscal buffers in which any miss on expenditure forecasting may undermine fiscal discipline. The literature identifies that comparative budget forecasting remains an underdeveloped capability in developing economies, providing a clear rationale for case study focused attention on Cambodia (Kara and Şengüler, 2026).

History of past expenditure, macroeconomic context, demographic shifts and policy commitments and institutional behaviour all Impact public expenditure. Adding demographic information into actual public expenditure forecasting, in South Africa, it is suggested that ARIMAX will not outperform a machine-learning model (Ghiassi and Simo-Kengne, 2021). The findings suggest that demographic monitoring is likely to be relevant to fiscal planning in specific contexts such as where there are rising expenditures on health, social protection and ageing. Public-sector studies also suggest that conventional approaches may not be adequate when expenditure series perfect seasonal patterns or have inertia and high frequency shocks (Drăgulin et al., 2025). In this sense, for Cambodia the implication is that government expenditure should be viewed as more of a dynamic fiscal time series with only the one-off shocks levelled down and not treated simply as a pure trend extrapolation.

2.2. Statistical Forecasting and ARIMA-Based Benchmarks

ARIMA model and similar time-series models are an obvious choice as a starting point for modelling government expenditure: they incorporate trend and persistence using historical values, autocorrelation, differencing. In general forecasting literature ARIMA is considered as a flat benchmark since it is not only transparent, but also theoretically grounded and well suited for many univariate forecasting problems (Petropoulos et al., 2022). This transparency has added a practical benefit for users of research in public finance: government users always know where the forecasts come from and how they perform against official budget assumptions.

Nevertheless, ARIMA-based models are limited due to their reliance on linear structures as such Khashei and Bijari (2011) state that in fact, while ARIMA can effectively capture linear components; when the real-world series exhibit both linearity and nonlinearity, then ARIMA does not perform well. This is particularly true for expenditure data that are subject to nonlinear dynamics such as the effect of policy shocks, election cycle effects, emergency spending and inflationary pressure. Evidence from expenditure-based healthcare studies also supports continued role of ARIMA as baseline. The ARIMA and Prophet were the best performing baseline models for predicting Medicaid expenditure while machine-learning models had an important role in discovering more complex relationships in the context of policies. Combining standard statistical forecasting methods with machine learning methods to analyse health service expenditure and show some advantage in comparing both these groups of models, rather than relying on a single model approach (Wang et al., 2024; Wang et al., 2025).

2.3. Machine Learning and Hybrid Forecasting Models

Machine learning are capable of representing nonlinear relationships, interactions among predictors and complex changes in the data-generating process, they have received considerable attention. This is a very promising pattern in the data emerging from studies about public expenditure and artificial intelligence. Ghiassi and Simo-Kengne (2021) found that a dynamic artificial

neural network proved to be superior to ARIMAX for forecasting the South African public expenditures. Yang et al. (2023) established a comparative experimental study using ARIMA, exponential smoothing, support vector regression, XGBoost, CNN and LSTM and proposed the use of a gated recurrent unit model to predict government expenditure achieving the highest overall accuracy. Drăgulin et al. (2025) also suggest that models guided by advanced machine learning, when evaluated with suitable performance metrics, can improve public expenditure forecasting and increase accountability in the public sector. Dai et al. (2025), for example, show how random forest, XGBoost and neural networks can aid macroeconomic forecasting as well scenario analysis of fiscal policy in China.

Tree based and kernel based methods are also supported from results in the more general forecasting literature. XGBoost is widely used in volatile forecasting where interaction and non-linearity between features are inherent, as it has been shown to perform well in modelling with the regularisation helping reduce overfitting (Tightiz et al., 2025). Random forest approaches are also associated with nonlinearities and the ability to do feature importance analysis; Shiraishi et al. (2024) propose extending random forest quantile regression to static settings, which is particularly useful when uncertainty around expenditure forecasts matter. Due to its learning of non-linear relations among data and generalisation ability in complex environments, support vector regression has been utilised in several demand forecasting contexts (Fan et al., 2023; Nie et al., 2024).

The literature has a strong theme around hybrid models being useful. The hybrid approaches combine both the interpretability and trend-capturing ability of statistical methods with the nonlinear learning capacity of machine learning. This theory was substantiated by Khashei and Bijari (2011) who are the first to bring a piece of evidence that an ARIMA-ANN hybrid can boost the accuracy for time-series forecasting, which follows some linear dynamic as well as nonlinear components. Abellana et al. (2021) use a hybrid SVR-SARIMA model to capture seasonal, linear and nonlinear components in tourism demand forecasting, while He and Qian (2026) combine STL decomposition, holt-winters, ARIMA and XGBoost to improve forecasting accuracy for seasonal tourism data. Hassanpouri Baesmat et al. (2025) also demonstrates that a sparse random forest and ARIMA hybrid can reduce prediction errors for electric load forecasting. Random forest lag selection strengthens ARIMA-LSTM forecasting of volatile monthly series, these applications are out of government spending but their methodological consequence has in fact a direct relevance (Ray et al., 2023). ARIMA may serve as a benchmark for hybrid machine-learning models to improve Cambodian expenditure forecast results.

2.4. Model Evaluation and Research Gap for Cambodia

The evaluation criteria are key to a credible comparison of ARIMA and machine learning methods. The systematic review analyse show that RMSE, MAE and MAPE all dominate on budgeting forecasting evaluation while directional errors are not that focused on (Kara and Şengüler, 2026). Forecasting competitions also show that robust benchmarking and out-of-sample testing are necessary because complex models do not always outperform simpler statistical methods, while RMSE can penalise very large fiscal forecast errors for government expenditure, MAE can provide an easily interpretable mean error by year and MAPE can help assess relative budget forecasting performance across expenditure categories. It is not just the size of an error that matters; a government may want to know whether expenditure is likely to be higher or lower (Makridakis et al., 2020; Petropoulos et al., 2022).

A pluralistic research design for Cambodia is supported by the literature. Machine learning models such as random forest, SVR or XGBoost should be applied to assess nonlinear forecasting, however, the ARIMA should still be a transparent benchmark. Given that fiscal series may be driven by trend, persistence, volatility and policy shocks, the contemplate hybrid specifications are also included in this research. Feature selection is also an important step when high-dimensional macroeconomic or digital indicators are added, like shown in Yang and Wu (2026) who introduce new external search variables to improve demand forecasting with Boruta feature selection. The most important gap, however, is not a lack of forecasting methods but a lack of data on what works best where expenditure forecasting in developing countries is concerned. To help remedy this shortcoming, a comparison in the Cambodia context that tests whether government expenditure forecasts beyond ARIMA materially benefit from machine learning using consistent out-of-sample evaluation and policy-relevant error metrics, is the focus of this paper.

3. METHODOLOGY

This study implements five forecasting methods for natural logarithm of the government expenditure (LNG) of Cambodia, which include autoregressive integrated moving average (ARIMA), hybrid ARIMA–random forest, random forest (RF), extreme gradient boosting (XGBoost) and support vector regression (SVR) methods. The summary of all models is presented in Table 1. The study considers the ARIMA model as a benchmark statistical model and include the machine learning models to investigate if incorporating nonlinear forecasting methods improves predictability. Comparing statistical and machine learning forecasting methods, since the relative performance can

Table 1: Model summary

Model	Type	Role in this study
ARIMA	Traditional time-series model	Benchmark model for capturing linear time-series structure
Hybrid ARIMA+ Random Forest	Hybrid model	Combines ARIMA linear forecasting with Random Forest residual learning
Random Forest	Machine learning model	Captures nonlinear relationships using lagged LNG, trend, and month
XGBoost	Machine learning model	Uses boosted trees to improve nonlinear forecasting accuracy
SVR	Machine learning model	Uses kernel-based regression to model nonlinear relationships

vary by time-series characteristics, sample size, stationarity and horizon, has been justified in recent literature (Makridakis et al., 2018; Makridakis et al., 2020; Petropoulos et al., 2022).

3.1. Data Splitting and Forecasting Design

The LNG series was divided chronologically for training and testing datasets. 80% of all observations were used to fit the model and 20% for evaluation on out-of-sample forecasting. It is important to note that the training dataset includes 166 observations from January 2009 to October 2022. The test dataset consists of 41 observations from November 2022 until March 2026. All data were collected from the database of the ministry of economy and finance. As time-series observations are in an ordered sequence over time, the data were not randomly shuffled. This method is suitable since the model for forecasting must be fitted using previous observations and then evaluated on future unseen observations. It is recommended that out-of-sample evaluation methods with preserved temporal ordering be used for non-stationary time-series data (Cerqueira et al., 2020).

3.2. ARIMA Model

The ARIMA model was used as a baseline statistical forecasting model for the LNG series. ARIMA-type models are the most popular benchmark and baseline models for comparison between statistical and machine learning approaches, particularly because these methods are more transparent and interpretable and were designed to operate on ordered time-series data (Makridakis et al., 2018; Makridakis et al., 2020; Petropoulos et al., 2022).

Let Y_t represent the LNG value at t. The general ARIMA model is written as:

$$ARIMA(p, d, q)$$

where p is the number of autoregressive terms, d is the degree of differencing required to make the series stationary, and q is the number of moving-average terms. The autoregressive component captures the relationship between the current LNG value and its previous values. The integrated component applies differencing to remove non-stationarity. The moving-average component captures the effect of past forecast errors.

More formally, the ARIMA model can be expressed as:

$$\phi(B)\,(1 - B)^{d} Y_t = c + \theta(B)\varepsilon_t$$

Where B is the lag operator, $\phi\,(B)$ represents the autoregressive polynomial, $(1 - B)^{d}$ represents differencing, $\theta(B)$ represents the moving-average polynomial, c is a constant term, and ε_t is the error term.

In order to estimate the ARIMA model, the LNG data series was split chronologically into training and testing datasets. The model was estimated using the first 80% of monthly observations, while the remaining 20% were used to evaluate in-sample forecasting performance. The test case data was time-ordered and not randomly shuffled since time-series forecasting must keep the order of observations. Nevertheless, this helps to prevent data leakage and makes sure that the model is assessed against future observations out of sample (i.e., no used in prediction). Cerqueira et al. (2020) state that preserving temporal order is very important when out-of-sample methods are applied to obtain forecasts for non-stationary time series.

The first step in ARIMA modelling was to check whether the LNG series was stationary or not. A time series that is stationary has a constant mean, variance and autocorrelation structure over time. Economic and fiscal time series often exhibit trend patterns, thus the LNG series may be non-stationary in level. The augmented dickey-fuller (ADF) test was used therefore to find out if a unit root exist. If the LNG series was stationary, it differenced. Moderate non-stationarity is possible to approach long-term trend and accordingly, DLNG which is first difference of LNG after each levels was used when above estimated values were required.

Once the differencing order d was decided, the autoregressive order p and the moving-average order q were chosen. The autoregressive term p reflects the direct impact of past LNG values on current LNG, while the moving-average term q captures the effects of past shocks or errors. Generally evolving with the data driven process along with information criterions such as AIC, AICc or BIC and also correlated by the auto-correlation function or partial auto correlation function. The preferred ARIMA specification is the model that offers a trade-off between forecast accuracy and model parsimony.

Due to the fact that the data are monthly, seasonal patterns may exist as well. A seasonal ARIMA specification may be considered when seasonality is found. The seasonal ARIMA model is usually denoted as follows:

The correct way to write the seasonal ARIMA model is:

$$ARIMA\ (p, d, q)\ (P, D, Q)_{12}$$

or, more specifically, it is often called:

$$SARIMA(p,d,q)\ (P,D,Q)_{12}$$

The general model form is:

$$\Phi_{p}\,(B^{12})\phi_{p}(B)(1 - B)^{d}\,(1 - B^{12})^{D} y_t = \Theta_{Q}\,(B^{12})\theta_{q}\,(B)\varepsilon_t$$

Where:
(p, d, q)

Represents the non-seasonal ARIMA part, and
$(P, D, Q)_{12}$

Represents the seasonal part with a seasonal period of 12, usually for monthly data.

Where P, D, and Q are the seasonal autoregressive (SAR), seasonal differencing (SD), and seasonal moving-average (SMA) terms, respectively, with 12 indicating monthly frequency. The LNG time series was treated as a monthly data, thus model estimation

permitted the time-series structure to reflect monthly patterns where appropriate.

Once the ARIMA model was fitted within the training data, it was also important to perform diagnostic checking to check that the residuals were white noise. Residuals left by a well-fitted ARIMA model should not show remaining systematic pattern. Residual diagnostics were visually inspected, autocorrelation was checked and the Ljung-Box test was performed to detect possible presence of autocorrelation. If the residuals, however, revealed that substantial autocorrelation was still present then this indicated a need to further reassess the model specification.

The estimated ARIMA model was used to forecast LNG during the test period. The forecasted LNG and actual LNG were compared to evaluate the accuracy of each forecast. the root mean squared error, mean absolute error and mean Absolute percentage error were the main evaluation measures. Those measures are frequently employed in empirical forecasting studies to evaluate the predictive accuracy of statistical and machine learning approaches (Makridakis et al., 2020; Petropoulos et al., 2022).

In this study, ARIMA was taken as the benchmark model in this study. It was meant to show a statistical baseline with which the performance of Hybrid ARIMA–random forest, random forest, XGBoost and SVR could be compared against. The main reason for this is that the nonlinear machine learning approaches indeed outperform ARIMA forecasting since much less errors are yielded by these models and in comparison with linear time-series modelling by traditional methods, such as the different variants of ARIMA.

3.3. Hybrid ARIMA-Random Forest Model

Based on the statistical and machine-learning method combination, the Hybrid ARIMA random forest model was created. ARIMA model was the first structure, which was fitted to this LNG training series is time-series based and it provide linear time series. Second, the ARIMA residuals were calculated and then modeled via random forest with lagged residuals as predictors. The last hybrid forecast was created by summing the ARIMA forecast and the residual forecast for Random Forest:

$$Hybrid\ Forecast_t = ARIMA\ Forecast_t + RF\ Residual\ Forecast_t$$

The hybrid model works on the basis that ARIMA picks up linear part of LNG and random forest picks up any residual nonlinearity in the residuals. This is why recent literature on forecasting has suggested that hybrid and ensemble approaches may be used since several forecasting models might pick different characteristics of a time series (Masini et al., 2023; Petropoulos et al., 2022).

3.4. Random Forest Model

In this study, the dependent variable is *LNG*, which refers to the logarithm of government expenditure. The Random Forest model is specified as an autoregressive machine-learning model with deterministic trend and monthly seasonality since the model only relies on historical tax revenue observations and temporal variables. The model is expressed as:

$$LNG_t = f_{RF}\left(LNG_{t-1}, LNG_{t-2},\ldots,LNG_{t-p}, Trend_t, Month_t\right) + \varepsilon_t$$

Where lagged values of *LNG* captures persistence in government expenditure, captures long term trend and captures seasonality monthly frequency. Random Forest prediction is the average of the predictions from all regression trees. Random Forest is suitable for forecasting because it can capture nonlinear relationships and interaction effects without requiring a strict parametric functional form. It captures nonlinear relationships, does not implying strict parametric functional form, and allow for interaction effects between predictors. Recent time-series forecasting research proposes tree-based methods such as Random Forest as significant nonlinear machine learning approaches for statistical forecasting (Masini et al., 2023).

3.5. XGBoost Model

Define the input vector as:

$$Z_t = \left(LNG_{t-1}, LNG_{t-2},\ldots, LNG_{t-p}, Trend_t, Month_t\right)$$

The general XGBoost model is:

$$LNG_t = f_{XGBoost}\left(LNG_{t-1}, LNG_{t-2},\ldots,LNG_{t-p}, Trend_t, Month_t + \varepsilon_t\right.$$

The predicted value is obtained by adding many boosted regression trees:

$$LNG_t = \sum_{m=1}^{M} \eta T_m\left(Z_t\right)$$

Where M is the number of boosted trees, η is the learning rate, and $T_m\left(Z_t\right)$ is the prediction from m^{th} regression tree.

XGBoost was also applied in the same supervised learning structure as mentioned behind for random forest. Inputs of the regressors are the lagged LNG variables, trend and month variables. This form of ensemble decision tree modeling, called XGBoost is a gradient-boosted tree algorithm that builds trees one at a time, where each new tree reduces the loss function from the existing decision model. It is also valuable for modelling complex nonlinear relationships, and boosting can improve predictive performance. Boosted trees are shown in recent machine learning forecasting studies and reviews to be a notable class of nonlinear forecasting models (Masini et al., 2023).

3.6. Support Vector Regression Model

A second machine learning method was support vector regression for the application of LNG forecasting. The SVR model also used lagged LNG values, along with the trend and month, on par with random forest and XGboost. In SVR a kernel function is applied to all input variables in order to map them into a higher-dimensional usually infinite-dimensional feature space in which it estimates the nonlinear functions. The radial basis function kernel was used in this study due to its effectiveness to model nonlinear relationships. SVR is added because recent forecasting studies show that it is a nonlinear

machine learning model with good generalisation capacity, particularly in multidimensional prediction contexts (Alsalem et al., 2025; Santoro et al., 2024).

3.7. Model Evaluation

Forecasting performance of each model was evaluated on out-of-sample test data using mean absolute error (MAE), root mean squared error (RMSE) and mean absolute percentage error (MAPE). These metrics were utilized to assess the closeness of fit between each models predicted LNG values and actual LNG values during the test period. The MAE and RMSE indicated the best forecasting performance of models, while MAPE gives us an understandable error in a percentage format. Out-of-sample accuracy comparison is a popular criterion used in forecasting competitions and empirical forecasting studies to assess competing statistical and machine learning methods (Makridakis et al., 2020; Cerqueira et al., 2020).

Mean absolute error

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|LNG_t - LNG_t\right|$$

Where

LNG_t: Actual value at time t

LNG_t : Forecast value at time t

n: Number of forecast observations

Root mean squared error

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(LNG_t - LNG_t\right)^2}$$

Mean absolute percentage error

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{LNG_t - LNG_t}{LNG_t}\right|$$

4. RESEARCH RESULTS

Table 2 presents logarithm (LNG) and the first difference of logarithm government expenditure (DLNG) summary statistics, from January 2009 until March 2026. LNG contains 207 monthly observations, while DLNG contains 206 observations due to the loss of one observation after first differencing. With respect to LNG the average is 6.75 (its median is 6.86), which likely indicates a slight left-skewed distribution. The interquartile range values are from 6.24 to 7.22, illustrating that the middle 50% of expenditure values on log scaled measurements are rather compact, although wider minimum and maximum measurements of the range of our sample data equate to 4.07 and 8.49 respectively indicating variation in individual streams across the whole period sampled. With a standard deviation of 0.78, the level series presents moderate dispersion. DLNG as a growth rate of government expenditure has mean and median of 1.90 and 4.68, respectively suggesting spending trends towards higher levels through time. The growth series, on the other hand, is extremely inconsistent with a standard deviation of 70.75 and that range from −244.18 to 169.56. The interquartile range, ranging from −24.63 to 39.85, additionally validates substantial month-to-month oscillations. Finally, and in contrast with LNG, which captures both the long-run behaviour of government expenditure, DLNG picks up short-run fluctuations potentially as consequences of irregular fiscal disbursements, seasonal budget execution, policy shocks or exceptional expenditure adjustments within threshold limits.

Presented in Table 3 are the results for the ADF unit root test for LNG and DLNG using 80% training dataset. Choosing this method is right since we check for stationarity on the estimated data that, as noted before again; in turn the remaining 20% is saved to test out of sample. The test of ADF is a statistical content that tested the null hypothesis no a unit root exists in id regions this means that the series would non-stationary. The ADF statistic for LNG is −5.9168 with a P = 0.01. Since the P-value is less than the conventional 5% significance level, we thus reject the null hypothesis of a unit root. This shows that the logarithm of government expenditure is stationary in the training sample. This also goes to DLNG where the ADF statistic is −8.0046 and P = 0.01 suggesting that the null hypothesis of unit root is rejected, hence, the growth rate series is stationary as well. A lag order of 5 suggests that the test controls for monthly data on short-term dynamic effects. In total, both the LNG and DLNG results suggest that neither series displays unit root behaviour; consequently permitting modelling with the training sample. This lessens the potential of spurious modelling results and lends some confidence to any forecasting analysis, which is based upon this training data.

The ARIMA results for forecasts, evaluated on the remainder 20% test sample using an 80% training sample, are shown in Table 4. In Panel A, both seasonal autoregressive components are statistically significant and negative: sar1 = −0.4610, P < 0.001, and sar2 = −0.3545, P < 0.001. This implies that the logged government spending series exhibits genuine seasonal persistence where past seasonal movements produce a corrective influence on future consumption behaviour. An especially large positive drift coefficient of 0.0117, P < 0.001), suggests an ongoing increase in LNG through the entire training period. In the test data, MAE = 0.485, RMSE = 0.570 and MAPE = 6.564% for ARIMA model as indicated in Panel B. The MAPE of below 10% implies a relatively good performance in out-of-sample forecasting. The RMSE being greater than the MAE indicates that there remain

Table 2: Summary statistics

Variable	Observations	Mean	Median	Q25	Q75	Minimum	Maximum	Standard deviation
LNG	207	6.75	6.86	6.24	7.22	4.07	8.49	0.78
DLNG	206	1.90	4.68	−24.63	39.85	−244.18	169.56	70.75

Q25 and Q75 refer to the 25th and 75th percentiles

some larger forecasting errors, which could be due to irregular fiscal disbursements or expenditure shocks.

These findings are in line with existing forecasting literature showing that ARIMA models are suitable for stationary and rather linear time-series data, especially for short-term forecasts (Vuong et al., 2024; Ramos et al., 2015). However, previous studies show evidence that ARIMA models do not perform as well as few hybrid or machine-learning approaches when the data generating process is known to be characterized by strong non-linearity, volatility and/or extreme shocks (Duy et al., 2024; Khashei and Bijari, 2011; Mahmud, 2026). Thus, ARIMA model can be a robust statistical baseline for LNG prediction but ideally, this accuracy needs to be compared with other approaches like SVR, RF, XGBoost and hybrid models surrounding ARIMA.

The result for LNG in Table 5 corresponding to the Hybrid ARIMA–random forest using the 80% training sample and tested on the remaining 20% test sample. In MSE change representation in Panel A, it is observed that prediction of LNG Lag1 contributes the most since its percent increase in MSE is the highest at 12.07%. This suggests that government spending in the immediately preceding month has the strongest predictive content for current expenditure. However, second lag is LNG Lag12 indicating annual seasonality in government expenditure. Another significant predictor is trend, confirming that the long-run path of spending does indeed continue to matter in explaining LNG. On the contrary, LNG Lag6 has a negative percentage change in MSE, signifying that this variable is not helpful at all in improving predictions.

Regarding to the hybrid model, the MAE is 0.383, RMSE is 0.480 and MAPE is 5.114% as shown above in Panel B section. The hybrid approach improves the MAE and RMSE by 21.03% and 15.79% respectively when compared to a standalone ARIMA model reported previously, while reducing MAPE by about 22.09% between both models. This suggests that the combined ARIMA and random forest captures both time-series structure and any possible nonlinear relationships in the data, thus improving forecasting accuracy.

Because of these two predictions, it is concluded that hybrid models can be more effective than only ARIMA because ARIMA captures the linear time-series patterns and machine-learning models have enhanced ability to model nonlinear structures in data (Zhang, 2003; Khashei and Bijari, 2011). Other recent research has also confirmed the effectiveness of hybrid forecasting frameworks for complex time-series data, while ARIMA is still described as useful in linear and stationary series but less powerful when nonlinear patterns are present (Vuong et al., 2024).

Table 6 deals with the random forest forecast for LNG, based on 80% training sample, with forecast accuracy assessed for a test sample of 20%. As depicted in Panel A, the most important predictor is LNG Lag12 based on both a percent increase in MSE (28.81) and node purity (25.26). This suggests that annual seasonality is the primary factor explaining monthly government expenditure. Next, month comes second; reinforcing the prominence of seasonal budgetary behavior, and trend ranks third suggests a significant contribution of longer-term expenditure growth to the forecast. Shorter lags such as LNG Lag1, LNG Lag2 and LNG Lag3 are still significant but play a less important role than the annual lag.

Panel B demonstrates that random forest forecasting out-of-sample performance is remarkably strong, at an MAE of 0.349, RMSE of 0.452 and MAPE of 4.604%. Random Forest reduces MAE, RMSE and MAPE by about 28.04%, 20.70% and 29.86% as compared to the ARIMA model respectively. Additionally, it beats the Hybrid ARIMA–random forest model with 8.88% less in MAE value, 5.83% lower RMSE value and 9.97% lower MAPE value. These results indicate that random forest is more successful than the benchmark models in capturing nonlinear patterns caused due to seasonal effects in LNG.

This result is also consistent with previous evidence suggesting that well-tuned random forest models can outperform ARIMA and ANN for time-series forecasting (Shakila and Babu, 2025). This is also backed by recent studies on random forest usefulness in time-series modelling (Shiraishi et al, 2024). However a couple of researchers suggest that ARIMA, hybrid approaches or deep-learning methods can outperformed Random Forest in some cases, especially when it comes to the differences between training data and test forecasting horizon, but it is difficult to compare without practical use-cases (Chauhan et al., 2023; Gerakoudi et al., 2026).

Table 3: ADF unit root test results for LNG and DLNG

Variable	Data used	ADF statistic	Lag order	P-value	Decision
LNG	80% training data	−5.9168	5	0.01	Stationary
DLNG	80% training data	−8.0046	5	0.01	Stationary

The ADF test evaluates the null hypothesis that the series has a unit root. If the P=0.05, the series is treated as stationary. The test was conducted using only the first 80% of the data as training data

Table 4: ARIMA model results for LNG

Panel A: ARIMA component estimation results					
Parameter	**Estimate**	**Standard error**	**t-value**	**P-value**	**Significance**
sar1	−0.4610	0.0785	−5.8732	0.000	***
sar2	−0.3545	0.0794	−4.4674	0.000	***
drift	0.0117	0.0009	13.0992	0.000	***
Panel B: Forecast accuracy results					
Model		**MAE**	**RMSE**	**MAPE**	
ARIMA		0.485	0.570	6.564	

Table 7 presents the XGBoost forecasting results of LNG using the 80% training sample and evaluated the prediction accuracy during the first 20% test sample. In the top of Panel A, it can be seen that LNG Lag12 is most important predictor with gain value = 0.749. This means the annual lag has the greatest impact on reducing forecast error, which indicates consistency with an underlying seasonal or year-end budgetary cycle of government expenditure. Trend followed closely with an increase of 0.134 confirming that long-run growth in expenditure is important. Month also adds to the model providing more evidence of seasonality in fiscal behaviour. Thus, while LNG Lag1 has a frequency value of 0.174, which is reasonably high, its lower gain (0.025) indicates that it is used frequently in the trees but does less to improve predictive accuracy than annual lag.

Table 5: Hybrid ARIMA+random forest forecast results for LNG

Panel A: Random forest variable importance			
Rank	**Variable**	**Percent increase in MSE**	**Increase in node purity**
1	LNG Lag1	12.07	0.98
2	LNG Lag12	10.34	1.29
3	Trend	10.04	1.24
4	LNG Lag2	3.69	0.98
5	Month	2.36	0.63
6	LNG Lag3	1.59	0.91
7	LNG Lag6	−1.69	1.09

Panel B: Forecast accuracy results			
Model	**MAE**	**RMSE**	**MAPE**
Hybrid ARIMA+Random Forest	0.383	0.480	5.114

Table 6: Random forest forecast results for LNG

Panel A: Random forest variable importance			
Rank	**Variable**	**Percent increase in MSE**	**Increase in node purity**
1	LNG Lag12	28.81	25.26
2	Month	21.70	7.26
3	Trend	17.69	15.37
4	LNG Lag1	14.23	5.48
5	LNG Lag3	9.41	8.50
6	LNG Lag2	8.25	7.24
7	LNG Lag6	6.04	5.09

Panel B: Forecast accuracy results			
Model	**MAE**	**RMSE**	**MAPE**
Random forest	0.349	0.452	4.604

Table 7: XGBoost forecast results for LNG

Panel A: XGBoost variable importance				
Rank	**Variable**	**Gain**	**Cover**	**Frequency**
1	LNG Lag12	0.749	0.246	0.209
2	Trend	0.134	0.169	0.152
3	Month	0.050	0.103	0.080
4	LNG Lag1	0.025	0.114	0.174
5	LNG Lag6	0.016	0.146	0.127
6	LNG Lag3	0.015	0.116	0.130
7	LNG Lag2	0.012	0.107	0.128

Panel B: Forecast accuracy results			
Model	**MAE**	**RMSE**	**MAPE**
XGBoost	0.344	0.441	4.569

In Panel B of Table 7 presents the out-of-sample evaluation results for XGBoost. The forecasted results of XGBoost reduces MAE, RMSE and MAPE for ARIMA by 29.07%, 22.63% and 30.39% respectively. It also slightly outperforms random forest, although the improvement is modest. This suggests that XGBoost is the best-performing model among the reported alternatives, but its advantage over random forest is relatively small.

These results are similar to the forecasting studies which concluded that modeling time series with seasonality, trend and nonlinear patterns require advanced methods such as XGBoostbased predictions. Recent studies show that optimized XGBoost achieves greater robustness in volatile markets, and verifies the better effectiveness of STL-XGBoost over benchmark models for seasonal and trending series (Tightiz et al., 2025; He and Qian, 2026). Similar work had previously confirmed competitive XGBoost performance compared to other algorithms with these methods and shows how XGBoost and Random Forest increasingly dominate in prediction settings (Ganguly et al., 2026; Yang and Wu, 2025).

Result of SVR for LNG is presented in Table 8 with trained samples and test sample as evaluated from the other 20%. Panel A shows the defined SVR model with radial basis kernel under ε-regression, cost = 10, gamma = 0.10 and epsilon = 0.10 as the defined specifications, i.e., used to create the linear problem addressed in this study. In total, across all 20 iterations of the cross-validation procedure, the model incorporated 119 support vectors from 156 training observations indicating that a large proportion of available data were used to specify the prediction function. This suggests that the LNG series comprises intricate relationships, necessitating significant backing from past observations.

As indicated by the results in Panel B, SVR achieved an MAE of 0.406, RMSE of 0.550 and MAPE of 5.288%. The MAPE of less than 10% suggests that the forecasting accuracy out-of-sample is good. SVR outperforms the ARIMA benchmark reducing MAE by 16.29%, RMSE by 3.51% and MAPE by 19.44% respectively when comparing with ARIMA breaking its performance ceiling significantly. The SVR proves to be useful also, as it can capture that nonlinear behaviour, beyond the traditional time-series models.

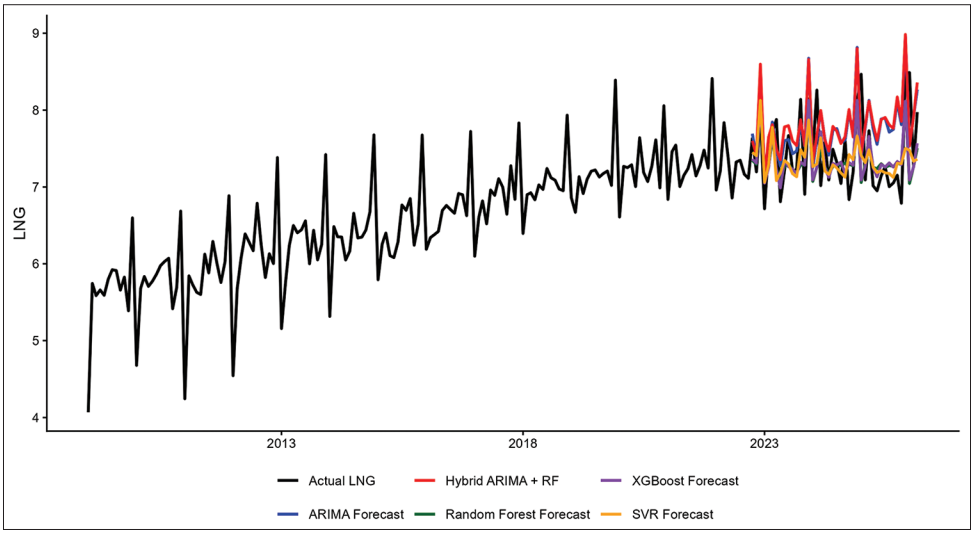
On the other hand, SVR does not perform better than models learnt with the more modern machine-learning approaches noted earlier. Similar to XGBoost, the MAE, RMSE and MAPE of SVR are 18.02%, 24.72% and 15.74% higher than respectively it is for reported in comparison. It also has a relatively low accuracy compared to Random Forest and the Hybrid ARIMA–random forest model. This result is in accordance with the previous findings that SVR is capable of learning nonlinearity and seasonality but good accuracy relies many on careful model specification, parameter tuning and feature design (Fan et al., 2023; Huang et al., 2022; Nie et al., 2024; Tay and Cao, 2001).

In this context, this research compare the actual versus predicted LNG values over the test set by using five different time-series models include the ARIMA, hybrid ARIMA + random forest,

Table 8: SVR forecast results for LNG

Panel A: Model specification								
Model	**Kernel**	**SVR type**	**Cost (C)**	**Gamma (γ)**	**Epsilon (ε)**	**Support vectors**	**Training obs.**	**Testing obs.**
Support vector regression	Radial basis	ε-regression	10	0.10	0.10	119	156	39
Panel B: Forecast accuracy results								
Model	**MAE**	**RMSE**	**MAPE**					
Support vector regression	0.406	0.550	5.288					

Figure 1: Integrated forecasting graph for LNG



The graph compares actual LNG with forecasts from ARIMA, Hybrid ARIMA + Random Forest, Random Forest, XGBoost, and SVR during the test period

random forest, XGBoost and SVR, as suggested in Figure 1. All models capture the upward trend in government expenditure on aggregate. Nonetheless, the study suggested that XGBoost dominates the major top performing models here with respect to predictive accuracy, as its predictions are much more align and able to capture short term movements in the data. ARIMA is smoother than the other machine learning models, and each of the models reaches different levels of promptness. In general, the results indicate that among those reviewed in this study, XGBoost appears to provide the best forecasting performance for LNG and is thus likely more appropriate for modeling complex expenditure dynamics.

5. CONCLUSION AND POLICY IMPLICATION

5.1. Conclusion

The results show a concrete proof that government spending, in the log form, can be predicted fairly successfully using both the traditional time-series and machine-learning methodologies. We demonstrate that the ARIMA model has statistically significant seasonal autoregressive components and a positive drift term, suggesting that government expenditure trends up with clear seasonality. Nonetheless, its predictive performance is lower relative to that of the machine-learning models, as evidenced by an MAPE of 6.564%.

In the comparison with alternative models, XGBoost gives the best forecasting performance, as well as the lowest MAE, RMSE and MAPE. Random Forest is slightly lower suggesting tree-based ensemble methods do a better job when modelling the structure of monthly government spending. The strong importance of LNG Lag12, Month and Trend across the machine-learning models indicates that expenditure behaviour is largely influenced by annual seasonality, monthly budget execution patterns and long-term expenditure growth. The Hybrid ARIMA-random forest model proposed explores the idea of combining statistical and machine-learned techniques and shows improvement upon the pure ARIMA, thus validating that it is beneficial to combine different forecasting approaches. Its performance, however, is still a little behind the base random forest and XGBoost models. Compared to SVR, the ARIMA is good but not as good as tree based models, this indicates that kernel based models are highly dependent on parameter fine-tuning and feature engineering.

The results suggest that, in general, machine-learning methods yield stronger predictive accuracy compared with government expenditure forecasts from the traditional ARIMA benchmark, namely XGBoost and Random Forest. The findings indicate that government expenditure is nonlinear, seasonal and trend-driven in nature, which makes ensemble-learning methods more capable of modeling such time series structures compared to a linear-only time-series specification.

5.2. Policy Implication

There are several policy implications of the results. First, fiscal authorities must better use data-driven forecasting tools for budget preparation, cash flow planning and expenditure monitoring. More

reliable projections of government spending can help improve liquidity management, mitigate the risk of cash shortfalls and better time fiscal disbursements. The second point is that the financial yearly lag and its monthly property are strong forecasters, indicating seasonal patterns of budget execution. This reinforces why the expenditure in a fiscal year should be smoothened over time for avoiding concentrating spending in few months of the particular financial year.

Third, volatile government expenditure growth reveals that it should do more to monitor short-term fluctuations in fiscal policy. Great modifications month-on-month may mirror irregular disbursement procedures, postponed task execution, emergency spending or management bottlenecks. Using forecasting models can thus be seen as a lead indicator for anomalous expenditure movements that warrant policy scrutiny.

Fourth, the outperformance of XGBoost and random forest indicates that public financial management systems can gradually integrate machine-learning based forecasting into traditional econometric models. While such models would not substitute for policy judgement, they can improve information for uses e.g., in budget planning, assessing fiscal risks and medium-term expenditure management. Finally, yet importantly, model performance should be routinely checked when a new data arrives. Updating, validating and comparing forecasting models based on evidence-based fiscal policymaking are required for time-to-time accuracy.

REFERENCES

Abellana, D.P.M., Rivero, D.M.C., Aparente, M.E., Rivero, A. (2021), Hybrid SVR-SARIMA model for tourism forecasting using PROMETHEE II as a selection methodology: A Philippine scenario. Journal of Tourism Futures, 7(1), 78-97.

Alsalem, K. (2025), A hybrid time series forecasting approach integrating fuzzy clustering and machine learning for enhanced power consumption prediction. Scientific Reports, 15, 6447.

Asimakopoulos, S., Paredes, J., Warmedinger, T. (2020), Real-time fiscal forecasting using mixed-frequency data. The Scandinavian Journal of Economics, 122(1), 369-390.

Carabotta, L., Claeys, P. (2024), Combine to compete: Improving fiscal forecast accuracy over time. Journal of Forecasting, 43(4), 948-982.

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.

Chauhan, B., Tabassum, R., Tomar, S., Pal, A. (2023), Analysis for the prediction of solar and wind generation in India using ARIMA, linear regression and random forest algorithms. Journal of Engineering Research, 47(2), 251-265.

Dai, L., Feng, Y., Chen, Y., Zhang, M. (2025), Enhancing Macroeconomic Forecasting and Fiscal Policy Analysis in China with Machine Learning. In: 2nd International Conference on Digital Economy, Blockchain and Artificial Intelligence (DEBAI 2025). p60-64.

Drăgulin, B., Ștefan, V., Tăbîrcă, A. I., Șcheau, M. C., Radu, V., Munteanu, V. (2025), AI-driven models for forecasting public expenditures in the digital era. Electronics, 14(20), 4047.

Duy, Q.P., Nguyen Thi, O., Le Thi, P.H., Pham Hoang, H.D., Luong, K.L., Nguyen Thi, K.N. (2024), Estimating and forecasting bitcoin daily prices using ARIMA-GARCH models. Business Analyst, 45(1), 11-23.

Fan, W., Wu, X., Shi, X.Y., Zhang, C., Hung, I.W., Leung, Y.K., Zeng, L.S. (2023), Support vector regression model for flight demand forecasting. International Journal of Engineering Business Management, 15, 1-9.

Ganguly, S., Sengupta, P., Aggarwal, D. (2026), Predicting EU emissions allowance prices using macroeconomic indicators and hybrid AI models. Journal of Forecasting, 1-21. https://doi.org/10.1002/for.70173.

Gerakoudi, K., Georgoulas, D., Stavroulakis, P.J. (2026), A novel hybrid ARIMA-LSTM model for maritime shipping stock forecasting: Comparative evidence against statistical and machine learning benchmarks. Maritime Business Review, 11(1), 2-20.

Ghannam, S. (2026), An explainable comparative study of statistical, machine learning, deep learning, and hybrid models for CO_2 emissions forecasting in Australia. Array, 29, 100639.

Ghiassi, M., Simo-Kengne, B.D. (2021), Machine learning for economic modeling: An application to South Africa's public expenditures. International Journal of Public Administration in the Digital Age, 8(1), 1-17.

Ghysels, E., Ozkan, N. (2015), Real-time forecasting of the US federal government budget: A simple mixed frequency data regression approach. International Journal of Forecasting, 31(4), 1009-1020.

Hassanpouri Baesmat, K., Shokoohi, F., Farrokhi, Z. (2025), SP-RF-ARIMA: A sparse random forest and ARIMA hybrid model for electric load forecasting. Global Energy Interconnection, 8, 486-496.

He, M., Qian, X. (2026), Forecasting tourist arrivals using STL-XGBoost method. Tourism Economics, 32(2), 408-436.

Huang, Y., Deng, C., Zhang, X., Bao, Y. (2022), Forecasting of stock price index using support vector regression with multivariate empirical mode decomposition. Journal of Systems and Information Technology, 24(2), 75-95.

Iskandar, I., Willett, R., Xu, S. (2018), The development of a government cash forecasting model. Journal of Public Budgeting, Accounting and Financial Management, 30(4), 368-383.

Kara, B., Şengüler, H. (2026), A comparative analysis of budget forecasting methods: A systematic literature review covering the 1983-2024 period. Public Budgeting and Finance, 46, 29-46.

Khashei, M., Bijari, M. (2011), A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. Applied Soft Computing, 11(2), 2664-2675.

Mahmud, A., Musa, K.I., Hamzah, F.M., Badrin, Z.M.Y., Nawi, M.A.A. (2026), Enhancing COVID-19 Forecasting accuracy in Malaysia using a hybrid ARIMA-LSTM model with exogenous variables: A time-series predictive study. Health Science Reports, 9, e72684.

Makridakis, S., Spiliotis, E., Assimakopoulos, V. (2018), Statistical and machine learning forecasting methods: Concerns and ways forward. PLoS One, 13(3), e0194889.

Makridakis, S., Spiliotis, E., Assimakopoulos, V. (2020), The M4 Competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting, 36(1), 54-74.

Marsudiyanto, A. (2025), Addressing fiscal gap between budget availability and financing needs for Indonesian defense strategies. Asian Journal of Accounting Research, 10(4), 430-455.

Masini, R.P., Medeiros, M.C., Mendes, E.F. (2023), Machine learning advances for time series forecasting. Journal of Economic Surveys, 37(1), 76-111.

Nie, R., Che, J., Yuan, F., Zhao, W. (2024), Forecasting peak electric load: Robust support vector regression with smooth nonconvex epsilon-insensitive loss. Journal of Forecasting, 43(6), 1902-1917.

Paredes, J., Pedregal, D.J., Perez, J.J. (2014), Fiscal policy analysis in the euro area: Expanding the toolkit. Journal of Policy Modeling, 36(5), 800-823.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M.Z., Barrow, D.K., Ben Taieb, S., Bergmeir, C., Bessa, R.J., Bijak, J., Boylan, J.E., Browell, J., Carnevale, C., Castle, J.L., Castle, P., & Ziel, F. (2022), Forecasting: Theory and practice. International Journal of Forecasting, 38(3), 705-871.

Ramos, P., Santos, N., Rebelo, R. (2015), Performance of state space and ARIMA models for consumer retail sales forecasting. Robotics and Computer-Integrated Manufacturing, 34, 151-163.

Ray, S., Lama, A., Mishra, P., Biswas, T., Das, S.S., Gurung, B. (2023), An ARIMA-LSTM model for predicting volatile agricultural price series with random forest technique. Applied Soft Computing, 149, 110939.

Reitano, V. (2018), An open systems model of local government forecasting. The American Review of Public Administration, 48(5), 476-489.

Salisu, A.A., Omotor, D.G. (2024), Forecasting expenditure components in Nigeria. Journal of Economic Studies, 51(4), 783-807.

Santoro, D., Ciano T, Ferrara M. (2024), A comparison between machine and deep learning models on high stationarity data. Scientific Reports, 14, 19409.

Shakila, C.V., Babu, S.K. (2025), Evaluating and contrasting machine learning and statistical techniques for time series forecasting with hyperparameter optimization. Results in Engineering, 27, 106275.

Shiraishi, H., Nakamura, T., Shibuki, R. (2024), Time series quantile regression using random forests. Journal of Time Series Analysis, 45(4), 639-659.

Sica, E., Altinbas, H., Marini, G. G. (2024), Public debt forecasts and machine learning: The Italian case. Journal of Economic Studies, 51(6), 1355-1370.

Sofianos, E., Betti, T., Papadimitriou, T., Barbier-Gauchard, A., Gogas, P. (2026), Using DSGE and machine learning to forecast public debt for France. Journal of Forecasting, 45, 2173-2185.

Tay, F.E.H., Cao, L. (2001), Application of support vector machines in financial time series forecasting. Omega, 29(4), 309-317.

Tightiz, L., Yoo, J., Al-Shibli, W.K. (2025), Strategic enhancements in electricity price forecasting: The role of XGBoost and error correction features. Results in Engineering, 26, 105609.

Vuong, P.H., Phu, L.H., Nguyen, T.H.V., Duy, L.N., Bao, P.T., Trinh, T.D. (2024), A bibliometric literature review of stock price forecasting: From statistical model to deep learning approach. Science Progress, 107(1), 1-31.

Wang, J., Qin, Z., Hsu, J., Zhou, B. (2024), A fusion of machine learning algorithms and traditional statistical forecasting models for analyzing American healthcare expenditure. Healthcare Analytics, 5, Article 100312.

Wang, J., Xu, S., Wang, Y., El Bouhissi, H. (2025), An analytics framework for healthcare expenditure forecasting with machine learning. Healthcare Analytics, 8, Article 100428.

Williams, D.W., Kavanagh, S.C. (2016), Local government revenue forecasting methods: Competition and comparison. Journal of Public Budgeting, Accounting and Financial Management, 28(4), 488-526.

Yang, C., Wu, J. (2026), Forecasting tourism demand with search query: Comparison of social media and search engine. Tourism Review, 81(1), 147-166.

Yang, C.H., Molefyane, T., Lin, Y.D. (2023), The forecasting of a leading country's government expenditure using a recurrent neural network with a gated recurrent unit. Mathematics, 11(14), Article 3085.

Zhang, G.P. (2003), Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175.