AIMS Mathematics, 11(11): 38721–38761.
DOI: 10.3934/math.20261494
Received: 10 July 2026
Revised: 10 October 2026
Accepted: 21 October 2026
Published: 19 November 2026

https://www.aimspress.com/journal/Math

Research article

Forecasting Saudi Arabia's monthly consumer price index using hybrid support vector regression models: Evidence from SARIMA-SVR and GM(1,1)-SVR

Zahrah Fayez Althobaiti*

Department of Statistics, Faculty of Science, University of Tabuk, 71491, Tabuk, Saudi Arabia

* **Correspondence:** Email: zalthebiti@ut.edu.sa.

Abstract: Accurate short-term forecasting of the consumer price index (CPI) is useful for monitoring changes in the general price level. This study compares three standalone forecasting models—seasonal autoregressive integrated moving average (SARIMA), grey model GM(1,1), and support vector regression (SVR)—with two residual correction hybrids, SARIMA–SVR and GM(1,1)–SVR, using 108 monthly Saudi CPI observations from January 2016 to December 2024. The data are divided chronologically into a 2016–2023 training period and a January–December 2024 out-of-sample holdout. Model selection, scaling, lag selection, and hyperparameter tuning are restricted to pre-2024 information, and all forecasts are generated by rolling one-step-ahead evaluation. A training-only seasonal differencing specification check resolves the near-boundary seasonal moving average estimate in the earlier SARIMA by selecting SARIMA(1,1,1)(0,0,1)(12). Forecast performance is evaluated using root mean square error (RMSE), mean absolute percentage error (MAPE), Theil's U, and Harvey–Leybourne–Newbold-adjusted Diebold–Mariano tests. GM(1,1)–SVR has the lowest 2024 RMSE (0.1421; MAPE = 0.1062%; U = 0.6427), while SARIMA–SVR has the lowest MAPE (0.1044%; RMSE = 0.1443; U = 0.6528); the difference between these two hybrids is not statistically significant (adjusted p = 0.849). Residual correction produces a large and stable gain for the weak GM(1,1) base, whereas the gain for SARIMA is smaller and not statistically established. The results therefore support residual predictability as an empirical screening principle for deciding whether a correction stage is worth adding, rather than supporting the universal superiority of the hybrids. Because the final Saudi holdout contains only 12 observations and the models are univariate, the findings remain short-horizon forecasting evidence rather than direct policy guidance.

Keywords: consumer price index (CPI); support vector regression (SVR); SARIMA; grey model GM(1,1); hybrid forecasting models; inflation forecasting; Saudi Arabia; vision 2030; Theil's U; Diebold–Mariano test
Mathematics Subject Classification: 60G25, 62M10, 91B70, 91B84

1. Introduction

The consumer price index (CPI) is a central measure of inflation, purchasing power, and cost of living dynamics in any economy. In Saudi Arabia, CPI forecasting plays a vital role in shaping fiscal and monetary policy, particularly as the Kingdom advances its Vision 2030 initiative aimed at economic diversification, reduced oil dependency, and sustainable development. With inflation rising to 2.3% in June 2025, up from 1.5% the previous year. Thus, accurate and timely CPI forecasts can provide useful short-term information for monitoring inflation and changes in the general price level [1,2].

Vision 2030 emphasizes economic diversification, reduced dependence on oil revenues, and sustainable economic development. These objectives require strong economic monitoring tools to anticipate inflation trends, stabilize markets, and guide strategic decision-making. CPI forecasting thus plays a pivotal role in aligning fiscal policies with the vision's long-term goals [3].

Traditional time-series models, such as the seasonal autoregressive integrated moving average (SARIMA) and grey model GM(1,1), have been widely applied to economic forecasting due to their effectiveness in modeling seasonality and their ability to operate with limited data. However, both approaches are constrained by linear assumptions and are often unable to capture the complex, nonlinear structures found in real-world CPI data, particularly under conditions of structural economic transformation. These limitations have led to the growing use of machine learning methods, especially support vector regression (SVR), which offers strong generalization ability and robustness in modeling nonlinear relationships [4–7].

Recent research has increasingly examined combinations of classical statistical methods and machine learning techniques. Although some hybrid models report lower forecasting errors, such findings do not imply that hybridization is automatically beneficial: A second-stage learner can improve a forecast only when the base model leaves a residual structure that is predictably exploitable out of sample. This conditional interpretation motivates the present comparison [8].

Despite growing interest in forecasting inflation, limited research has examined residual correction hybrids for Saudi Arabia's monthly CPI under a strictly chronological evaluation design. Existing Saudi studies have mainly relied on standalone statistical or machine learning models, while the potential value of using SVR to model predictable residual structure left by SARIMA or GM(1,1) remains underexplored in the Saudi context.

This study addresses this gap by developing and comparing two residual correction hybrid models, SARIMA–SVR and GM(1,1)–SVR, against their standalone counterparts. The analysis uses official monthly CPI data from January 2016 to December 2024, with January 2016 to December 2023 used for model estimation, and January to December 2024 reserved for out-of-sample evaluation. Forecasting performance is assessed using root mean square error (RMSE), mean absolute percentage error (MAPE), Theil's U, and the Diebold–Mariano (DM) test to evaluate both error magnitude and statistical differences in predictive accuracy.

The main objective of this study is to evaluate the conditional value of residual correction for short-horizon Saudi CPI forecasting. Under a strictly chronological protocol, the study asks which base-model residual processes contain a reproducibly predictable structure that is sufficient to support a correction stage, whether a chronologically trained residual learner translates that predictability into out-of-sample forecast improvement, and whether any numerical gain is statistically distinguishable from strong standalone and naive benchmarks.

1.1. Contributions and novelty

First, the study provides a Saudi-specific empirical comparison of standalone and residual correction forecasting models for monthly CPI using recent official data. The contribution is empirical and procedural rather than a claim of a new forecasting theory: The study applies a transparent rolling-origin evaluation framework to compare classical, machine learning, and hybrid structures under Saudi Arabia's recent CPI dynamics.

Second, the study evaluates two established residual correction hybrid structures, SARIMA–SVR and GM(1,1)–SVR, under a common leakage-controlled forecasting protocol. The novelty does not lie in inventing the additive hybrid architecture itself; rather, it lies in testing these structures on Saudi monthly CPI while enforcing genuine rolling-origin residual generation, time-series cross-validation, comparison with naïve benchmarks, and small-sample-adjusted forecast accuracy testing.

Third, the study moves beyond numerical error ranking by combining RMSE and MAPE with Theil's U against a naive benchmark, small-sample-adjusted DM tests, and a formal structural break assessment. This evaluation design allows the study to distinguish a lower point forecast error from a statistically established improvement.

Fourth, the findings are interpreted cautiously as short-horizon monitoring evidence rather than as a new macroeconomic decision model. The study therefore positions its contribution as a reproducible Saudi CPI forecasting comparison and validation exercise, not as a new theoretical hybrid model architecture.

However, transferable findings revealed that the revised analysis does not propose hybridization as a universally superior architecture. Instead, it formulates an empirical screening principle: residual correction is worth considering when the pre-holdout residuals contain reproducibly predictable structure, and a chronologically evaluated correction model improves on a zero-correction benchmark. Conventional diagnostics such as auto-correlation function (ACF), Ljung–Box, and Brock-Dechert-Scheinkman (BDS) test are supporting evidence rather than formal proofs of unpredictability. In the final specification, GM(1,1) produces strongly predictable residuals and benefits greatly from correction, whereas SARIMA shows weak conventional residual dependence but a smaller, less stable correction gain. The evidence therefore supports a conditional base-model-specific interpretation rather than a universal rule.

The remainder of this paper is structured as follows. Section 2 presents a review of prior research on CPI forecasting using traditional and machine learning methods. Section 3 outlines the methodology, data sources, and modeling approaches used in the study. Section 4 discusses the empirical results and a comparative analysis of forecasting accuracy. Section 5 concludes with the key findings, policy implications, and suggestions for future research.

2. Literature review

2.1. CPI forecasting in Saudi Arabia

Several studies have focused on modeling Saudi Arabia's CPI using both traditional time-series techniques and machine learning. Nyoni [9], for example, built an autoregressive integrated moving average (ARIMA) model using annual Saudi CPI data (1963–2017), finding an ARIMA(0,2,1) to be optimal and projecting relatively high inflation in the following decade. Mahgoub and Alam [10] later applied a seasonal ARIMA to Saudi monthly CPIs (2013–2020); their best model, SARIMA(0,1,0)(12,0,12), effectively captured the CPIs' seasonal pattern and yielded stable forecasts with a steady upward trajectory for 2020–2021.

More recently, researchers have explored machine learning methods for exploring Saudi CPIs. Sibai et al. [11], evaluated a range of machine learning algorithms, including SVR, decision trees, k-nearest neighbors, linear regression, neural networks, and random forests, using Saudi CPI data from 2013 to 2020. They reported that a neural network produced the most accurate predictions, followed by k-nearest neighbors. Notably, all the machine learning models, including SVR, outperformed a simple trend-extrapolation baseline. These studies indicate that while ARIMA models provide a strong baseline, modern machine learning techniques such as SVR can improve accuracy by capturing nonlinear trends, motivating the integration of machine learning components into forecasting frameworks.

2.2. SARIMA and SVR approaches internationally

Internationally, a substantial body of literature has applied ARIMA-family models to forecast the CPI or inflation in various countries. For example, ARIMA models have been successfully used for inflation prediction in Nigeria, Bangladesh, and Ghana. These studies confirm that Box–Jenkins ARIMA techniques often yield accurate forecasts for price indices by capturing the underlying linear autocorrelation structure and any seasonality present in the series. However, pure ARIMA models have well-known limitations, notably, an inability to capture nonlinear patterns or sudden structural changes in economic data.

To overcome these limitations, researchers have increasingly used machine learning. For instance, [12] compared ARIMA with machine learning models for US inflation and found that machine learning models outperformed ARIMA in accuracy. SVR, the regression form of support vector machine (SVM), has emerged as a strong tool for time-series prediction, with studies showing that it can outperform even generalized autoregressive conditional heteroskedasticity models in short-term forecasting. These findings highlight SVR's strength in modeling nonlinear relationships and have supported the rise of hybrid approaches combining classical and machine learning models.

Several studies demonstrate that combining models often improves forecasting accuracy. Lawrence et al. [13] and Ashton and Ashton [14] showed that aggregated forecasts perform better than individual ones. Hybrid models, such as those proposed by [15–17], often merge ARIMA or SARIMA with artificial neural networks or SVR. These studies establish a basis for exploring SARIMA–SVR combinations.

2.3. Grey and hybrid modeling approaches

Grey forecasting models have also been explored for CPI prediction, particularly in contexts with limited data. The GM(1,1) model, which requires only a short historical series, has, in some cases, produced competitive forecasts for economic indices. For example, one study in China found that a GM(1,1) model achieved prediction accuracy on par with an ARIMA model for short-term CPI forecasting. To capitalize on different modeling strengths, [18] proposed a novel "divide and conquer" approach for China's CPI that integrated an ARIMA component to capture linear trends, a GM(1,1) model to handle small-sample trends, and a backpropagation neural network for complex nonlinear patterns. This hybrid system improved forecasting accuracy over any single model alone, illustrating that combining complementary models can yield more robust CPI predictions.

Hybrid forecasting combines distinct modeling components, whereas residual correction is a more specific architecture in which a second model learns from the errors left by a base forecast. Pai and Lin [19] combined ARIMA with SVR and reported improved forecasting performance relative to the component models. Such evidence motivates residual learning but it does not imply that the remaining base model residual is intrinsically nonlinear. In the present study, SVR is used as a flexible residual learner for potentially predictable error structures, which may arise from omitted linear structures, trend misspecification, nonlinear dependence, parameter estimation errors, or other systematic sources.

Grey-based hybrids have also produced useful empirical results. Zhao et al. [20] combined GM(1,1) with a support vector approach for electricity forecasts, while [18] combined grey and machine learning components for forecasting CPI. These studies motivate testing complementary model components, but their positive results are not treated here as evidence that hybridization is universally beneficial or that the residual structure is necessarily nonlinear. The incremental value of a correction stage must instead be established through chronological out-of-sample evaluation.

In recent years, forecasting studies have increasingly adopted deep learning and hybrid methods in response to global volatility. For example, [21] compared forecasting models for Ecuador's CPI, while [22] and [23] applied advanced models such as long short-term memory (LSTM), gated recurrent unit (GRU), and extreme gradient boost (XGBoost) to CPI in the US and Turkey. These studies confirm that no single model is universally best; hybrid frameworks generally achieve better performance by integrating complementary model strengths.

2.4. Recent methodological developments

Recent CPI forecasting research further supports the need to compare classical, machine learning, and hybrid structures rather than assuming that one model class is universally superior. Sibai et al. [11] compared six machine learning methods for the Saudi Arabian CPI and found that performance depended materially on model choice and tuning. Gur [22] evaluated LSTM, multivariate adaptive regression splines (MARS), XGBoost, and hybrid LSTM-based combinations for the US CPI and showed that hybridization can improve forecasting when complementary predictive information is present. Şimşek [23] similarly compared machine learning and deep learning approaches for the Turkish CPI using macroeconomic predictors and reported strong performance for SVR. These studies motivate empirical comparison, but their positive hybrid results are not treated here as a universal mechanism. The present study instead tests the conditional proposition that a residual learner can help

only when the base model's error process contains reproducibly predictable information. Table 1 summarizes the selected studies and their remaining research gaps.

Table 1. Comparative positioning of selected CPI and hybrid forecasting studies.

Study	Setting	Methods	Evaluation focus	Relevance/remaining gap
[9]	Saudi Arabia; annual CPI	ARIMA	Long-horizon univariate forecast	Saudi benchmark, but annual frequency and no hybrid residual correction.
[10]	Saudi Arabia; monthly CPI	SARIMA	Seasonal fit and forecast	Monthly Saudi evidence, but no machine learning residual correction comparison.
[11]	Saudi Arabia; CPI 2013–2020	Six machine learning models incl. SVR	Model comparison accuracy	Strong machine learning benchmark; no rolling residual correction hybrid.
[18]	China; CPI	ARIMA + grey model (GM) + backpropagation neural network (BPNN)	Hybrid forecast accuracy	CPI hybrid precedent; different architecture and evaluation setting.
[22]	United States; CPI	LSTM, MARS, XGBoost, hybrids	Machine learning/hybrid comparison	Shows that hybrid benefits are model- and information-dependent.
Present study	Saudi Arabia; monthly CPI 2016–2024	SARIMA, GM(1,1), SVR; residual hybrids	Rolling origin, naive benchmark, Harvey–Leybourne–Newbold-adjusted DM, break test, residual predictability	Tests when residual correction is useful rather than assuming the hybrid's superiority.

2.5. Positioning against prior CPI studies

The literature indicates three unresolved issues that motivate the present study. First, several CPI forecasting studies rely on annual data, relatively short monthly samples, or evaluation designs that do not fully preserve temporal ordering for strict out-of-sample assessment. Second, both standalone machine learning models and hybrid models can perform well, so hybridization should not be presumed to be superior; its incremental value must be tested against tuned standalone benchmarks and the predictability of the base residual process. Third, many studies emphasize numerical error reduction without examining whether the improvement is statistically meaningful or better than a simple persistence benchmark. The present study addresses these issues by using monthly Saudi CPI data, generating genuine rolling-origin residuals, tuning SVR models by time-series cross-validation, comparing standalone and residual correction models, and evaluating forecasts using RMSE, MAPE, Theil's U, small-sample-adjusted DM tests, and pre-holdout residual predictability diagnostics.

In summary, the research gap addressed here is not the absence of hybrid forecasting as a general concept. Rather, it is the limited evidence on when established residual correction hybrids add value for Saudi monthly CPI forecasts under a strict chronological protocol, relative to tuned standalone

models and a naive benchmark. The study therefore contributes a controlled empirical evaluation and a cautious residual predictability screening principle rather than a new hybrid forecasting theory.

3. Materials and methods

3.1. Data description

Monthly Saudi Arabia CPI observations were obtained from the official General Authority for Statistics (GASTAT) CPI series for January 2016–December 2024. The target is the national monthly CPI index reported on the study's stated base-year scale of 2018 = 100, yielding 108 observations. The sample is divided chronologically into January 2016–December 2023 for model development and January–December 2024 for out-of-sample evaluation. The manuscript reports the responsible statistical agency, series definition, frequency, base year, sample period, and training–testing split. Because the source files supplied for this revision do not contain a persistent GASTAT table identifier or direct series URL, no unverified URL or series code has been added. Table 2 encapsulates the study's data characteristics and forecasting design.

Table 2. Data description and forecasting design.

Item	Description
Target variable	Saudi Arabia monthly CPI
Source	GASTAT, Saudi Arabia
Frequency	Monthly
Base year	2018 = 100
Full sample	January 2016–December 2024
Number of observations	108
Training period	January 2016–December 2023
Test period	January 2024–December 2024
Forecasting horizon	One-month-ahead forecasts over the 2024 holdout period

Preprocessing was performed in chronological order. The monthly date index was checked for continuity, the CPI values were checked for missing observations, and no random shuffling was applied. The target CPI remained on its original index scale. For all SVR-based models, predictor variables were standardized using z-score standardization implemented through a standard scaler as follows:

$$x_{t,j}^{*} = \frac{x_{t,j} - \mu_{j,train}}{\sigma_{j,train}},$$

where $\mu_{j,train}$ and $\sigma_{j,train}$ denote the mean and standard deviation of the predictor j estimated only from the corresponding training fold or rolling training window. The same training-derived transformation was then applied to the validation or test observation. Only predictor variables were standardized. Neither the standalone CPI target nor the GM(1,1) and SARIMA residual targets were rescaled before fitting SVR; all targets remained in their original CPI or residual units. Accordingly, ε is interpreted in the native units of each target, and the model-specific ε values were selected separately by time-series cross-validation rather than imposed as a common tolerance across differently scaled targets. The

scaler was never fitted on the 2024 holdout data, and no inverse transformation was required before computing RMSE, MAPE, Theil's U, or the DM statistics.

The first hybrid model, SARIMA–SVR, combines a SARIMA model with SVR. SARIMA provides a structured base forecast for linear and seasonal dependence, while SVR is used as a flexible learner for any residual structure that remains predictably exploitable after the base forecast. Previous hybrid forecasting studies motivated testing this architecture, but the present analysis does not presume that the remaining residual component is nonlinear or that adding SVR must improve the forecast accuracy [24].

The second hybrid model, GM(1,1)–SVR, integrates the grey model GM(1,1) with SVR. GM(1,1) can capture the overall trend of a series using a small sample size without requiring complex computations or strict statistical assumptions. The strategy is for GM(1,1) to generate an initial forecast on the basis of its simple pattern, then for SVR to learn and correct the remaining errors, potentially enhancing forecast accuracy [25].

These two hybrid approaches were selected to test whether integrating machine learning techniques with statistical models can yield measurable improvements in inflation forecasts' accuracy compared with using each model alone.

3.2. Individual models

In this section, we describe the three individual forecasting models used as benchmarks. First, we introduce the GM(1,1) model and explain its basic formulation and strengths. Next, we present the SARIMA model and outline how it captures linear trends and seasonality. Finally, we cover the SVR model and discuss its ability to learn complex nonlinear patterns. Each model's setup and parameter choices are detailed before moving on to the hybrid approaches.

3.2.1. The GM(1,1) grey model

The abbreviation GM(1,1) represents a grey prediction model with one variable and one order. Developing a grey prediction model involves the following general steps.

Phase 1: Establish an initial series based on observed data.

$$x^{(0)} = \left(x^{(0)}(1), x^{(0)}(2), \ldots\ldots, x^{(0)}(n)\right), \tag{1}$$

where indicates the starting point data for the time.

The non-negative sequence is $x^{(0)}$, and the sample size is$n$. The GM(1,1) model can be estimated from a minimum of four observations.

Phase 2: Construct the first-order accumulated generating operation (AGO) series from the original series

$$x^{(1)} = \left(x^{(1)}(1), x^{(1)}(2), \ldots\ldots, x^{(1)}(n)\right), \quad \text{n} \geq 4. \tag{2}$$

The following formula is used to derive $x^{(1)}(k)$

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i). \tag{3}$$

Phase 3: Calculate the background value sequence of the first-order AGO.

The average series generator is described as follows:

$$z^{(1)} = \left(z^{(1)}(1), z^{(1)}(2), \ldots\ldots, z^{(1)}(n)\right).$$

The sequential data's average value $z^{(1)}(k)$ is defined as follows:

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1), \quad k = 2,3,\ldots,n. \tag{4}$$

Phase 4: Consider the following first-order differential equation for the series:

$$\frac{dx^{(1)}}{dt} + a\, x^{(1)} = b. \tag{5}$$

The following is its difference equation:

$$x^{(0)}(k) + a\, z^{(1)}(k) = b. \tag{6}$$

Phase 5: Estimate the parameters, which are calculated using ordinary least squares

$$\hat{a} = [a, b]^T = (B^T B)^{-1} B^T Y. \tag{7}$$

Phase 6: The solution of the grey differential equation under the initial condition $x^{(1)}(1) = x^{(0)}(1)$ yields:

$$\hat{x}^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{b}{a}\right] \mathrm{e}^{-ak} + \frac{b}{a}. \tag{8}$$

Phase 7: Obtain the forecast values using the first-order inverse AGO (IAGO).

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k). \tag{9}$$

3.2.2. SARIMA models

The ARIMA model is one of the most widely used statistical models for time series forecasts. Its basic principle is to transform a nonstationary time series into a stationary series through differencing when required. As a result, the dependent variable will be described as a model that only yields its lag value, as well as the actual and lag values of the random error term. The following are the steps in the prediction phase.

Phase 1: Difference the time series as needed to achieve stationarity. In theory, stationarity ensures that the fitted curve formed by sampling a time series can continue inertially along the present form in the future, i.e., the data's mean and variance should not be significantly changed.

Phase 2: Construct an autoregressive (AR) model. The autoregressive model is a method of forecasting itself using the variable's historical result data, and it illustrates the link between current value and

previous value. It has the following formula:

$$y_t = \mu + \sum_{i=1}^{p} \phi_i \, y_{t-i} + \varepsilon_{t,} \tag{10}$$

where y_t represents the current value, indicates the constant term, denotes the order, is the autocorrelation coefficient, and represents the error.

Phase 3: Construct a moving average (MA) model. In the autoregressive model, the moving average model concentrates on the accumulation of error components. The MA component models dependence on past forecast errors. It has the following formula:

$$y_t = \mu + \sum_{i=1}^{q} \theta_i \, \varepsilon_{t-i} + \varepsilon_t, \tag{11}$$

where θ_i is the MA formula's correlation coefficient.

Phase 4: Construct an autoregressive moving average (ARMA) model by combining the AR and MA components. The following is the exact formula. The orders of the AR and MA models, respectively, are p and q in this formula. The correlation coefficients of the two models, ϕ_i and θ_i respectively, must be solved.

The seasonal ARIMA model belongs to a family of flexible linear time series models that can be used to model many different types of seasonal as well as nonseasonal time series.

Box and Jenkins generalized the ARIMA model to deal with seasonality. Their suggested model is known as the SARIMA model. In the most popular multiplicative form, the SARIMA model can be expressed as follows:

$$\Phi(B^s)\,\varphi(B)\,(1-B)^d\,(1-B^s)^D\,y_t = \Theta(B^s)\,\theta(B)\,\varepsilon_t\,, \tag{12}$$

where $\varphi(B)$ and $\theta(B)$ are the nonseasonal AR and MA polynomials, $\Phi(B^s)$ and $\Theta(B^s)$ are the seasonal AR and MA polynomials, d and D are the nonseasonal and seasonal differencing orders, s is the seasonal period, y_t is the CPI series, and ε_t is a zero-mean white noise error term. The revised SARIMA specification contains no unconditional drift term.

For SARIMA$(p,d,q)(P,D,Q)_s$, p and q are the nonseasonal AR and MA orders, P and Q are the seasonal AR and MA orders, d and D are the nonseasonal and seasonal differencing orders, and s is the seasonal period (s = 12 for monthly CPI).

Seasonal differencing subtracts the observation from the value one seasonal period earlier and is defined as

$$\nabla_s = x_t - x_{t-s}. \tag{13}$$

With s = 4, as in quarterly data, the seasonal difference is

$$(1-B^4)x_t = x_t - x_{t-4.} \tag{14}$$

With s = 12, as in monthly data, the seasonal difference is

$$(1-B^{12})x_t = x_t - x_{t-12}. \tag{15}$$

A SARIMA model reduces to pure ARIMA(p,d,q) when there is no seasonal influence, while a

pure ARIMA model reduces to ARMA(p,q) when the time series dataset is stationary. The behavior of the ACF and partial autocorrelation function (PACF) in pure SARMA models are displayed in Table 3.

Table 3. General behavior of the ACF and PACF for pure SARMA models.

	$AR_s(P)$	$MA_s(Q)$	$ARMA_s(p,q)$
ACF	Tails off after lags ks, $k = 1,2,\ldots$	Cuts off after lags Qs	Tails off at lags ks
PACF	Cuts off after lags Ps	Tails off at lags ks	Tails off at lags ks

After estimating the SARIMA model, diagnostic tests are conducted to ensure its reliability and validity. First, we assess the significance of the predicted parameters and residuals derived from the fitted SARIMA model. This involves examining whether the estimated coefficients and residuals exhibit statistical importance, indicating their reliability in capturing the underlying patterns in the data.

Second, we validate the model's assumptions by comparing the actual observations with the values predicted by the candidate model. This validation process helps determine the model's accuracy and adequacy in representing the observed data. If the model passes these diagnostic tests, it is considered to be suitable for generating future forecasts. However, if the model fails to meet the criteria set by these tests, indicating potential inadequacies or deficiencies in its specification, we initiate a corrective process.

This involves setting up a new model and repeating the estimation process, diagnostic tests, and validation steps outlined above. By iteratively refining the model according to diagnostic feedback, our goal is to develop a robust and reliable forecasting framework that accurately captures the underlying dynamics of the data.

3.2.3. The SVR model

SVR is the regression form of SVM. In regression tasks, the output is a real or continuous numerical value rather than a class label. SVR is useful for time-series forecasting because it can control overfitting through the ε-insensitive loss function and the regularization parameter C, which supports good generalization performance [26,27]. Unlike classification SVM, which separates observations into classes, SVR estimates a regression function that keeps the prediction errors within an ε tube while penalizing larger deviations. This makes SVR suitable for modeling nonlinear CPI movements that cannot be fully captured by purely linear statistical models [28].

SVR can also be interpreted geometrically as a regression hyperplane surrounded by an ε-insensitive tube. Observations near or outside the tube act as support vectors, and the objective is to estimate a regression function that minimizes the forecasting error while preserving generalization ability. This geometric interpretation helps explain why SVR is effective for modeling nonlinear CPI patterns in a controlled and flexible way as W^T is the weight vector l dimension, where $\varphi(x_i)$ is a function that maps to the feature space with l dimensions, b is the bias term.

The function of $W^T\varphi(x_i) - y_i + b \leq \varepsilon$ shows a point in the higher dimensional feature space, which is the result of the mapping of the input vector in the lower dimensional input space.

The coefficients are estimated by minimizing the risk function. Therefore, to maximize the margin,

a minimum is needed. Optimization of problem-solving can be carried out as shown in Function (16) with the condition

$$y_i - W^T\varphi(x_i) - b \leq \varepsilon, for\ i = 1,\ldots,l\ and\ W^T\varphi(x_i) - y_i + b \leq \varepsilon, for\ i = 1,\ldots,l,$$

where y_i is the actual value of period i, and $W^T\varphi(x_i) - b$ is the estimated value of period i.

The term controls regularization. Minimizing will make a function as thin (flat) as possible so that it can control the functional capacity. All points outside the margin/limit ε will be penalized. Furthermore, the optimization problem above can be formulated into the function shown in (16)

$$min\frac{1}{2}\| W\|^2+\mathrm{C}\sum_{i=1}^{\ell}(\zeta_i + \zeta_i^*), \tag{16}$$

with the condition

$$y_i - W^T\varphi(x_i) - b - \zeta_i \leq \varepsilon, for\ i = 1,\ldots,l,$$
$$W^T\varphi(x_i) - y_i + b - \zeta_i^* \leq \varepsilon, for\ i = 1,\ldots,l,$$
$$\zeta_i\ ,\zeta_i^* \geq 0.$$

The constant $C > 0$ determines how far the error deviation is from the tolerable limit $\varepsilon\varepsilon$. The formula above is a convex quadratic optimization problem which functions to minimize the quadratic function to be converted into a constraint. This limitation can be solved by using the Lagrange multiplier function. The dual formulation is obtained using Lagrange multipliers. After going through mathematical stages, a new equation is obtained with the function

$$f(x) = \sum_{i=1}^{l}(a_i - a_i^*)\cdot(x_i \cdot x) + b, \tag{17}$$

where the support vector is the test vector. The linear formulation can be extended to nonlinear problems by applying a kernel function. For nonlinear problems, the values of xi and x are first transformed into a high-dimensional feature space by mapping the vectors and into the kernel function so that the final function becomes

$$f(x) = \sum_{i=1}^{l}(a_i - a_i^*)\cdot K(x_i \cdot x) + b. \tag{18}$$

In our SVR implementation, we focus exclusively on the radial basis function (RBF) kernel. For reference, Table 4 lists the common linear and RBF kernels.

Among the various kernel functions, the RBF kernel is widely used in nonlinear regression, since it handles nonlinear relationships without requiring prior knowledge of the data's structure. Thus, we adopt only the RBF form in this study because of its ability to represent nonlinear relationships in monthly CPI data.

Table 4. Common kernel functions for SVR.

Kernel	Function
Linear	$K\left(\vec{x_i}, \vec{x_j}\right) = \left(\vec{x_i}, \vec{x_j}\right)$
Radial basis function	$K\left(\vec{x_i}, \vec{x_j}\right) = exp\left(-\frac{\left\|\vec{x_i} - \vec{x_j}\right\|^2}{2\sigma^2}\right)$

3.3. Hybrid model construction

The additive hybrid framework follows a residual correction architecture. At each forecast origin, the base model first produces a one-step-ahead CPI forecast. The difference between the observed CPI and the corresponding genuine rolling-origin base forecast forms the residual series. SVR is then refitted on the lagged residuals available up to that origin to estimate the potentially predictable residual structure that remains after the base model forecast. The hybrid forecast is obtained by adding the predicted SVR residual correction to the base model's forecast. The residual is not assumed to be intrinsically nonlinear; it may reflect an omitted linear structure, trend misspecification, nonlinear dependence, parameter estimation error, or unpredictable noise. The SVR component therefore models whatever predictable residual signal is present without assigning a nonlinear interpretation a priori.

For the GM(1,1)–SVR model, the GM(1,1) one-step-ahead forecast is generated first. The corresponding base model residual is defined as

$$e_t^{\mathrm{GM}} = y_t - \hat{y}_t^{\mathrm{GM}}, \tag{19}$$

where y_t is the observed CPI value and $\hat{y}_t^{GM}$ is the GM(1,1) forecast. SVR uses lagged genuine rolling-origin residuals to estimate the residual correction $\hat{e}_t^{sVR}$. The final hybrid forecast is

$$\hat{y}_t^{\mathrm{GM-SVR}} = \hat{y}_t^{\mathrm{GM}} + \hat{e}_t^{\mathrm{SVR}}. \tag{20}$$

For the SARIMA–SVR model, the SARIMA one-step-ahead forecast is generated first, and its rolling-origin residual is defined as

$$r_t^{\mathrm{SARIMA}} = y_t - \hat{y}_t^{\mathrm{SARIMA}}, \tag{21}$$

where $\hat{y}_t^{sARIMA}$ is the SARIMA forecast. SVR then predicts the residual correction $\hat{r}_t^{sVR}$ from the lagged genuine rolling-origin SARIMA forecast residuals. The final SARIMA–SVR forecast is

$$\hat{y}_t^{\mathrm{SARIMA-SVR}} = \hat{y}_t^{\mathrm{SARIMA}} + \hat{r}_t^{\mathrm{SVR}}. \tag{22}$$

Equations (19)–(22) use one consistent notation for both hybrid models: The base forecast is generated first, and the SVR-predicted residual correction is then added to obtain the final one-step-ahead hybrid forecast.

3.3.1. Residual correction algorithm

For each forecast origin *t*, the residual correction procedure is implemented sequentially. The

mathematical quantities used in the algorithm are defined as follows.

Standalone base forecast definition: At the forecast origin t, the one-step-ahead base forecast is defined explicitly as

$$\hat{y}_{t|t-1}^{base} = M_{base}(\mathcal{F}_{t-1}),$$

where M_{base} denotes the selected base model (GM(1,1) or SARIMA) fitted using only the information set $\mathscr{F}_\{t-1\}$. The base forecast is formed before z_t, and the SVR residual correction and the hybrid forecast are computed.

$$z_t = (e_{t-l_1}, \ldots, e_{t-l_k})^{\mathrm{T}}.$$

$$\hat{e}_t^{SVR} = f_{SVR}(z_t).$$

$$\hat{y}_t^{hybrid} = \hat{y}_{t|t-1}^{base} + \hat{e}_t^{SVR}.$$

$$e_t = y_t - \hat{y}_{t|t-1}^{base}.$$

The base forecast is generated using information available through $t-1$. After the actual value y_t becomes available, the new base model residual e_t is appended to the residual history and becomes available only for subsequent forecast origins. This ordering prevents use of the current period's residual before the corresponding forecast has been made.

Forecast risk basis and residual predictability principle

Let $e_t = y_t - \hat{y}_t^{base}$ denote the one-step-ahead error from a base model and let $\hat{e}_t$ denote a residual correction predicted from information available at $t-1$. The hybrid error is $e_t - \hat{e}_t$. Under squared-error loss, the difference in mean squared forecast risk is

$$MSE_{Hybrid} - MSE_{Base} = E[\hat{e}_t^2] - 2E[e_t\,\hat{e}_t]. \tag{23}$$

Therefore, residual correction improves the base forecast if and only if

$$2E[e_t\,\hat{e}_t] > E[\hat{e}_t^2]. \tag{24}$$

This condition makes it clear that a correction is useful only when it is sufficiently aligned with the future base-model error.

$$m_t = E[\,e_t \mid \mathcal{F}_{t-1}\,]. \tag{25}$$

Writing the optimal predictable component as m_t and using the orthogonality of $e_t - m_t$ to all $t-1$ information gives

$$MSE_{Base} - MSE_{Hybrid} = E[m_t^2] - E[(\hat{e}_t - m_t)^2]. \tag{26}$$

Hence, the predictable signal energy in the residuals must exceed the estimation error of the correction model. This identity motivates an empirical prehybrid screening principle rather than a formal diagnostic equivalence: Before adding an SVR correction, we must assess whether genuine

rolling-origin base model residuals show reproducible out-of-sample predictability under chronological validation. ACF, Ljung–Box, and BDS statistics are descriptive diagnostics, and a nonsignificant result does not prove that $E[e_t|\mathcal{F}_{t-1}] = 0$. The most direct empirical check used here is whether a chronologically refitted residual learner reduces residual RMSE relative to a zero-correction benchmark.

Here $m_t = E[e_t|F_{t-1}]$ denotes the theoretical conditional expectation, i.e., the optimal predictable component under squared-error loss. The SVR output $\hat{e}_t$ is only an estimated correction constructed from finite training data and is not identified with m_t itself; its usefulness depends on how accurately it approximates that predictable component out of sample.

3.3.2. Analysis of the GM(1,1)–SVR hybrid model

Below are the main steps used to construct and evaluate the GM(1,1)–SVR hybrid, from rolling base forecasts and genuine one-step-ahead forecast residuals to monthly SVR refitting and final CPI prediction.

(i) Data preprocessing: Split into training (2016–2023) and test (2024) sets.

(ii) GM(1,1)'s initial forecast: Accumulate the raw CPI series to form the "1-AGO" sequence.

- Estimate the GM(1,1)'s whitening equation parameters (a, b) via least squares.
- Generate the time-response function and compute the GM(1,1)'s fitted values for the training set.
- Produce the GM(1,1) one-step-ahead forecasts for both the training and test periods.

(iii) Residual calculation: Compute the genuine rolling-origin one-step-ahead GM(1,1) forecast residuals.

$$e_t = y_t - \hat{y}_t^{GM}, \tag{27}$$

where y_t is the actual CPI and $\hat{y}_t^{GM}$ is the GM(1,1) one-step-ahead forecast produced using information available in $t-1$.

(iv) SVR residual modeling: Construct the input features from lagged rolling-origin residuals.

- The RBF kernel can represent flexible mappings among lagged residuals, but the residual signal is not assumed to be nonlinear. A price shift first appears as a base model forecast error; once observed, that error enters the residual history available at the next forecast origin. The monthly refitted SVR maps the updated residual history to a correction for the following month. This mechanism can exploit persistent predictable deviations after they appear, but it cannot anticipate an unobserved shock before information about that shock becomes available.
- At each forecast origin, refit the SVR residual correction model on the expanding set of genuine rolling-origin residuals available before the target month. Lag structures and hyperparameters are fixed from chronologically ordered time-series cross-validation within 2016–2023; random cross-validation is not used.
- Use the monthly refitted SVR to predict the residual correction for the next forecast origin.

(v) Hybrid forecast generation: For each forecast horizon in the test set, do the following:

- Compute the GM(1,1) forecast $\hat{y}_t^{GM}$;
- Generate the SVR-predicted residual $\hat{e}_t$;

- Combine

$$\hat{y}_t^{SVM-GM} = \hat{y}_t^{GM} + \hat{e}_t. \tag{28}$$

(vi) Performance evaluation: Compute the error metrics (RMSE, MAPE) for the following:

- GM(1,1) alone;
- SVR alone (trained directly on CPI data);
- Hybrid GM(1,1)–SVR forecast.

Compare the metrics to quantify the improvement gained by the hybrid.

3.3.3. Analysis of the SARIMA–SVR hybrid model

The SARIMA–SVR hybrid is constructed sequentially. SARIMA first generates a one-step-ahead forecast using the information available in the preceding month. The corresponding genuine rolling-origin SARIMA forecast residuals are stored through time. Lagged residuals are then supplied to an SVR that is refitted monthly on the expanding residual history and predicts a potentially predictable residual component for the next forecast. The final forecast is the SARIMA forecast plus the SVR-predicted correction. After the actual CPI becomes available, the new SARIMA forecast residual is appended for use at the subsequent origins.

(i) Data preprocessing: Split into training (2016–2023) and test (2024) sets.

(ii) SARIMA initial forecast: Fix the final SARIMA order using training-only stationarity and seasonal specification checks, then re-estimate its coefficients at each rolling forecast origin.

- Fit the SARIMA model on the training set to capture the linear and seasonal structure.
- Generate one-step-ahead SARIMA forecasts for both the training and test periods.

(iii) Residual calculation: Compute genuine rolling-origin one-step-ahead SARIMA forecast residuals

$$r_t = y_t - \hat{y}_t^{SARIMA}, \tag{29}$$

where y_t is the actual CPI and $\hat{y}_t^{sARIMA}$ is the one-step-ahead SARIMA forecast produced using information in $t-1$.

(iv)SVR residual modeling: Create feature vectors from the lagged genuine rolling-origin residuals.

- Choose the RBF kernel and select the residual lag structure and SVR hyperparameters using chronologically ordered time-series cross-validation restricted to the 2016–2023 training period.
- Refit the SVR at each forecast origin on the expanding residual history. The residual lag structure and SVR hyperparameters remain fixed at the values selected by training period's time-series cross-validation

(v) Hybrid forecast generation: For each forecast point in the test set, do the following:

- Compute the SARIMA forecast $\hat{y}_t^{SARIMA}$.
- Predict the next residual correction with the monthly refitted SVR.
- Combine

$$\hat{y}_t^{SVM-SARIMA} = \hat{y}_t^{SARIMA} + \hat{r}_t. \tag{30}$$

(vi) Performance evaluation: Calculate the RMSE and MAPE for the following:

- SARIMA alone;
- SVR alone (trained directly on CPI data);
- Hybrid SARIMA–SVR.

Compare the results to assess how much the hybrid reduces forecast errors.

As illustrated in Figure 1, the construction of the hybrid model comprises six main stages, beginning with data preparation and concluding with performance evaluation.

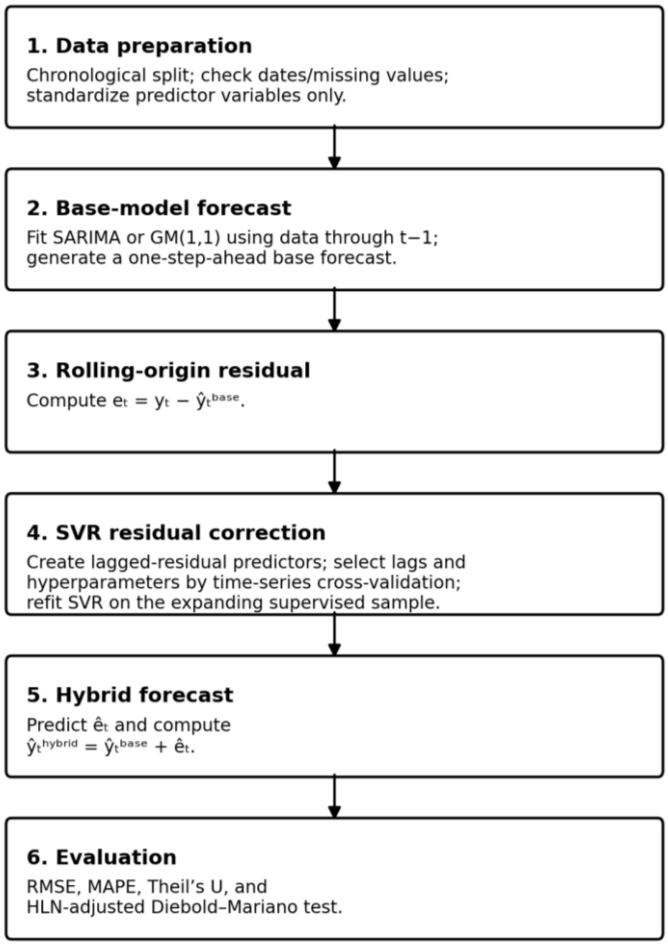


Figure 1. Revised rolling-origin construction of the GM(1,1)–SVR and SARIMA–SVR residual correction models.

3.4. Experimental design and assessment of the models' performance

The experimental design follows a strictly chronological forecasting protocol. Both model selection and tuning are restricted to the January 2016–December 2023 training period, and all reported test results are computed only on the unseen January–December 2024 observations. For the hybrid models, genuine one-step-ahead residuals are generated within the training period using an expanding rolling-origin procedure rather than residuals from a base model fitted to the full training sample. Beginning with an initial 36-month estimation window, each base model is estimated using

observations available only in time $t-1$, a one-step-ahead forecast for time t is produced, and the rolling-origin residual is calculated as described below.

Training-only SARIMA specification choice: All stationarity and seasonal specification diagnostics were recomputed using January 2016–December 2023 only. The level series is nonstationary (ADF = −0.1991, $p = 0.939$; 5% critical value = −2.8925), while the first difference is stationary (ADF = −9.3031, $p < 0.001$; 5% critical value = −2.8928), so $d = 1$ is retained. Because a seasonal ACF spike does not by itself establish a seasonal unit root, $D = 1$ was formally checked against $D = 0$ using the same SARIMA(1,1,1)(0,D,1)[12] structure. With $D = 1$, Akaike information criterion (AIC) = 204.893 and Bayes information criterion (BIC) = 214.568, but the seasonal MA estimate is $\Theta_{12} = -0.9996$, a near-invertibility-boundary warning, and chronological 2021–2023 one-step-ahead RMSE is 0.8164. With $D = 0$, AIC = 210.536, BIC = 220.752, $\Theta_{12} = 0.0992$, and rolling RMSE falls to 0.3297. Because $D = 0$ removes the boundary warning and is substantially better under the prespecified chronological forecasting criterion, the final SARIMA specification is SARIMA(1,1,1)(0,0,1)[12] without unconditional drift. This choice is fixed before the 2024 holdout.

Rolling implementation. The final SARIMA order SARIMA(1,1,1)(0,0,1) [12] is fixed after the pre-2024 selection step, but its coefficients are re-estimated at every forecast origin using the expanding history through month $t-1$. The GM(1,1) parameters a and b are likewise re-estimated at every origin. For SVR, the lag structures and hyperparameters are fixed after the training period's time-series cross-validation, while the SVR model itself is refitted monthly on the expanding supervised sample. After the realized CPI for month t is observed, the new CPI value and genuine base model forecast residual become available only for the next forecast origin. Thus, the fixed hyperparameters are clearly distinguished from monthly model refitting.

$$e_t = y_t - \hat{y}_{t|t-1}.$$

The estimation window is then expanded by one observation and the procedure is repeated. During the 2024 holdout period, forecasts are generated sequentially one month ahead. At each forecast origin, only information available through the preceding month is used; after the actual CPI observation becomes available, it is added to the information set for the next forecast. This design prevents look-ahead bias.

Model selection was implemented entirely within the 2016–2023 training period: (i) Nonseasonal differencing d was selected from the training-only stationarity diagnostics; (ii) the seasonal differencing choice D was checked by comparing $D = 1$ and $D = 0$ under the information criteria, the seasonal MA boundary estimate, and chronological rolling validation; (iii) the GM(1,1) parameters were estimated by ordinary least squares after the AGO; and (iv) the SVR lag structures and hyperparameters were selected by chronologically ordered time-series cross-validation. All scaling parameters were estimated within the corresponding training folds, and the observations for 2024 were not used for specification, feature scaling, lag selection, or hyperparameter tuning.

3.4.1. SVR hyperparameter selection

SVR's performance depends on the penalty parameter C, the ε-insensitive margin ε, the RBF kernel parameter γ, and the lag structure. These quantities have distinct mathematical roles. The parameter C controls the penalty assigned to deviations outside the ε-insensitive tube: larger C places

greater weight on fitting the training errors, whereas a smaller C increases regularization. The parameter ε defines the width of the loss-insensitive tube around the regression function; errors with an absolute magnitude no greater than ε do not contribute to the ε-insensitive loss. For the RBF kernel shown below, γ controls the spatial decay of similarity: Small γ values produce smoother functions with broader influence, while larger γ values allow more localized and flexible nonlinear fits. Because these parameters interact with the scale and serial dependence of the CPI predictors, there is no closed-form parameter combination that is mathematically optimal for this dataset. Therefore, the final values were selected empirically by chronologically ordered time-series cross-validation using validation by RMSE, with all tuning restricted to the 2016–2023 training period

$$K(x_i, x_j) = \exp(-\gamma \, ||x_i - x_j||^2).$$

3.4.2. Reproducibility specification

The revised analysis is fully specified by the following elements: The chronological training period (2016–2023), the unseen 2024 holdout period, rolling one-step-ahead evaluation, a 36-month initial window for generating genuine rolling-origin base model forecast residuals, SARIMA(1,1,1)(0,0,1)[12] without unconditional drift, GM(1,1) estimated by ordinary least squares after the AGO, and RBF-kernel SVR models selected by time-series cross-validation. Predictor scaling for every SVR specification uses training-only z-score standardization. The standalone CPI target and both hybrid residual targets remain unscaled in their native units; only the predictor features are standardized.

The selected SVR specifications are as follows: Standalone SVR with CPI lags (1,12), $C = 100$, $\gamma = 0.01$, and $\varepsilon = 0.05$; SARIMA residual SVR with residual lags (1,2,12), $C = 10$, $\gamma = 0.001$, and $\varepsilon = 0.01$; and GM(1,1) residual SVR with residual lags (1,6), $C = 100$, $\gamma = 0.01$, and $\varepsilon = 0.05$. The lag sets and hyperparameters are fixed from pre-2024 time-series cross-validation, whereas each SVR is refitted at every rolling forecast origin. The monthly predictions, selected settings, and analysis code are supplied as reproducibility materials.

Forecast intervals: To quantify forecasts' uncertainty without using the 2024 holdout observations, empirical 95% forecast intervals were constructed from 40 genuine pre-2024 rolling-origin forecast errors. These 40 errors correspond to complete one-step-ahead model forecasts after satisfying the initial estimation window and lag requirements. At each origin, the base model and, for a hybrid, the residual correction SVR were fitted only with the information available before the target month; the target month's residual was not used in fitting the correction that forecast that same month. The 2.5th and 97.5th percentiles of each model's resulting rolling error distribution were added to the point forecasts for 2024. The intervals are descriptive finite-sample uncertainty bands and should be interpreted cautiously because the pre-2024 error sample is limited.

Forecast performance was assessed using absolute error, relative error, relative benchmark accuracy, and statistical forecast comparison tests. This follows the forecasting evaluation tradition of [29] and the accuracy measure concerns discussed by [30]. RMSE and MAPE measure the magnitude of the forecast error, Theil's U compares each model against a naive random walk benchmark, and the DM test assesses whether error differences between competing models are statistically meaningful. They are calculated as shown in the next section.

3.4.3. Mean absolute percentage error

$$MAPE = \frac{1}{n}\left(\sum_{i=1}^{n}\left|\left(\frac{x_i^{(0)} - \hat{x}_i^{(0)}}{x_i^{(0)}}\right)\right|\right) \times 100\%. \tag{31}$$

MAPE refers to mean absolute percentage error, $\hat{x}_i^{(0)}$ represents the predicted value, $x_i^{(0)}$represents the actual value, and $n$is the total number of data observations.

3.4.4. Root mean square error

Various statistical techniques are used to assess the estimated results in order to determine forecasting accuracy, including the RMSE.

$$RMSE = \sqrt{\sum_{i=1}^{n}\left(x_i^{(0)} - \hat{x}_i^{(0)}\right)^2}/n. \tag{32}$$

The obtained MAPE and RMSE should be as small as possible for an accurate forecast. As indicated in Table 5, [31] provided criteria for MAPE.

Table 5. MAPE accuracy criteria [31].

Criterion (%)	Forecasting power
>50	Inaccurate
20–50	Reasonable
10–20	Good
<10	Highly accurate

3.4.5. Theil's U statistic

Theil's U statistic was used as a relative accuracy measure by dividing each model's test RMSE by the RMSE of a naive random walk forecast. A value below 1 indicates that the model improves on the naive benchmark, while a value above 1 indicates weaker performance than the naive benchmark. This benchmark is important for forecasting CPIs because the CPI is persistent, and a model can produce a low MAPE while still offering limited improvement over a simple previous-month forecast.

3.4.6. The DM test

The DM test compares predictive accuracy using the forecast error's loss differential. Because the 2024 holdout contains only 12 monthly observations, conventional DM inference may be

unreliable in small samples. Accordingly, the revised analysis applies the Harvey–Leybourne–Newbold small-sample adjustment to the DM statistic, following [32], and interprets all significance results cautiously.

For each model pair, the loss differential is defined using squared forecast errors over the 2024 holdout period. The conventional DM statistic is first computed from the mean and variance of the loss differential. For the one-step-ahead horizon, the Harvey–Leybourne–Newbold small-sample correction is written as

$$DM_{HLN} = DM \cdot \frac{\sqrt{(T-1)}}{\sqrt{T}},$$

where $T = 12$ for the 2024 holdout. The corrected statistic is evaluated using Student's t distribution with $T-1$ degrees of freedom. This correction provides a more conservative small-sample comparison, although it cannot overcome the limited power of a 12-observation holdout.

$$d_t = \left(y_t - \hat{y}_{1,t}\right)^2 - \left(y_t - \hat{y}_{2,t}\right)^2, \tag{33}$$

and this forms the DM statistic

$$DM = \frac{\bar{d}}{\sqrt{\hat{\sigma}_{\bar{d}}^2/\mathrm{T}}}. \tag{34}$$

The null hypothesis is equal predictive accuracy between the two competing models. The revised interpretation is based on the small-sample-adjusted p-value rather than the unadjusted p-value.

$$H_0: E[d_t] = 0 \quad \text{vs} \quad H_1: E[d_t] \neq 0.$$

Because only 12 forecast observations are available, statistical significance is treated as supporting evidence rather than as the sole basis for ranking the models.

4. Empirical results

The empirical analysis uses January 2016–December 2023 for model development and January–December 2024 for final holdout evaluation. Figure 2 shows a nonmonotonic CPI trajectory rather than a uniformly rising inflation regime. During 2016, the index was relatively stable around 105–106, then declined through 2017 and fell sharply to 100.9 in January 2018 before reaching a trough near 97.4 in March 2019. A pronounced level shift occurred in mid-2020, after which the index entered a sustained upward phase and reached 112.28 in December 2024. The post-2020 rise therefore characterizes the second half of the sample only; the earlier years contain a substantial decline.

After the mid-2020 level shift, the CPI increased gradually through 2024. From roughly 104 in mid-2020 to 112.28 in December 2024, the average increase was about 0.5 index points per quarter, not 1 point per quarter.

This contrast between the pre-2020 decline and the post-2020 upward regime motivates an evaluation across different temporal regimes rather than describing the full sample as a single smooth upward trend.

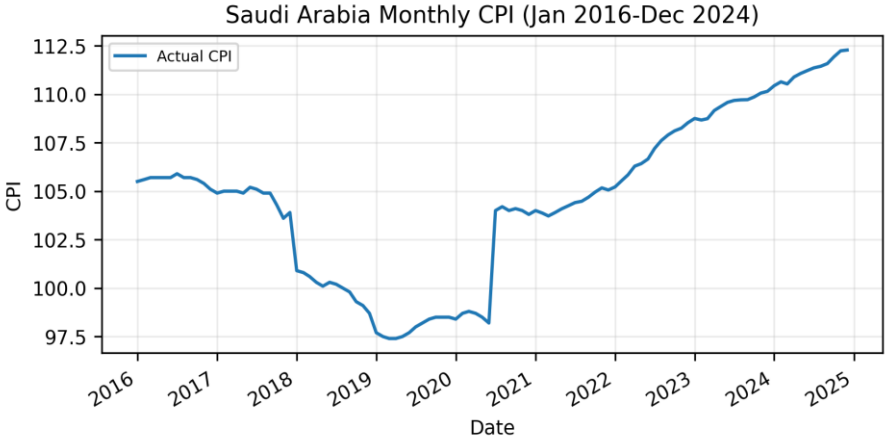


Figure 2. Saudi Arabia's monthly CPI (January 2016–December 2024).

4.1. CPI time series data

The monthly CPI data cover the study period from January 2016 to December 2024. The series has a mean of 104.63 index points and a standard deviation of approximately 4.26 points. The lowest recorded value was 97.4 points in March 2019, while the highest recorded value was 112.28 points in December 2024. The index stood at 100.9 points in January 2018 and then declined further toward the March 2019 trough. This pattern reflects a period of price deflation followed by a gradual and sustained rise in prices after 2019, with the index surpassing its prior peak and reaching new highs by the end of 2024. Consequently, Figure 2 show that the CPI series experienced significant fluctuations over the study years, indicating its sensitivity to economic cycles and various fiscal measures during that period.

The distribution of the index values is slightly left-skewed (skewness = –0.12), indicating a relative dominance of lower values in the series. Its kurtosis (2.06) is lower than that of a normal distribution (≈ 3), indicating a moderately flat peak and lighter tails, which implies a moderate spread of values around the mean without extreme concentration in any single value.

Besides, Table 6 shows that the 2024 test period has a higher mean CPI level and a narrower standard deviation than the training period. This indicates that the test set represents a continuation of the post-2020 upward CPI regime rather than a return to the lower CPI levels observed during 2018–2019. Therefore, evaluating the models on 2024 data provides useful evidence on their ability to forecast during a sustained upward inflationary phase.

Table 6. Descriptive statistics for the training and test CPI series.

Period	Mean	Std. deviation	Minimum	Maximum
Training data, 2016–2023	103.80	3.75	97.40	110.15
Test data, 2024	111.30	0.63	110.44	112.28

The 2024 holdout is also substantially less volatile than the training period: Its standard deviation is 0.63 compared with 3.75 for 2016–2023. This comparatively smooth test window can make short-

horizon point forecasting easier and may contribute to the low RMSE and MAPE values reported for the best models. For that reason, the error metrics are interpreted together with Theil's U against a naive previous-month benchmark. In the final evaluation, SARIMA (U = 0.7613), SVR (U = 0.9256), SARIMA–SVR (U = 0.6528), and GM(1,1)–SVR (U = 0.6427) all achieve U < 1; only the standalone GM(1,1) remains above the naive benchmark.

4.2. Benchmark model performance

4.2.1. Individual models

In this subsection, we outline the core structure and fitted parameters of each standalone forecasting model, namely SARIMA, GM(1,1), and SVR, and highlight the most significant insights obtained from their respective estimations. By examining these values, we reveal each model's capacity to capture the underlying trend, seasonality, and nonlinear dynamics of the CPI series before proceeding to a detailed error-based evaluation.

4.2.1.1. Analysis of the SARIMA model

In this subsection, we extend the ARIMA framework by incorporating seasonal dynamics to capture both intrayear cycles and long-run trends in the CPI series.

Stationarity testing

Stationarity is assessed using the training period only (January 2016–December 2023), preserving the unseen 2024 holdout. At the level, the ADF statistic is −0.1991 ($p = 0.939$), which is above the 5% critical value of −2.8925, so nonstationarity is not rejected. After first differencing, the ADF statistic is −9.3031 ($p < 0.001$), which is below the 5% critical value of −2.8928. Therefore, $d = 1$ is retained in the final SARIMA specification.

Model identification and selection

The differenced training series shows seasonal dependence around lag 12, but a seasonal ACF spike alone does not establish a seasonal unit root. We therefore compare $D = 1$ and $D = 0$ for SARIMA(1,1,1)(0,D,1)[12] using only pre-2024 data. Table 7 reports the information criteria, the seasonal MA estimate, and chronological one-step-ahead validation over 2021–2023. Although $D = 1$ gives a lower AIC/BIC, it produces a near-boundary seasonal MA estimate ($\Theta_{12} = -0.9996$) and substantially worse rolling RMSE. $D = 0$ removes the boundary warning and reduces the rolling RMSE from 0.8164 to 0.3297; it is therefore selected for the final forecasting analysis.

Table 7. Training-only seasonal differencing specification check.

Specification	Training fit	Chronological validation / decision
SARIMA(1,1,1)(0,1,1)[12]	AIC = 204.893; BIC = 214.568; $\Theta_{12} = -0.9996$	2021–2023 rolling RMSE = 0.8164; near-boundary seasonal MA warning
SARIMA(1,1,1)(0,0,1)[12]	AIC = 210.536; BIC = 220.752; $\Theta_{12} = 0.0992$	2021–2023 rolling RMSE = 0.3297; selected
Final choice	SARIMA(1,1,1)(0,0,1)[12], no drift	Boundary warning is removed and chronological forecast validation is materially improved

Parameter estimates

For the final SARIMA(1,1,1)(0,0,1)[12] specification, the order is fixed from training-only selection, and no unconditional drift is included. The December 2023 reference fit gives AR(1) = 0.9268 (standard error (SE) = 0.5643), MA(1) = −0.8842 (SE = 0.7056), and seasonal MA(12) = 0.0992 (SE = 0.1565). These are reference estimates only; the coefficients are re-estimated at every rolling forecast origin during evaluation.

Diagnostic checking

We assess the final $D = 0$ SARIMA residuals descriptively for remaining serial structure.

- The ACF/PACF of the final SARIMA training fit residuals are shown in Figure 3.
- Ljung–Box testing at Lag 12 does not indicate remaining serial autocorrelation after initialization ($p > 0.99$).
- Residual normality is not assumed for the forecast comparison; the training residual distribution remains affected by large structural movements in the sample.

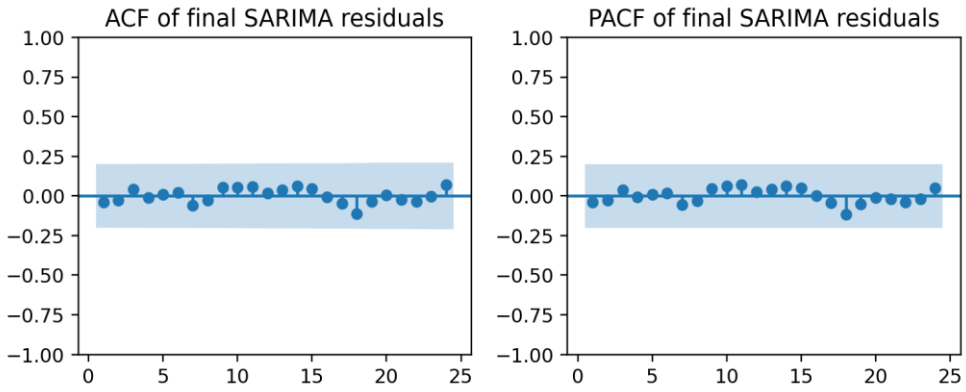


Figure 3. ACF and PACF of residuals from the final SARIMA(1,1,1)(0,0,1)[12] training fit.

These diagnostics indicate that the final $D = 0$ specification captures the principal serial

dependence without the near-boundary seasonal MA behavior produced by $D = 1$. Residual distributional departures are interpreted cautiously and are not used as evidence of the forecasts' superiority.

To examine the seasonal patterns, the time series was decomposed into trend, seasonal, and residual components using seasonal trend loess (STL) decomposition. The result, shown in Figure 4, confirms the presence of a stable seasonal pattern with a 12-month cycle, supporting the use of seasonal models such as SARIMA.

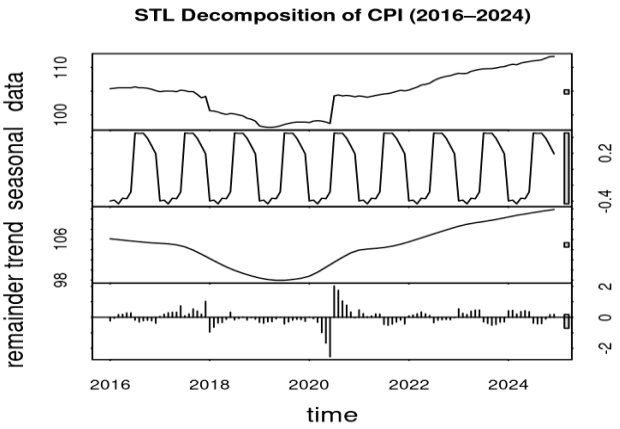


Figure 4. Seasonal trend decomposition of the CPI time series.

Nevertheless, Figure 5 provides the training-only chronological specification check required for seasonal differencing. Over January 2021–December 2023, the $D = 0$ specification tracks the one-step-ahead CPI path more closely than the $D = 1$ alternative. The corresponding RMSE values are 0.3297 for $D = 0$ and 0.8164 for $D = 1$.

The $D = 1$ model has a lower in-sample AIC/BIC but its seasonal MA estimate is essentially −1, whereas $D = 0$ removes the boundary warning and performs substantially better under the study's chronological forecast validation criterion. The final 2024 evaluation therefore uses $D = 0$; the holdout is not used to make this choice.

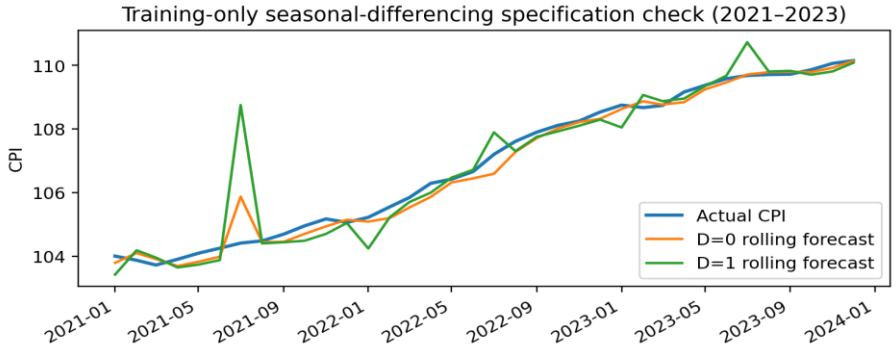


Figure 5. Training-only rolling comparison of the $D = 0$ and $D = 1$ SARIMA specifications (2021–2023).

4.2.1.2. Analysis of the GM(1,1) model

For reproducibility, the GM(1,1) parameters estimated on the full January 2016–December 2023 training sample immediately before the holdout were $a = -0.000576$ and $b = 100.903$. These are reference estimates only; both parameters are re-estimated at every rolling forecast origin. The coefficient b is the grey input term in the whitening equation and should not be interpreted as the long-run mean of CPI.

In GM(1,1), the development coefficient a controls the exponential evolution of the accumulated series, while b is the grey input term. Their interpretation is model-specific; b's numerical proximity to the CPI level is not evidence that it anchors a long-run mean. The empirical issue relevant to hybridization is instead that the rolling GM(1,1) forecasts leave large, strongly predictable residual errors, which the residual correction model can exploit.

4.2.1.3. SVR model analysis

The revised standalone ε-SVR uses an RBF kernel with CPI lags of 1 and 12. Time-series cross-validation selected the following parameter configuration:

$$C = 100, \quad \gamma = 0.01, \quad \varepsilon = 0.05, \quad \mathcal{L} = \{1, 12\}.$$

In this fitted specification, the relatively large C penalizes larger forecast deviations strongly, $\gamma = 0.01$ maintains a comparatively smooth nonlinear response, and $\varepsilon = 0.05$ ignores very small deviations within 0.05 CPI index points. These values are empirically selected tuning parameters rather than theoretically universal constants.

Under the revised rolling one-step-ahead evaluation, standalone SVR achieved a 2024 RMSE of 0.2046 and a MAPE of 0.1501%, outperforming the naive benchmark with Theil's U = 0.9256.

After constructing CPI lags (1,12), the standalone SVR was fitted on 84 effective supervised training cases rather than all 96 monthly training observations; 67 of these 84 cases were support vectors (79.8%). The support vector proportion is reported descriptively and was not used as a model selection criterion.

4.2.2. Hybrid model parameterization and interpretation

In this subsection, we examine the parameterization of GM(1,1)–SVR and SARIMA–SVR to test whether residual correction improves CPI forecasting relative to each base model. The hybrid construction is not assumed to improve accuracy *a priori*; its value depends on whether the base model's residuals contain a predictable structure.

Analysis of the hybrid GM(1,1)–SVR model

In this subsection, we present the key parameter estimates for the GM(1,1)–SVR hybrid and explain how the base model's parameters and SVR settings determine the residual correction procedure. The SVR stage is interpreted as modeling potentially predictable residual structures; its gain is not attributed to a nonlinear structure unless supported by evidence. Table 8 lists these parameters and their corresponding interpretations.

Table 8. Hybrid GM(1,1)–SVR parameters.

Parameter	Estimate	Interpretation
GM(1,1) a (development)	-0.000576*	December 2023 reference development coefficient; re-estimated at each forecast origin.
GM(1,1) b (grey input)	100.903*	December 2023 reference grey input term; not a long-run mean; re-estimated at each origin.
SVR C (penalty)	100	Selected by time-series cross-validation.
SVR γ (RBF width)	0.01	RBF kernel parameter selected within the training period.
SVR ε (insensitivity)	0.05	ε-insensitive margin selected within the training period.
Residual lags	(1, 6)	Lagged rolling-origin GM(1,1) residual predictors.

*The GM(1,1) base model coefficients are reference estimates from the December 2023 training fit; a and b are re-estimated at each rolling forecast origin.

- Base forecast (GM(1,1)): The December 2023 reference estimates are $a = -0.000576$ and $b = 100.903$. These values are re-estimated at every forecast origin. The empirical limitation of the grey model base is not inferred from the coefficient b as a long-run mean; rather, its rolling forecasts leave large systematic forecast residuals that can be modeled by the correction stage.
- Residual correction (SVR): The revised GM(1,1)–SVR uses residual lags (1,6), with $C = 100$, $\gamma = 0.01$, and $\varepsilon = 0.05$ selected by time-series cross-validation. The SVR is trained on genuine rolling-origin one-step-ahead GM(1,1) residuals.
- The rolling-origin procedure generated 60 pre-2024 GM(1,1) residuals after the 36-month initialization window. Applying residual lags (1,6) leaves 54 effective SVR training cases; 39 of these 54 cases are support vectors (72.2%). This corrected denominator replaces the earlier percentage based incorrectly on the full 96-month training period.
- Combined effect: The residual correction component substantially reduces the systematic forecast error of GM(1,1). In the 2024 holdout, GM(1,1)–SVR achieves the lowest RMSE (0.1421), while SARIMA–SVR has a slightly lower MAPE (0.1044% versus 0.1062%). The large GM(1,1)–SVR improvement should not be interpreted as direct evidence of nonlinear complementarity. It first demonstrates that the weak GM(1,1) base leaves a large amount of predictable systematic error that can be reconstructed from lagged rolling-origin residual information; the present evidence does not identify that predictable component as intrinsically nonlinear.

Analysis of the hybrid SARIMA–SVR model

This subsection tests whether adding SVR residual correction to the final SARIMA(1,1,1)(0,0,1)[12] base improves forecasts' accuracy. The SARIMA order is fixed before the 2024 holdout, while its coefficients are re-estimated at each rolling origin. The SVR lag structure and hyperparameters are fixed from pre-2024 time-series cross-validation, and the SVR is refitted monthly on the expanding residual history. A summary of these parameters and their interpretations is outlined in Table 9.

Table 9. Hybrid SARIMA–SVR parameters.

Parameter	Estimate	Interpretation
SARIMA (p, d, q)	(1,1,1)	Nonseasonal ARMA terms with first differencing.
SARIMA (P, D, Q, s)	(0,0,1,12)	Seasonal MA term at lag 12; no seasonal differencing.
phi1 (AR1)	0.9268*	December 2023 reference AR(1) estimate; re-estimated at each forecast origin.
theta1 (MA1)	−0.8842*	December 2023 reference MA(1) estimate; re-estimated at each forecast origin.
Theta12 (SMA)	0.0992*	December 2023 reference seasonal MA estimate; boundary warning removed.
Drift	None	No unconditional drift in the final specification.
SVR C (penalty)	10	Selected by time-series cross-validation.
SVR γ (RBF width)	0.001	Selected by time-series cross-validation.
SVR ε	0.01	Selected within the training period.
Residual lags	(1, 2, 12)	Lagged genuine rolling-origin SARIMA forecast residual predictors.

*SARIMA coefficient values are reference estimates from the December 2023 $D = 0$ training fit; the order is fixed pretest, but the coefficients are re-estimated at each rolling forecast origin.

- Base forecast (SARIMA): The final SARIMA(1,1,1)(0,0,1)[12] uses first nonseasonal differencing and a seasonal MA term without seasonal differencing, avoiding the near-boundary seasonal MA estimate produced by $D = 1$.
- Residual correction (SVR): For the final $D = 0$ SARIMA residual series, time-series cross-validation selected the residual lags (1,2,12), $C = 10$, $\gamma = 0.001$, and $\varepsilon = 0.01$. These settings are fixed pre-2024, while the SVR itself is refitted monthly on the expanding set of genuine rolling-origin residuals.
- The rolling-origin procedure generated 60 pre-2024 SARIMA forecast residuals after the 36-month initialization window. Applying residual lags (1,2,12) leaves 48 effective SVR training cases; all 48 were support vectors (100%). The support vector fraction is descriptive and is not used as a model selection criterion.
- Combined effect: Under the final $D = 0$ specification, SARIMA–SVR improves numerical the RMSE of 2024 from 0.1683 for SARIMA to 0.1443, and the MAPE from 0.1298% to 0.1044%. However, the Harvey–Leybourne–Newbold-adjusted DM comparison does not establish a statistically significant difference (adjusted $p = 0.400$). The correction gain is therefore described as a modest numerical improvement rather than statistical superiority.

The two base models show different degrees of residual predictability rather than a simple success/failure dichotomy. GM(1,1) produces strong conventional dependence and a very large reduction in the chronological residual prediction's RMSE. The final $D = 0$ SARIMA residuals show weak ACF/Ljung–Box/BDS evidence, yet the chronological residual learner still reduces the residual RMSE and produces modest forecast gains in several annual windows. This reinforces the screening

principle that direct chronological residual prediction performance should be considered alongside conventional diagnostics. Table 10 presents the pre-2024 base residual predictability diagnostics.

Table 10. Pre-2024 base residual predictability diagnostics (January 2021–December 2023).

Base model	Lag-1 ACF	Ljung–Box p (12)	BDS p (dim. 2)	Chronological residual RMSE: Zero → SVR	Implication
SARIMA	0.080	0.216	0.935	0.3297 → 0.1971	Weak conventional dependence, but chronological residual prediction improves RMSE.
GM(1,1)	0.918	<0.001	<0.001	4.6900 → 0.1698	Strong predictable residual structure; correction produces a large gain.

The diagnostics illustrate, but do not mathematically prove, the residual predictability principle in Section 3.3. For the final SARIMA, ACF, Ljung–Box, and BDS do not show strong conventional dependence, yet chronological residual prediction reduces the RMSE from 0.3297 to 0.1971. For GM(1,1), both the conventional dependence diagnostics and chronological residual prediction indicate a strong predictable structure. The results show why nonsignificant residual tests should not be treated as proof of complete unpredictability; the direct rolling residual prediction benchmark provides complementary evidence.

Parameter timing note: The asterisked GM(1,1) and SARIMA coefficient values in Tables 8 and 9 are reference estimates from the December 2023 training fit. They are not held fixed during 2024. At each monthly forecast origin, the GM(1,1) and SARIMA coefficients are re-estimated on the expanding history and each SVR is refitted on the expanding supervised sample, while the model orders, lag sets, and SVR hyperparameters remain fixed from the pre-2024 selection.

4.3. Visual comparison

Figures 6 and 7 compare the rolling one-step-ahead forecasts with actual CPI data only for the January–December 2024 holdout. They are used to assess month-by-month tracking within the final test window; they do not display the 2019 trough or the 2020 level shift. Historical regime behavior is evaluated separately through the cross-regime rolling backtests reported in Section 4.5.

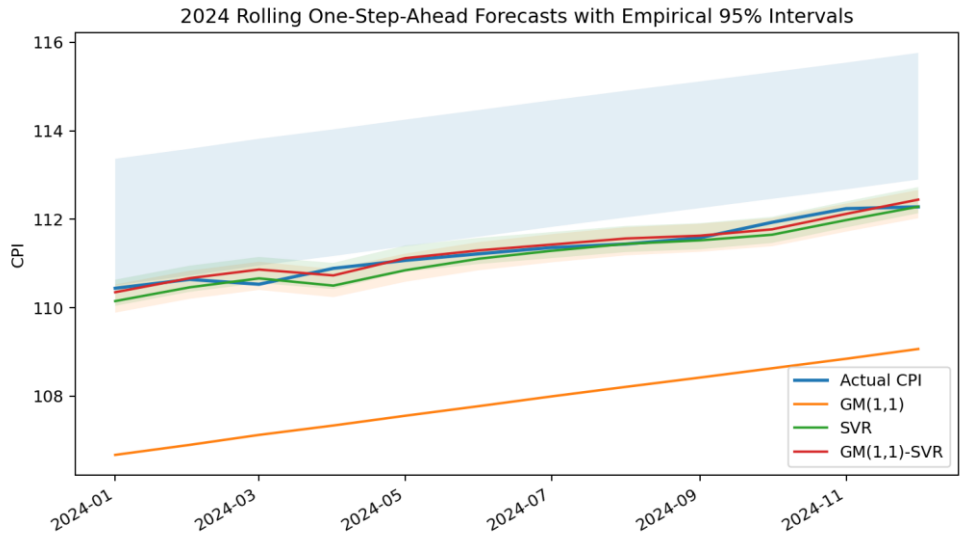


Figure 6. Actual CPI and revised GM(1,1), SVR, and GM(1,1)–SVR rolling one-step-ahead forecasts for 2024 with empirical 95% forecast intervals.

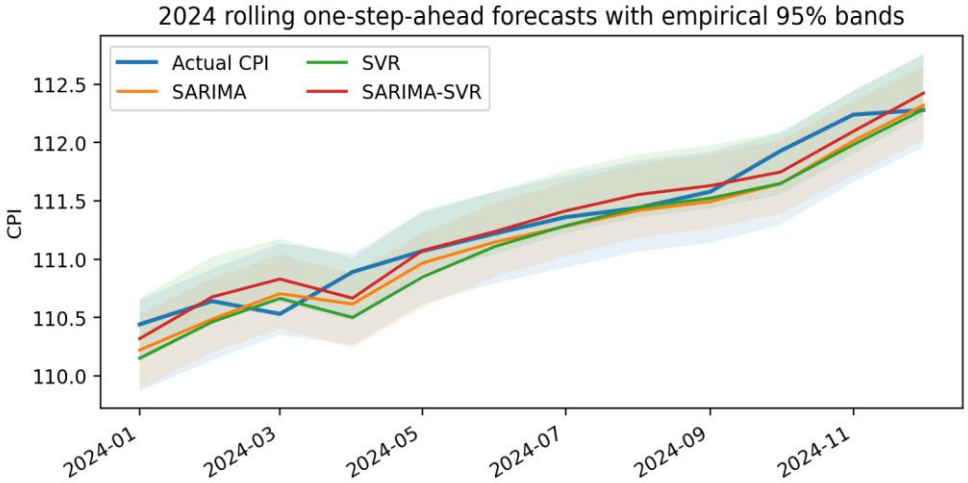


Figure 7. Actual CPI and final $D = 0$ SARIMA, SVR, and SARIMA–SVR rolling one-step-ahead forecasts for 2024 with empirical 95% forecast bands.

The revised 2024 prediction files were also checked for missing values and consistent month-to-month date alignment. All reported forecasts and observed values are on the original CPI's scale, so no inverse transformation is required before error calculation. RMSE and MAPE are computed directly from the 12 observed and predicted CPI values for January–December 2024.

Moreover, Figures 6 and 7 present rolling one-step-ahead forecasts for the unseen 2024 holdout. Figure 7 uses the final training-selected $D = 0$ SARIMA specification. Under this specification, both SARIMA and SARIMA–SVR outperform the naive benchmark, and SARIMA–SVR shows a small numerical improvement over SARIMA, although the adjusted DM comparison is not statistically significant.

4.4. Diagnostic analysis

This section presents descriptive residual diagnostics for the 2024 holdout. Because only 12 monthly errors are available, ACF, histogram, and quantile–quantile (QQ) displays are used as limited diagnostic illustrations rather than as formal evidence that the residuals are white noise or normally distributed.

First, we compare the actual CPI values with each hybrid forecast to describe the month-by-month deviations. We then examine the residuals over time and display the ACF, histogram, and QQ plot as descriptive diagnostics. With only 12 holdout errors, these displays are not used to prove residual independence or normality.

The following diagnostic plots use the observed CPI values and hybrid forecasts for January–December 2024.

Figure 8 compares the observed CPI values with the two revised hybrid forecasts, and Figure 9 displays the corresponding residual sequences. Figures 10–12 focus on GM(1,1)–SVR because it has the lowest numerical RMSE for 2024. With $T = 12$, the ACF, histogram, and QQ plot cannot support strong distributional conclusions; at most, they show no obvious persistent directional pattern within this short holdout.

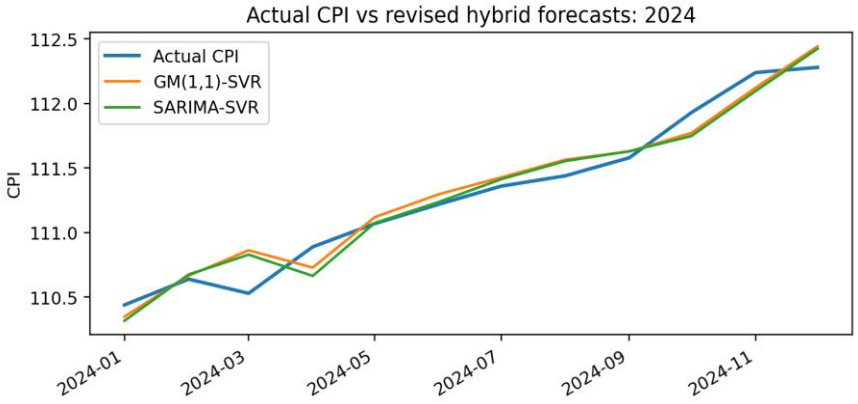


Figure 8. Actual CPI and revised hybrid forecasts (January–December 2024).

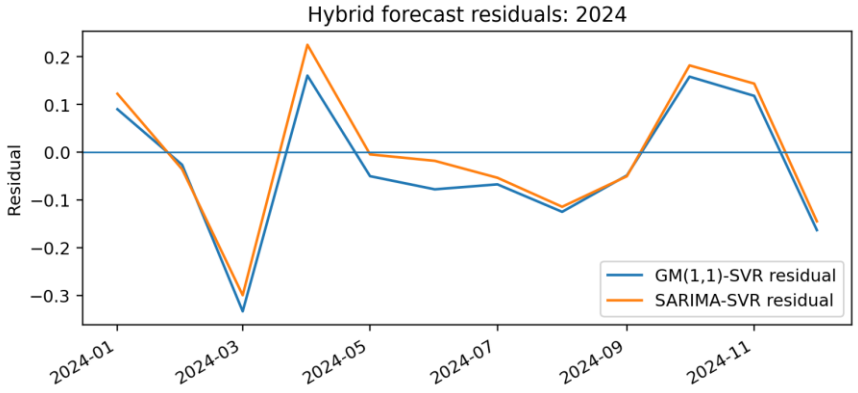


Figure 9. Revised hybrid model residuals over time (January–December 2024).

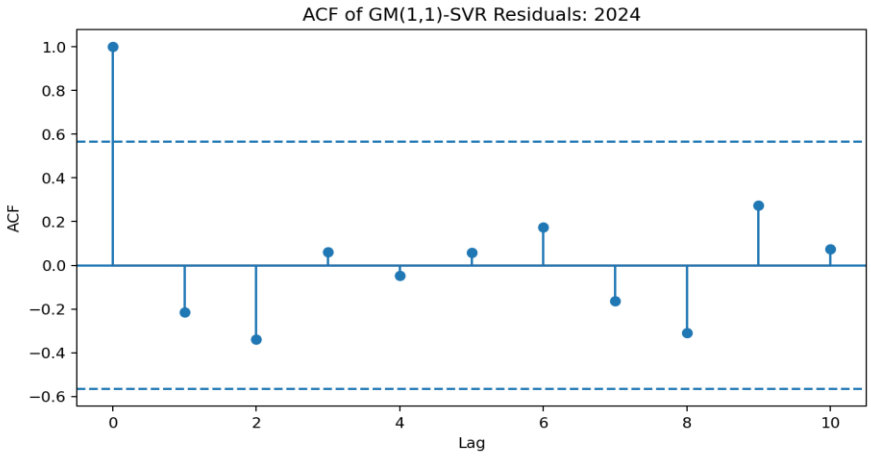


Figure 10. ACF of GM(1,1)–SVR residuals.

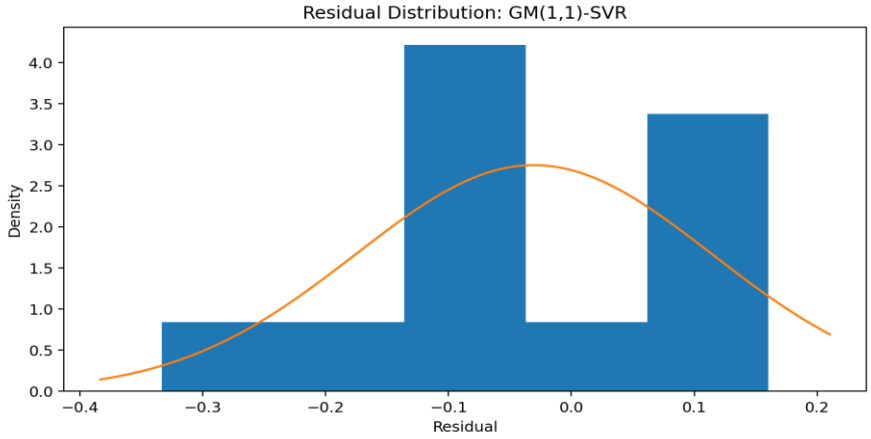


Figure 11. Residual distribution for GM(1,1)–SVR.

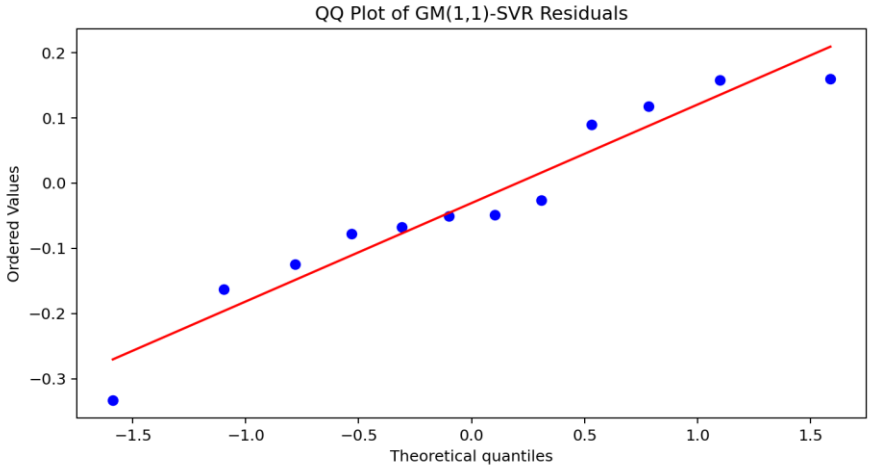


Figure 12. QQ plot of GM(1,1)–SVR residuals.

Taken together, the plots are consistent with the error metric results but are too short to establish residual independence or normality. Formal claims therefore rely primarily on the chronological forecast comparison and are stated cautiously.

4.5. Forecasting accuracy

Forecast accuracy is evaluated on the unseen January–December 2024 holdout period using RMSE and MAPE. In the revised analysis, the training period's fitted-error metrics are no longer reported in Table 11 because they are not directly comparable with genuine out-of-sample forecast errors. This change also removes the previously reported SARIMA training RMSE of 12.55, which was inconsistent with the CPI scale and with the holdout period's errors. The revised table reports only rolling one-step-ahead holdout performance after checking the evaluation window, date alignment, and CPI scale.

Table 11. Forecast accuracy: MAPE and RMSE.

Model	Period	MAPE	RMSE
SARIMA	2024 holdout	0.1298%	0.1683
SVR	2024 holdout	0.1501%	0.2046
SARIMA–SVR	2024 holdout	0.1044%	0.1443
GM(1,1)	2024 holdout	3.0760%	3.4278
GM(1,1)–SVR	2024 holdout	0.1062%	0.1421

Table 11 shows that GM(1,1)–SVR has the lowest RMSE for 2024 (0.1421), while SARIMA–SVR has the lowest MAPE (0.1044%). Their RMSE values are very close (0.1421 versus 0.1443), and their adjusted DM comparison is not significant. Standalone SARIMA improves substantially under the final $D = 0$ specification (RMSE = 0.1683; MAPE = 0.1298%) and has a lower numerical RMSE and MAPE than the standalone SVR (RMSE = 0.2046; MAPE = 0.1501%). GM(1,1) remains the weakest base model. All values use the same January–December 2024 rolling one-step-ahead holdout window.

Figure 13 summarizes the final 2024 error measures. GM(1,1)–SVR has the lowest RMSE, while SARIMA–SVR has the lowest MAPE. The small difference between the two hybrids should not be interpreted as established superiority because the corresponding adjusted DM comparison is not significant.

Residual correction remains most dramatic and stable for GM(1,1), whose base forecasts leave large predictable errors. For the final SARIMA specification, residual correction yields a smaller numerical improvement in 2024 and in several annual rolling windows, but the 2024 improvement is not statistically significant. Thus, the hybrid's value remains conditional on the base model and the reproducible predictability of its residual process.

Because the test window contains only 12 monthly observations, the numerical ranking should not be interpreted as proof of general superiority. Statistical forecast comparison tests are therefore considered together with the RMSE, MAPE, and Theil's U. Table 12 reports forecast accuracy relative to the naive benchmark.

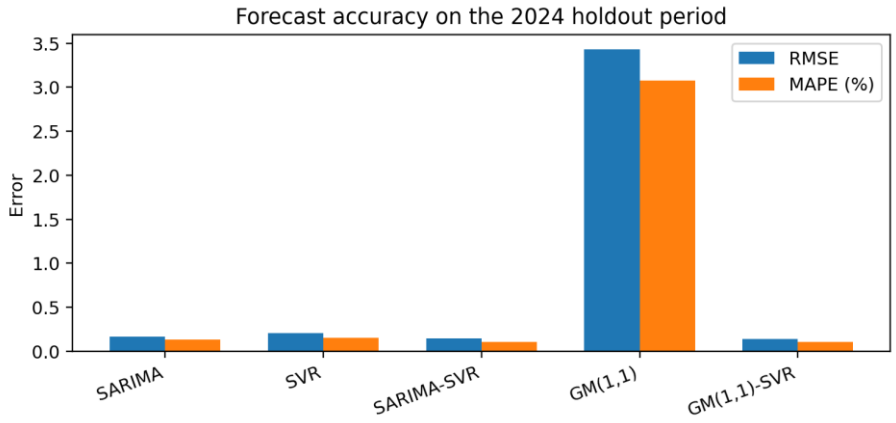


Figure 13. RMSE and MAPE for rolling one-step-ahead forecasts on the 2024 holdout period.

Table 12. Relative accuracy against a naive benchmark.

Model	MAPE (%)	RMSE	Theil's U	Note
SARIMA	0.1298	0.1683	0.7613	Better than naive
SVR	0.1501	0.2046	0.9256	Better than naive
SARIMA–SVR	0.1044	0.1443	0.6528	Better than naive
GM(1,1)	3.0760	3.4278	15.5051	Above naive
GM(1,1)–SVR	0.1062	0.1421	0.6427	Better than naive

Theil's U compares each model with a naive previous-month forecast. GM(1,1)–SVR has the lowest U (0.6427), followed closely by SARIMA–SVR (0.6528), while the standalone SARIMA (0.7613) and standalone SVR (0.9256) also outperform the naive benchmark. GM(1,1) remains well above 1 (15.5051). These relative results reinforce that the final $D = 0$ SARIMA is materially stronger than the earlier $D = 1$ specification.

Differences in forecast accuracy were assessed using squared-error DM tests over the 12-month 2024 holdout. Table 13 presents the Harvey–Leybourne–Newbold small-sample-adjusted statistic and adjusted p-value. A nonsignificant result is interpreted only as insufficient evidence of a difference, not as evidence of equal accuracy. Conversely, the very large GM(1,1) versus GM(1,1)–SVR statistic is not treated as an effect-size measure; it is driven by the much larger squared errors of GM(1,1) in a very small sample. All inference is therefore treated as supporting rather than definitive evidence.

Table 13. Diebold–Mariano test results.

Comparison	Harvey–Leybourne–Newbold-adjusted DM	Adjusted p-value
GM(1,1) vs GM(1,1)–SVR	28.740	<0.001
SVR vs GM(1,1)–SVR	1.278	0.227
SARIMA–SVR vs GM(1,1)–SVR	0.195	0.849
SARIMA vs SARIMA–SVR	0.875	0.400
SVR vs SARIMA–SVR	1.512	0.159

After the Harvey–Leybourne–Newbold small-sample correction, GM(1,1)–SVR remains significantly more accurate than GM(1,1) (adjusted $p < 0.001$). Its difference from standalone SVR is not statistically significant (adjusted $p = 0.227$), and its difference from SARIMA–SVR is also not significant (adjusted $p = 0.849$). SARIMA–SVR is numerically more accurate than standalone SARIMA in 2024, but the adjusted comparison is not significant ($p = 0.400$); its difference from standalone SVR is likewise not significant ($p = 0.159$). With only 12 holdout observations, these tests are interpreted as supporting evidence rather than proof of the models' equivalence or superiority.

4.5.1. Cross-regime and external robustness checks

To examine whether the residual correction pattern is tied only to the stable 2024 holdout, the final $D = 0$ model structures were evaluated in annual rolling windows for 2021–2024. Table 14 shows that GM(1,1)–SVR consistently removes the large GM(1,1) base error. SARIMA–SVR also lowers SARIMA's RMSE in 2021, 2022, and 2024, but not in 2023, and its gains are much smaller. This mixed pattern supports a conditional residual predictability interpretation rather than the hybrid's universal advantage.

Table 14. Cross-regime rolling one-step-ahead RMSE on the Saudi CPI.

Year	SARIMA	SVR	SARIMA–SVR	GM(1,1)	GM(1,1)–SVR
2021	0.4687	0.2142	0.2105	4.8969	0.1766
2022	0.2964	0.1844	0.2297	4.8266	0.1738
2023	0.1363	0.1755	0.1397	4.3259	0.1586
2024	0.1683	0.2046	0.1443	3.4278	0.1421

As a limited external series transferability check, the final Saudi-selected model structures and SVR hyperparameters were transferred without retuning to the US's seasonally adjusted CPI-U series (CPIAUCSL) from the US Bureau of Labor Statistics via Federal Reserve Economic Data (FRED). The external training window is January 2016–December 2023 and the available holdout is January–September 2024. The check is supplementary and is not interpreted as evidence of universal cross-country validity. Table 15 displays the accuracy measures for this external US CPI check.

Table 15. External US CPI robustness check, January–September 2024.

Model	RMSE	MAPE (%)	Theil's U
SARIMA	0.5158	0.1296	0.6301
SVR	0.6473	0.1785	0.7908
SARIMA–SVR	0.5621	0.1213	0.6866
GM(1,1)	4.7698	1.4327	5.8266
GM(1,1)–SVR	0.6946	0.1659	0.8485

The external check has the same cautious lesson: Correcting GM(1,1) removes a very large base model error, but GM(1,1)–SVR is not the best external model; standalone SARIMA has the lowest external RMSE. SARIMA–SVR does not improve SARIMA's RMSE, although its MAPE is slightly

lower. Because the external holdout contains only 9 months and uses a differently defined national CPI series, the exercise is an illustration of transferability only. It supports screening residual correction empirically rather than assuming that a hybrid architecture will dominate across settings.

4.6. Economic and policy interpretation

The revised results support only a limited monitoring interpretation. GM(1,1)–SVR produced the lowest numerical forecast errors over the 2024 holdout period, but the holdout contains only 12 monthly observations and its advantage over standalone SVR is not statistically significant after small-sample correction. Accordingly, the models may be used as short-horizon statistical monitoring tools, but the present evidence is insufficient to justify direct claims about budget indexation, policy optimization, or formal decision support.

The unusually low volatility of the 2024 holdout (standard deviation = 0.63 versus 3.75 in the training period) is an additional reason for caution: Part of the very small point-forecast errors may reflect an easier forecasting window rather than the model's architecture alone. Theil's U is therefore used to judge whether the models add predictive skill beyond simple persistence.

The model comparison clarifies that residual correction can add forecasting value without implying the hybrid's universal superiority. GM(1,1)–SVR substantially improves on a weak GM(1,1) base, while the final SARIMA and SARIMA–SVR are both strong short-horizon forecasters relative to the naive benchmark. However, these models do not include policy or macroeconomic covariates. Future structural specifications should consider tax policy interventions such as the July 2020 increase in value-added tax, administered energy price measures, exchange rate conditions, housing costs, interest rates, and other relevant price drivers. Forecast accuracy in the present study should therefore not be interpreted as causal policy evidence.

4.7. Structural break assessment

Because the CPI series displays a pronounced level change in mid-2020, the revised analysis formally tests a prespecified break at July 2020. To reduce the risk of spurious inference from the nonstationary CPI level, the break test is applied to monthly CPI changes rather than to the undifferenced index.

A Chow-type parameter-stability test was applied to a linear regression of monthly CPI changes on an intercept and time trend, allowing the coefficients to differ before and from July 2020 onward. The test rejects parameter stability at the 5% level ($F = 3.891$, $p = 0.023$). The average monthly change in CPI was –0.138 index points before July 2020 and 0.261 afterward; a Welch comparison of these mean monthly changes also indicates a difference ($p = 0.002$).

These results support treating the 2020 episode as a meaningful instability in the CPI-generating process. The timing coincides with the increase in Saudi Arabia's standard VAT rate from 5% to 15% on 1 July 2020, which provides an economically plausible policy context for the observed level shift. However, the present univariate design does not identify the VAT change as the sole causal source of the break; contemporaneous pandemic-related and other economic factors may also have contributed. The break date is therefore treated as a prespecified empirical instability point rather than as a causal attribution. Figure 14 illustrates the monthly CPI changes and marks the prespecified July 2020 structural break date.

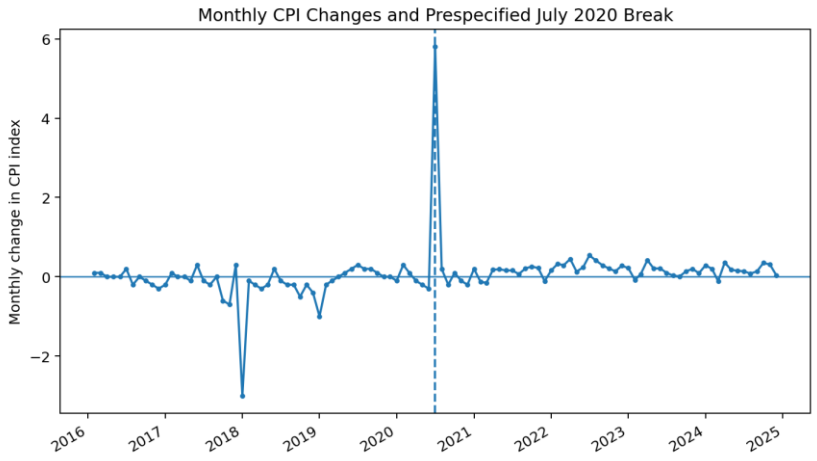


Figure 14. Monthly CPI changes and the prespecified July 2020 structural break date.

4.8. Limitations and threats to validity

First, the full sample contains 108 monthly observations, while the out-of-sample evaluation period contains only 12 monthly forecasts. This design preserves one complete annual cycle for evaluation but provides limited statistical power for the model comparison tests. To reduce overstatement, the revised analysis applies the Harvey–Leybourne–Newbold small-sample correction to the DM statistic, distinguishes numerical ranking from statistical significance, and avoids interpreting borderline *p*-values as strong evidence. The 12-month holdout should therefore be viewed as an initial short-horizon validation window. Stronger conclusions require additional future observations, longer rolling holdouts, and repeated evaluation of the forecast origin.

Second, the models are intentionally univariate and use only lagged CPI information. This design isolates the predictive contribution of the time-series structure, but it does not constitute a macroeconometric model and cannot identify the causal effects of external drivers. Important Saudi-specific factors omitted from the present specification include the increase in the standard VAT rate from 5% to 15%, effective 1 July 2020; administered energy price policies such as the domestic gasoline price cap introduced in July 2021; and the Saudi riyal's fixed exchange rate arrangement against the US dollar. These factors can affect the level or transmission of consumer prices and should be incorporated explicitly in future multivariate or intervention-based models. Accordingly, the present forecasts should be interpreted as conditional time-series projections rather than as structural explanations of inflation [29,33].

Third, the revised analysis identifies evidence of parameter instability around the prespecified July 2020 break in monthly changes in CPI. This formal test strengthens the treatment of the 2020 episode but it does not establish that July 2020 is the only break date. Future research should examine multiple-break procedures and evaluate forecasts' stability across alternative regimes.

Fourth, the forecast results depend on the selected lag structures, scaling procedure, rolling-origin initialization window, and SVR hyperparameters. The revised analysis reports these choices explicitly and provides reproducible prediction outputs and code. Nevertheless, additional sensitivity analysis with longer holdout periods would be valuable.

Future research may also compare the additive residual correction structure used in this study

with multiplicative hybrid models, especially if the changes in CPI exhibit proportional rather than additive nonlinear behavior.

5. Conclusions

This study evaluates standalone and residual correction hybrid models for short-term monthly CPI forecasting in Saudi Arabia using a strictly chronological design. A training-only seasonal-specification check selects SARIMA(1,1,1)(0,0,1)[12]; the earlier $D = 1$ alternative produces a seasonal MA estimate near −1 and much worse rolling validation, whereas $D = 0$ removes the boundary warning and improves the pre-2024 forecasting performance. In the 2024 holdout, GM(1,1)–SVR has the lowest RMSE (0.1421) and SARIMA–SVR has the lowest MAPE (0.1044%), but their adjusted DM difference is not statistically significant. GM(1,1)–SVR produces a large and stable correction gain over the weak GM(1,1) base. SARIMA–SVR produces a smaller numerical improvement over the final SARIMA in 2024, but the adjusted comparison provides insufficient evidence of statistical superiority, and the cross-regime analysis shows that SARIMA–SVR does not improve over SARIMA in every year. The external US transferability check likewise shows that standalone SARIMA has a lower RMSE than SARIMA–SVR. The main transferable finding is therefore conditional rather than architectural: Residual correction is potentially useful when the base model's error process contains sufficient reproducible out-of-sample predictability. Conventional residual diagnostics and direct chronological residual-prediction performance should be considered jointly, because nonsignificant diagnostic tests do not prove complete unpredictability. The SVR correction is an estimator of potentially predictable residual structure, not proof of nonlinear complementarity. The July 2020 break documents historical instability within training but does not demonstrate *ex ante* shock anticipation. With only 12 Saudi holdout observations, the conclusions remain short-horizon empirical evidence rather than a general theorem of hybrid models' superiority.

The findings provide short-horizon evidence relevant to monitoring monthly inflation in Saudi Arabia, but they should not be interpreted as direct policy guidance. Because the evaluation period contains only 12 observations and the models are univariate, broader fiscal or monetary applications would require validation using longer holdout periods and multivariate specifications with relevant economic and policy drivers. A natural extension is to compare the present univariate benchmarks with seasonal autoregressive integrated moving average exogenous (SARIMAX) dynamic-regression, or machine learning models that include exogenous policy and macroeconomic variables, with intervention indicators for major policy changes such as the July 2020 VAT reform.

Use of generative AI tools declaration

The author declares that no generative AI tools were used in the creation of this article.

Availability of data and material

The Saudi monthly CPI data used in this study were obtained from GASTAT, Saudi Arabia. For reproducibility, the revised submission includes the processed Saudi CPI series, chronological training–testing split, monthly model predictions, selected lag and hyperparameter settings, and analysis code. The external portability check uses the CPIAUCSL series as distributed by FRED. The

external check is supplementary to the Saudi analysis and does not alter the primary 2024 holdout conclusions.

Acknowledgments

The authors sincerely thank the University of Tabuk in Saudi Arabia. However, this research received no external funding.

Conflict of interest

The author declares that there are no competing interests.

References

1. D. S. Metwally, M. El-Morshedy, H. S. Shahen, A. S. M. A. Elshafei, The impact of financial stress on Saudi Arabia government spending: Using machine learning, *J. Radiat. Res. Appl. Sci.*, **18** (2025), 101582. https://doi.org/10.1016/j.jrras.2025.101582
2. *Saudi Arabia's annual inflation accelerates to 2.3% in June, Reuters*. (n.d.), 2025. Available from: https://www.reuters.com/world/middle-east/saudi-arabias-annual-inflation-accelerates
3. D. Havrlant, A. Darandary, *Economic diversification under Saudi Vision 2030*, 2021. https://doi.org/10.30573/KS--2021-DP06
4. R. Guido, S. Ferrisi, D. Lofaro, D. Conforti, An overview on the advancements of support vector machine models in healthcare applications: A review, *Information*, **15** (2024), 235. https://doi.org/10.3390/info15040235
5. G. E. P. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, *Time series analysis: Forecasting and control*, 5 Eds., Wiley, 2016.
6. J. L. Deng, Introduction to grey system theory, *J. Grey Syst.*, **1** (1989), 1–24. https://dl.acm.org/doi/10.5555/90757.90758
7. V. N. Vapnik, An overview of statistical learning theory, *IEEE Trans. Neural Netw.*, **10** (1999), 988–999. https://doi.org/10.1109/72.788640
8. D. N. Shah, M. Thaker, Forecasting stock index using a hybrid ARIMA-SVM model, *IOSR J. Econ. Fin.*, **15** (2024), 15–21. https://doi.org/10.9790/5933-1502011521
9. T. Nyoni, Predicting consumer price index in Saudi Arabia, 2019, 1–9.
10. A. Mahgoub, T. Alam, Modelling and predicting the consumer price index in Saudi Arabia, *J. Infras. Policy. Dev.*, **8** (2024), 5270. https://doi.org/10.24294/jipd.v8i9.5270
11. F. N Sibai, A. El-Moursy, A. Sibai, Forecasting the consumer price index: A comparative study of machine learning methods, *Int. J. Com. Dig. Sys.*, **15** (2024), 487–497. https://doi.org/10.12785/ijcds/150137
12. V. Ülke, A. Sahin, A. Subasi, A comparison of time series and machine learning models for inflation forecasting: empirical evidence from the USA, *Neural Comput. Appl.*, **30** (2018), 1519–1527. https://doi.org/10.1007/s00521-016-2766-x
13. M. J. Lawrence, R. H. Edmundson, M. J. O'Connor, The accuracy of combining judgemental and statistical forecasts, *Manage. Sci.*, **32** (1986), 1521–1532. https://dl.acm.org/doi/10.5555/2782812.2782813

14. A. H. Ashton, R. H. Ashton, Aggregating subjective forecasts: Some empirical results, *Manage. Sci.*, **31** (1985), 1499–1508. https://doi.org/10.1287/mnsc.31.12.1499
15. J. T. Luxhøj, J. O. Riis, B. Stensballe, A hybrid econometric—neural network modeling approach for sales forecasting, *Int. J. Prod. Econ.*, **43** (1996), 175–192. https://doi.org/10.1016/0925-5273(96)00039-4
16. J. Du Preez, S. F. Witt, Univariate versus multivariate time series forecasting: An application to international tourism demand, *Int. J. Forecasting*, **19** (2003), 435–451. https://doi.org/10.1016/S0169-2070(02)00057-2
17. G. P. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing*, **50** (2003), 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0
18. Y. Du, Y. Cai, M. Chen, W. Xu, H. Yuan, T. Li, A novel divide-and-conquer model for CPI prediction using ARIMA, Gray Model and BPNN, *Procedia Comput. Sci.*, **31** (2014), 842–851. https://doi.org/10.1016/j.procs.2014.05.335
19. P. F. Pai, C. S. Lin, A hybrid ARIMA and support vector machines model in stock price forecasting, *Omega*, **33** (2005), 497–505. https://doi.org/10.1016/j.omega.2004.07.024
20. Z. Zhao, Y. Zhang, Y. Yang, S. Yuan, Load forecasting via grey model-least squares support vector machine model and spatial-temporal distribution of electric consumption intensity, *Energy*, **255** (2022), 124468. https://doi.org/10.1016/j.energy.2022.124468
21. J. Riofrío, O. Chang, E. J. Revelo-Fuelagán, D. H. Peluffo-Ordóñez, Forecasting the consumer price index (CPI) of Ecuador: A comparative study of predictive models, *Int. J. Adv. Sci. Eng. Inf. Technol.*, **10** (2020), 1078–1084. https://doi.org/10.18517/ijaseit.10.3.10813
22. Y. E. Gur, Development and application of machine learning models in US consumer price index forecasting: Analysis of a hybrid approach, *Data Sci. Finance Econ.*, **4** (2024), 469–513. https://doi.org/10.3934/DSFE.2024020
23. A. I. Şimşek, Forecasting consumer price index using macroeconomic variables: A comparative analysis of machine learning and deep learning approaches, *Bingöl Üniver. Sosyal Bilimler Enstitüsü Dergisi*, **14** (2024), 15–29. https://doi.org/10.29029/busbed.1394983
24. H. Nie, G. Liu, X. Liu, Y. Wang, Hybrid of ARIMA and SVMs for short-term load forecasting, *Energy Procedia*, **16** (2012), 1455–1460. https://doi.org/10.1016/j.egypro.2012.01.229
25. J. Zhang, Z. Jiang, A new grey quadratic polynomial model and its application in the COVID-19 in China, *Sci. Rep.*, **11** (2021), 12588. https://doi.org/10.1038/s41598-021-91970-1
26. B. Scholkopf, A. J. Smola, *Learning with kernels: Support vector machines, regularization, optimization, and beyond*, MIT Press, 2001. https://doi.org/10.7551/mitpress/4175.001.0001
27. S. Abe, *Support vector machines for pattern classification*, London: Springer, 2005. https://doi.org/10.1007/1-84628-219-5
28. N. R. Babu, B. J. Mohan, Fault classification in power systems using EMD and SVM, *Ain Shams Eng. J.*, **8** (2017), 103–111. https://doi.org/10.1016/j.asej.2015.08.005
29. J. A. Mincer, V. Zarnowitz, The evaluation of economic forecasts, In: *Economic forecasts and expectations: Analysis of forecasting behavior and performance*, NBER, 1969, 3–46.
30. R. J. Hyndman, A. B. Koehler, Another look at measures of forecast accuracy, *Int. J. Forecasting*, **22** (2006), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

31. C. D. Lewis, *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*, 1982.
32. D. Harvey, S. Leybourne, P. Newbold, Testing the equality of prediction mean squared errors, *Int. J. Forecasting*, **13** (1997), 281–291. https://doi.org/10.1016/S0169-2070(96)00719-4
33. R. Giacomini, H. White, Tests of conditional predictive ability, *Econometrica*, **74** (2006), 1545–1578. https://doi.org/10.1111/j.1468-0262.2006.00718.x



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)