Proceedings of CONF-BPS 2026 Symposium: Innovation, Finance, and Governance for Sustainable Global Growth
DOI: 10.54254/2754-1169/2026.BJ30973

Forecast Breakdown of Inflation Models in the Post-COVID Era in the United States: Evidence from Classical Time-Series Methods

Sihan Wei

Finance Management School, Shanghai University of International Business and Economics, Shanghai, China
23053034@suibe.edu.cn

Abstract. The unprecedented inflation surge following the COVID-19 pandemic has raised serious concerns about the reliability of standard forecasting tools. This paper investigates whether inflation has become more difficult to forecast in the post-COVID era. Using monthly consumer price inflation data for the United States from 2000 to 2025, we compare the out-of-sample forecasting performance of several classical time-series models, including random walk specifications, exponential smoothing (ETS), and seasonal ARIMA models. We implement a rolling-window forecasting framework and evaluate one- to twelve-month-ahead predictions across pre-COVID and post-COVID subsamples. Our empirical results document a substantial deterioration (e.g., forecast errors roughly doubled for some specifications) in the forecast accuracy of standard inflation models after 2020. For most specifications, both the mean and variance of forecast errors increase significantly in the post-COVID subsample, indicating a breakdown in forecasts relative to the pre-COVID period. While more sophisticated models, such as ETS and SARIMA, outperform simple random walk benchmarks in normal times, their relative gains largely disappear during the high-inflation and high-uncertainty environment that followed the COVID-19 shock. These findings suggest that the inflation process has undergone significant structural changes, characterized by higher volatility and reduced predictability in recent years. This provides systematic evidence for a post-COVID' forecast breakdown' in classical univariate benchmarks. Consequently, the results have implications for central banks and professional forecasters, highlighting the challenges of relying on historical regularities and purely time-series-based models in an environment of frequent and large macroeconomic shocks.

Keywords: Inflation Forecast, Forecast Breakdown, Time-Series Models, ARIMA, COVID-19

1. Introduction

Inflation forecasting is a central component of modern macroeconomic policy. Central banks that target low and stable inflation rely on forecasts to set interest rates, guide expectations, and communicate with the public. The post-COVID-19 period, however, has seen one of the largest and

DOI: 10.54254/2754-1169/2026.BJ30973

most persistent inflation surges in advanced economies in decades, together with unusually large forecast errors. Benchmark models used by policy institutions significantly underpredicted the speed and scale of the inflation run-up in 2021–2022 [1]. Policymakers such as European Central Bank President Christine Lagarde have emphasized that frequent large shocks—including the COVID-19 pandemic and Russia's invasion of Ukraine—have made inflation more volatile and less predictable [2]. This experience raises a critical question: has the predictability of inflation fundamentally deteriorated in the post-pandemic era, even when standard, purely statistical time-series models are used? While structural shocks are a key part of the narrative, this paper focuses on a more foundational issue: the performance of purely statistical, univariate time-series models, which serve as critical benchmarks in the forecasting literature.

Recent work documents a marked rise in inflation uncertainty. Acharya, Hillenbrand, Venkateswaran, and Underwood construct a composite U.S. inflation-uncertainty index using news-based and market-based indicators, showing that uncertainty in the post-COVID period reached levels comparable to those of the 1970s and early 1980s [3]. Elevated inflation uncertainty is associated with higher risk premia and weaker real activity, underscoring the practical importance of understanding how well inflation can be forecast in this new environment [3].

The literature on inflation forecasting has traditionally followed two main lines. One uses structural or semi-structural models, especially Phillips-curve-type relationships and their New Keynesian extensions. The other relies on univariate time-series models—such as random walks, ARIMA, exponential smoothing (ETS), and unobserved components models—that treat inflation as a function of its own past values. A robust result from the pre-pandemic literature is that simple time-series benchmarks often perform surprisingly well. A very simple "no-change" forecast based on recent inflation frequently matches or outperforms more elaborate Phillips-curve models for U.S. inflation [4]. Stock and Watson report that Phillips curves add limited and unstable predictive content relative to univariate benchmarks, and that their performance varies substantially over time [5]. For the euro area, Bańbura and Bobeica document similar forecast instability for Phillips-curve-type models, even before the pandemic [6]. This surge in uncertainty directly challenges forecasting models. In such an environment, it is particularly valuable to examine the performance of parsimonious univariate models, as they provide a baseline against which the added value of more complex structural explanations can be measured.

The idea that inflation has become harder to forecast predates the COVID-19 pandemic. Stock and Watson show that, starting in the mid-1980s, the mean and volatility of U.S. inflation declined, but its predictability deteriorated: a range of models—including Phillips curves, factor models, and univariate time-series specifications—generated forecast errors that became larger and more unstable over time [7]. Giacomini and Rossi formalize this notion as forecast breakdown, defined as a deterioration in out-of-sample performance that cannot be reconciled with the model's in-sample fit, typically due to a structural change in the data-generating process [8]. Their framework is particularly well-suited for the post-COVID period, characterized by potential structural breaks in inflation dynamics due to unprecedented fiscal and monetary stimulus, supply chain disruptions, and geopolitical events. Their framework links forecast breakdowns directly to structural breaks and time variation in economic relationships.

Despite the rapidly growing body of literature on pandemic-era inflation, a systematic assessment of how classical univariate time-series models have fared remains limited. The evidence on benchmarks such as the random walk with drift, ARIMA, and ETS in the post-COVID period is particularly scarce. Simple univariate benchmarks continue to be widely used as baseline forecasts

DOI: 10.54254/2754-1169/2026.BJ30973

in policy institutions and applied work, yet systematic post-pandemic assessments of their performance and potential forecast breakdown are scarce.

The analysis utilizes monthly U.S. CPI inflation to estimate a set of widely used univariate models, including the random walk with drift, ARIMA, and trend–cycle decompositions, within rolling- and expanding-window frameworks. Pseudo-real-time forecasts at multiple horizons are generated before and after the COVID-19 shock and are evaluated using standard loss functions such as mean squared forecast error. To assess whether any deterioration in performance is statistically significant, the study employs forecast-breakdown-type tests, in the spirit of Giacomini and Rossi, together with the Diebold–Mariano test for equal predictive accuracy [8,9].

The paper makes three contributions. First, it provides updated, post-pandemic evidence on the absolute and relative forecast performance of classical univariate time-series models for inflation. Second, by combining rolling-window forecast evaluation with formal breakdown tests, it quantifies whether the COVID-19 episode marks a statistically significant deterioration in the performance of benchmark time-series models. Third, by comparing forecast horizons, it assesses whether the predictability loss is primarily a short-term phenomenon, driven by volatile supply shocks, or if it permeates the medium-term horizon that is crucial for monetary policy calibration, thereby having more profound implications for policymakers.”

2. Methodology

2.1. Data and computational setup

All computations were performed using the forecast package in R [10]. ETS models were estimated using the ets( ) function, while ARIMA/SARIMA models were estimated using the auto () function. Arima (), random walk forecasts via rwf (), and trend-cycle variant models via stlm (). Forecast results, accuracy metrics, and DM tests were obtained using the forecast (), accuracy (), and dm.test () functions, respectively [11].

Within this environment, the analysis utilizes monthly U.S. Consumer Price Index (CPI) data. The baseline series is seasonally adjusted headline CPI from the FRED database (CPIAUCSL), starting in January 2000 and extending to the latest available observation. Inflation is measured as year-over-year (YoY) log changes of the price level,

$$\pi_t = 100 \times [log\,(P_t) - log\,(P_{t-12})] \quad (1)$$

so that π_t represents the exact 12-month percentage change in CPI.

The first 12 observations are dropped, and all models are estimated on the univariate series $\{\boldsymbol{\pi_t}\}$. A break date of January 2020 separates “pre-COVID” and “post-COVID” forecasts targets. This date is chosen to precede the major economic disruptions caused by the pandemic and associated policy responses in the United States, thereby ensuring the post-COVID sample captures the full period of elevated inflation dynamics.

2.2. Univariate time-series models

The forecasting exercise focuses on standard univariate benchmarks widely used in the inflation forecasting literature [4,5,7].

This study evaluates the performance of three standard classes of univariate time-series models, chosen for their established role as benchmarks in the literature.

DOI: 10.54254/2754-1169/2026.BJ30973

Random walk with drift. The first benchmark is a random walk with drift,

$$\pi_t = \mu + \pi_{t-1} + \varepsilon_t \tag{2}$$

with constant drift μ and i.i.d. innovations ε_t. The special case μ=0 corresponds to the "no-change" forecast and has been shown to be difficult to beat for U.S. inflation [4,7].

Exponential smoothing (ETS). The second class consists of exponential smoothing models estimated in state-space form [10,11]. ETS specifications represent the series via unobserved level, trend, and seasonal components, with additive or multiplicative error. Models in the ETS(E,T,S) family (including simple exponential smoothing, Holt, and Holt–Winters) are estimated by maximum likelihood, and an automatic selection procedure chooses the structure that minimizes the corrected Akaike information criterion (AICc) [10,11]. Given that the series under study is the year-over-year inflation rate, which inherently removes seasonal patterns, this study restricts the model selection to non-seasonal ETS specifications (i.e., the seasonal component is set to S='N').

ARIMA and SARIMA. The third class is autoregressive integrated moving-average (ARIMA) and seasonal ARIMA (SARIMA) models. In compact notation, $\mathrm{ARIMA}(p,d,q)(P,D,Q)_{12}$ uses non-seasonal and seasonal autoregressive and moving-average polynomials with monthly seasonality [5]. Orders and differencing degrees are selected using the Hyndman–Khandakar automatic algorithm, which combines unit-root testing with a stepwise search over (p,q,P,Q) to minimize AICc [10]. The algorithm configuration allows for a maximum of one non-seasonal difference (d ≤ 1) and one seasonal difference (D ≤ 1), with the need for differencing determined by successive KPSS unit root tests. Parameters are estimated using maximum likelihood, and residual diagnostics verify the approximate white noise innovations [10,11].

2.3. Forecast evaluation and breakdown analysis

Forecast accuracy is summarized using squared and absolute errors. For model m and horizon h, over the set of targets $\mathcal{T}_h$, the mean squared forecast error (MSFE) and mean absolute error (MAE) are

$$\mathrm{MSFE}_h^{(m)} = \frac{1}{|\mathcal{T}_h|} \sum_{t+h\in\mathcal{T}_h} \left(e_{t+h}^{(m)}\right)^2,\ \mathrm{MAE}_h^{(m)} = \frac{1}{|\mathcal{T}_h|} \sum_{t+h\in\mathcal{T}_h} |\, e_{t+h}^{(m)} \,| \tag{3}$$

These measures are reported for the full sample and separately for pre-COVID and post-COVID subsamples, classified by the forecast target date.

Pairwise comparisons of predictive accuracy use the Diebold–Mariano (DM) test [10]. For two models, 1 and 2, and horizon h, the loss differential

$$d_{t+h} = L\left(e_{t+h}^{(1)}\right) - L\left(e_{t+h}^{(2)}\right) \tag{4}$$

has null hypothesis $\mathbb{E}[d_{t+h}] = 0$. The DM statistic is constructed as the sample mean of d_{t+h} divided by a HAC-robust standard error and is asymptotically standard normal [10]. Tests are applied on the full sample and on pre- and post-COVID subsamples.

The forecast breakdown, as assessed by Giacomini and Rossi, is evaluated by comparing MSFE across regimes. For each model and horizon, the MSFE is computed separately for the pre-COVID and post-COVID subsamples. A forecast breakdown is inferred if the post-COVID MSFE shows a statistically significant increase relative to the pre-COVID period. This is formally assessed using a

DOI: 10.54254/2754-1169/2026.BJ30973

two-sample Fisher's F-test for the equality of variances (MSFE) and supplemented by visual analysis of rolling-window MSFE plots to identify the timing of any performance deterioration [8].

3. Empirical result

3.1. Data & descriptive evidence

The descriptive statistics and visual evidence presented in Figure 1 and Table 1 highlight a significant shift in U.S. inflation dynamics around the year 2020. The sample, covering January 2000 onwards, is divided into two periods: pre-COVID (January 2000 – January 2020) and post-COVID (February 2020 – present) for analysis.

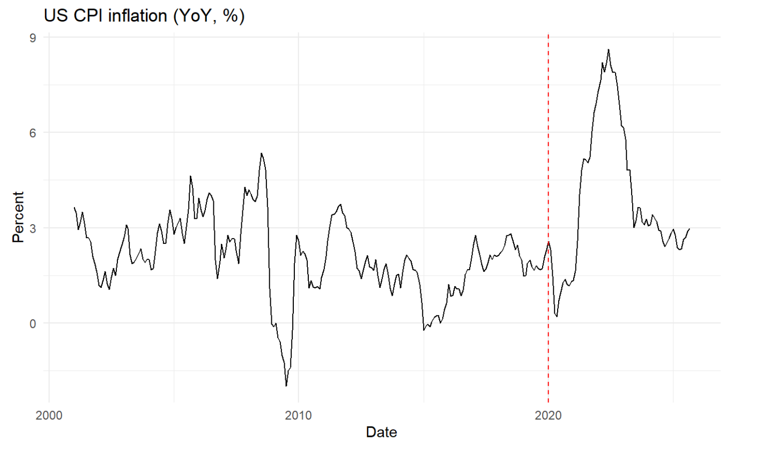

Figure 1. U.S. CPI inflation, 2000–2025

Prior to the global financial crisis, inflation rates fluctuated primarily between 2% and 4%. During the crisis, inflation briefly turned negative in 2009 before returning to the 1% to 3% range for most of the 2010s. By contrast, the post-COVID era has witnessed an unprecedented surge, as inflation rapidly climbed from near-zero levels in 2020 to a peak of over 8% in 2022. Although it has since moderated, it remains above pre-COVID averages. Inflation since 2020 has not only been higher on average but also significantly more volatile.

Table 1. Summary statistics for U.S. CPI inflation

	N	Mean	SD	Min	Max
Full Sample	297	2.504324	1.705541	-1.9781990	8.617126
Pre-COVID	228	2.080185	1.215096	-1.9781990	5.351718
Post-COVID	69	3.905829	2.266518	0.1980051	8.617126

Throughout the observation period, headline CPI inflation averaged approximately 2.5% annually, with a standard deviation of 1.7 percentage points, fluctuating broadly between -2.0% and 8.6%. Prior to the COVID-19 pandemic, the mean was lower and volatility was more contained, with average inflation standing at approximately 2.1%, and a standard deviation of 1.2 percentage points, consistently remaining within the range of [–2.0%, 5.4%]. After 2020, both indicators rose significantly. The post-COVID mean rose to approximately 3.9%, nearly double the pre-COVID average, while the standard deviation expanded to about 2.3 percentage points. The peak during this period occurred in 2022 when inflation exceeded 8%, while the trough remained positive, indicating

DOI: 10.54254/2754-1169/2026.BJ30973

no deflationary episodes in the post-pandemic sub-sample. These characteristics confirm that inflation outcomes in the post-pandemic era exhibit higher levels and greater dispersion.

The descriptive evidence points to a clear regime change around 2020, U.S. headline inflation becomes both higher on average and markedly more volatile. This provides a natural context for examining whether standard univariate time-series models experience a deterioration in forecast performance when moving from the pre-COVID to the post-COVID environment.

3.2. Rolling one-step-ahead forecast performance

Given the documented rise in inflation volatility, we next examine whether this led to a systematic deterioration in the forecasting performance of the univariate benchmarks.

Table 2. Rolling one-step-ahead forecast accuracy for headline CPI inflation

Model	Period	MAE	RMSE
RW	Full	0.2259045	0.2770018
RW	Pre-COVID	0.3297311	0.4475213
RW	Post-COVID	0.2663793	0.3534002
ETS	Full	0.2117522	0.2658036
ETS	Pre-COVID	0.3125211	0.3972569
ETS	Post-COVID	0.2510350	0.3234652
ARIMA	Full	0.1693516	0.2262567
ARIMA	Pre-COVID	0.2933275	0.6281961
ARIMA	Post-COVID	0.2176812	0.4302032

Table 2 reports the rolling one-step-ahead forecasting accuracy for the Random Walk (RW), ETS, and ARIMA models. The evaluation sample is divided into pre-COVID and post-COVID periods based on the target forecast date, while also presenting complete results from 2010 to the latest period. The table lists the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) for each model and subsample.

All three models exhibit a decline in predictive performance following the COVID-19 pandemic. Taking the random walk benchmark model as an example, its MAE increased from 0.23 to 0.33 prior to the COVID-19 pandemic, while its RMSE rose from 0.28 to 0.45. The ETS model exhibited a similar trend, with the MAE increasing from 0.21 to 0.31 and the RMSE rising from 0.27 to 0.40. The ARIMA model exhibited the most pronounced relative deterioration: its pre-COVID MAE of approximately 0.17 was the lowest among all models, yet it rose to about 0.29 post-COVID; RMSE surged from approximately 0.23 to 0.63. Thus, despite ARIMA's superior pre-COVID performance, its prediction dispersion showed the largest increase post-COVID.

When examining the full sample period from 2010 to the present, the ETS model and RW benchmark model exhibit comparable accuracy, with mean errors ranging from 0.25 to 0.27 and root mean square errors ranging from 0.32 to 0.35. The ARIMA model maintained a slightly lower mean error (approximately 0.22) throughout the entire sample period; however, its RMSE significantly increased to around 0.43, reflecting several large forecast errors in the post-COVID subsample. Overall, the rolling evaluation indicates that all univariate models exhibit systematically higher one-step forecast errors in the post-COVID environment, and their relative rankings are unstable across

DOI: 10.54254/2754-1169/2026.BJ30973

different subsamples—the ARIMA model performed excellently before the pandemic but saw its advantage significantly diminish afterward.

3.3. Diebold–Mariano tests and evidence of forecast breakdown

To statistically evaluate whether the observed changes in forecast accuracy represent a significant deterioration in relative model performance, this study employs the Diebold-Mariano test.

Table 3. Diebold–Mariano tests results

Model 1	Model 2	Sample	DM-stat	P-value
ETS	RW	Full	-1.6227022	0.1064429
ETS	RW	Pre-COVID	-0.7379686	0.4621490
ETS	RW	Post-COVID	-1.4511584	0.1513336
ARIMA	RW	Full	0.5030258	0.6155751
ARIMA	RW	Pre-COVID	-2.2215849	0.0284170
ARIMA	RW	Post-COVID	0.6329796	0.5288700

Table 3 presents the results of the Diebold–Mariano (DM) test, which compares ETS and ARIMA models against a random walk (RW) benchmark model. The loss difference is defined as the mean squared error of the first model minus the mean squared error of the RW model; thus, a negative DM statistic indicates that the first model is expected to yield lower losses. The results cover the full evaluation sample as well as pre- and post-COVID subsamples.

In the ETS versus RW comparison, the DM statistic is negative across all samples, with values of approximately -1.62 for the full sample, -0.74 for the pre-COVID period, and -1.45 for the post-COVID period. This indicates that ETS generally produces slightly lower losses than the random walk model. However, the corresponding p-values did not fall below conventional significance thresholds: approximately 0.11 for the full sample, 0.46 for the pre-COVID period, and 0.15 for the post-COVID period. Thus, ETS cannot be statistically proven to outperform the random walk benchmark significantly.

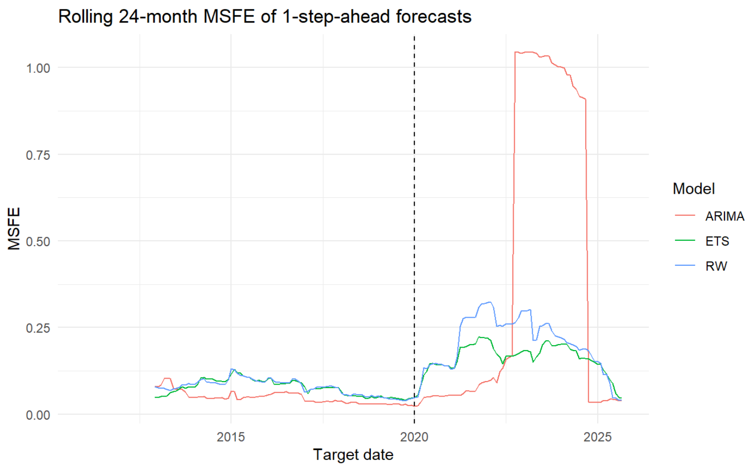

Figure 2. Rolling 24-month MSFE of one-step-ahead inflation forecasts

In the comparison between the ARIMA model and the random walk (Figure 2), the differences across subsamples exhibit more pronounced asymmetry. During the pre-pandemic phase, the

DOI: 10.54254/2754-1169/2026.BJ30973

difference statistic was approximately -2.22 with a p-value of around 0.03, indicating that the ARIMA model achieved significantly lower prediction losses than the random walk benchmark prior to the outbreak. Conversely, the DM statistics for the full sample and post-pandemic period were approximately 0.50 and 0.63, respectively, with p-values around 0.62 and 0.53. Both values were positive and far from significant, indicating that the ARIMA model no longer significantly outperformed the RW model after incorporating post-pandemic observations. RW and ETS also exhibit higher MSFE after 2020, but their increases are more moderate and broadly similar to each other.

Taken together, the evidence suggests a breakdown in the forecast for the ARIMA model [8]. This is most clearly seen in the DM tests, where ARIMA's significant pre-COVID advantage (p-value < 0.05) vanishes entirely in the post-COVID sample, resulting in no significant difference from the naive RW benchmark. While the RW and ETS models also became less accurate in absolute terms after 2020, their relative performance remained stable. This suggests that simpler, more parsimonious benchmarks demonstrated greater robustness to the structural shifts characterizing the pandemic era.

4. Discussion

4.1. Key findings: a post-COVID forecast breakdown

The central message of our empirical analysis is unequivocal: the COVID-19 pandemic precipitated a significant deterioration, or "breakdown," in the predictive power of standard univariate models for U.S. inflation. The empirical results indicate that U.S. headline inflation was meaningfully forecastable by simple univariate models in the pre-COVID period, but that this forecastability deteriorated once the pandemic-era inflation surge began. Before 2020, the ARIMA specification achieved the lowest one-step-ahead forecast errors among the benchmarks, with substantially smaller MAE and RMSE than the random-walk model and ETS. This pattern is consistent with earlier evidence that, during periods of relatively low and stable inflation, parsimonious time-series models can exploit autoregressive structure in prices to deliver non-trivial gains over naive benchmarks [4]. At the same time, the difference between ETS and the random walk is modest and rarely statistically significant, underlining the well-known difficulty of extracting large gains over simple models when inflation is already relatively well anchored [5].

After 2020, forecast errors rise sharply for all three models, in line with the descriptive evidence of higher mean inflation and greater volatility. The MAE and RMSE roughly double for the random-walk and ARIMA specifications, and rolling MSFE plots reveal a pronounced spike in loss around the 2021–2022 inflation surge. In relative terms, the ARIMA model—previously the clear winner—suffers the largest deterioration in RMSE, reflecting several extreme forecast misses. Diebold–Mariano tests confirm that ARIMA significantly outperforms the random walk before COVID-19, but no longer delivers statistically superior accuracy in the post-COVID subsample. In contrast, ETS remains statistically indistinguishable from the random walk throughout [9]. These patterns are characteristic of a forecast breakdown in the sense of Giacomini and Rossi, where a model that once dominated a benchmark fails to retain its advantage after a structural change in the data-generating process [8].

DOI: 10.54254/2754-1169/2026.BJ30973

4.2. Interpreting the breakdown: from shocks to structural change

Unprecedented Shock Composition: The breakdown of pre-pandemic forecasting relationships is naturally interpreted against the backdrop of the pandemic-era inflation episode. Recent work attributes the surge in U.S. inflation to a combination of large sectoral supply shocks, shifts in relative demand, and exceptionally strong fiscal support, all of which arrived in rapid succession [1,2]. Such shocks differ markedly from the disturbances that shaped inflation dynamics during the Great Moderation, and their propagation through supply chains and labor markets is unlikely to be well captured by fixed-parameter univariate processes. The single-equation models considered here implicitly assume that the persistence and variance of inflation innovations are stable over time; the post-COVID data contradict this assumption by exhibiting both a higher mean and more frequent large price changes [3,7].

Elevated Volatility and Fat Tails: Increased volatility and fat-tailed disturbances provide a second channel through which the models can break down. The descriptive statistics and rolling standard deviation plots indicate a significant increase in short-horizon volatility after 2020, while the rolling MSFE series displays a sharp and persistent rise in forecast loss, particularly for the ARIMA model. This pattern echoes earlier evidence that the predictability of macroeconomic series is tightly linked to the broader macroeconomic environment: periods of greater macroeconomic stability tend to be associated with more predictable inflation and output, whereas turbulent regimes reduce the relative importance of systematic components in favor of idiosyncratic shocks [12]. In other words, the forecast ability of the series—the proportion of variance explainable by its own past—appears to have declined. In the present context, the univariate ARIMA specification appears to overfit the stable pre-pandemic regime. Then it generates large errors when confronted with the rare, extreme shocks of the pandemic era. The random-walk and ETS benchmarks, being more parsimonious and less tightly parameterized, react more slowly to regime changes but also produce fewer extreme failures when historical relationships no longer hold.

Parameter Instability and Structural Change: Finally, the regime-dependent performance of ARIMA provides a direct link to the literature on forecast breakdown. The evidence that ARIMA yields significantly lower loss than the random walk pre-COVID but not post-COVID is precisely the type of regime-dependent performance that Giacomini and Rossi emphasize as diagnostic of breakdowns in predictive relationships [8]. This parameter instability, as evidenced by the regime-dependent performance of ARIMA, is indicative of a larger structural change in the inflation process, as formalized in the forecast breakdown literature [8]. The post-COVID era may therefore represent not merely a sequence of large shocks, but a shift to a new, more volatile inflation regime with distinct dynamic properties. The rolling MSFE profiles reinforce this interpretation by showing that the deterioration is not confined to a single outlying episode but reflects a sustained increase in loss. In that sense, the post-COVID period resembles earlier historical episodes in which apparent gains from sophisticated models vanished once the underlying inflation process shifted [12].

4.3. Implications for forecasting practice and policy

From a policy perspective, the results highlight the risks of relying mechanically on models calibrated to pre-pandemic data when forming inflation projections in a radically altered environment. The pandemic-era surge in prices has already prompted extensive debate on the appropriate interpretation of inflation shocks and on the responsiveness of monetary policy [1,2]. The evidence here suggests that even simple, historically successful time-series models substantially understated the magnitude and persistence of post-COVID inflation. For central banks and policy institutions, this underscores the need to complement model-based forecasts with real-time judgment and broader information sets, particularly when the economy is hit by shocks that lie outside the range observed in the estimation sample.

At the same time, the relative robustness of the random-walk and ETS benchmarks carries a practical message for forecast users. This pattern underscores a classic trade-off in forecasting: sophistication for efficiency during calm periods versus parsimony for robustness during turbulent ones. Although ARIMA dominates in tranquil periods, its post-COVID performance is fragile. In contrast, the random-walk and ETS models, while never dramatically superior, maintain more stable performance across regimes. For real-time applications in uncertain environments, the stability of simpler benchmarks may outweigh the occasional superior accuracy of more complex models. This pattern aligns with the view that simple benchmarks should play a central role in forecast evaluation and communication, both as transparent reference points and as safeguards against overfitting in more complex specifications [11].

The findings also highlight the design of models intended for real-time use when the underlying environment is evolving. This underscores a broader concern about robustness in macroeconometrics: models optimized for a specific historical regime may fail catastrophically during regime shifts. Recent work on estimating macroeconomic VARs after March 2020 emphasizes the need to treat pandemic observations as extreme events that can distort standard estimation procedures unless appropriately down-weighted or modeled separately [13]. A similar logic applies to univariate inflation models: the inclusion of a small number of unprecedented shocks can substantially alter parameter estimates and degrade out-of-sample performance if the model structure does not accommodate regime changes or time-varying volatility.

Consequently, the quest for reliable inflation forecasts in the post-pandemic world may necessitate a fundamental re-evaluation of modeling paradigms, moving beyond stationary, linear univariate frameworks towards approaches that explicitly account for structural breaks, time-varying parameters, and the integration of heterogeneous data sources.

5. Conclusion

This paper investigates whether U.S. headline consumer price inflation has become more challenging to forecast in the post-COVID period when employing standard univariate time-series models. Using monthly CPI inflation since 2000 and a rolling pseudo–real-time evaluation, the analysis compared the performance of a random-walk benchmark, an ETS specification, and an ARIMA model across pre- and post-COVID subsamples. Before the pandemic, ARIMA delivered the lowest one-step-ahead forecast errors and significantly outperformed the random walk, according to Diebold–Mariano tests, indicating that inflation was meaningfully predictable from its own lagged values in a relatively stable environment. After 2020, however, forecast errors rise sharply for all models, and ARIMA's advantage disappears; its RMSE even becomes the highest of the three. ETS and the random walk show more modest increases in loss and remain statistically indistinguishable from each other and from ARIMA in the post-COVID subsample. Rolling MSFE profiles corroborate these findings by revealing pronounced and persistent spikes in forecast loss around the recent inflation surge, particularly for ARIMA models.

These results align with broader evidence that the pandemic and related shocks have fundamentally altered inflation behavior. Beyond confirming the altered inflation landscape, this study makes a distinct methodological contribution—recent work documents exceptionally high inflation uncertainty, with post-COVID levels comparable to those of the 1970s. Structural analyses attribute the inflation surge to a combination of pandemic-related supply constraints, energy price

DOI: 10.54254/2754-1169/2026.BJ30973

shocks, and expansionary fiscal and monetary policies, which are difficult to capture with models estimated on pre-pandemic data. The findings presented here complement the existing literature by demonstrating that even purely statistical benchmarks—often considered robust “workhorse” tools in central banks—have experienced a form of forecast breakdown in the sense of Giacomini and Rossi—a situation where a model’s out-of-sample performance deteriorates significantly relative to its in-sample fit, typically signaling an underlying structural shift: where a model that once dominated a benchmark fails to retain its advantage after a structural change in the data-generating process.

From both a policy and practical perspective, the evidence highlights the limitations of relying on fixed-parameter models calibrated during the Great Moderation to guide inflation forecasts in the current regime. When the underlying inflation process is subject to large and infrequent shocks, simple but parsimonious benchmarks such as random-walk and exponential-smoothing models may offer greater robustness than more tightly parameterized ARIMA specifications, even if they are less efficient in tranquil periods. At the same time, the large post-COVID forecast errors for all models underline the need for approaches that explicitly allow for time variation and structural change, including time-varying-parameter or regime-switching models, as well as frameworks that integrate economic structure—such as Phillips curves, expectations, and relative-price indicators—into multivariate forecasting systems.

The analysis is subject to several limitations. It focuses on a single country, a single inflation measure, and a narrow set of univariate models, relying on ex-post revised data rather than real-time vintages. Promising avenues include: (1) extending the analysis to core inflation and international comparisons; (2) evaluating more flexible model classes like time-varying parameter or machine learning approaches; and (3) examining whether incorporating high-frequency indicators or expectations data can mitigate the post-COVID forecast deterioration. Future work could extend the framework to core and alternative inflation measures, to other advanced economies, and to richer model classes, including nonlinear and machine-learning methods, while also evaluating density forecasts and tail risks. Nonetheless, the evidence presented here conveys a clear message: in the aftermath of COVID-19, U.S. inflation has become significantly more challenging to forecast, even for well-established univariate benchmarks, and model choice and forecast evaluation must be reconsidered in light of this new, more uncertain environment.

References

[1] Bernanke, B. S., & Blanchard, O. (2023). What caused the U.S. pandemic-era inflation? Brookings Papers on Economic Activity, 2023(Spring).

[2] Gopinath, G. (2023). Crisis and monetary policy. Finance & Development, 60(1), 15–18.

[3] Acharya, V. V., Hillenbrand, S., Venkateswaran, V., & Underwood, M. (2025). Inflation uncertainty: Measurement, causes, and consequences (Working paper). SSRN.

[4] Atkeson, A., & Ohanian, L. E. (2001). Are Phillips curves useful for forecasting inflation? Federal Reserve Bank of Minneapolis Quarterly Review, 25(1), 2–11.

[5] Stock, J. H., & Watson, M. W. (2008). Phillips curve inflation forecasts (NBER Working Paper No. 14322). National Bureau of Economic Research.

[6] Bańbura, M., & Bobeica, E. (2023). Does the Phillips curve help to forecast euro area inflation? International Journal of Forecasting, 39(1), 364–390.

[7] Stock, J. H., & Watson, M. W. (2007). Why has U.S. inflation become harder to forecast? Journal of Money, Credit and Banking, 39(s1), 3–33.

[8] Giacomini, R., & Rossi, B. (2009). Detecting and predicting forecast breakdowns. Review of Economic Studies, 76(2), 669–705.

[9] Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3), 253–263.

DOI: 10.54254/2754-1169/2026.BJ30973

[10] Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast Package for R. Journal of Statistical Software, 27(3), 1–22.
[11] Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and Practice (3rd ed.). OTexts.
[12] D'Agostino, A., Giannone, D., & Surico, P. (2006). (Un)Predictability and macroeconomic stability (ECB Working Paper No. 605). European Central Bank.
[13] Lenza, M., & Primiceri, G. E. (2022). How to estimate a vector autoregression after March 2020. Journal of Applied Econometrics, 37(4), 688–699.