Lund University
School of Economics and Management

Fuel for Thought: Predicting Gasoline Prices

Alexander Eriksson Byström, Anselm Etzelmüller, Juan Sepulveda Odriozola

Replication Code: https://github.com/Alexerby/NEKN34-A1-Submission

1 Introduction

The average consumer in the United States spends 17% of their total annual expenditures on transportation, of which roughly one fifth—3.2% of total spending or $2,449 per year—goes towards gasoline (Bernard 2024). For lowest-income households, the share of consumption spent on gasoline is much higher and amounts to 18 per cent of their income (Vaidyanathan 2021).

Meanwhile, the retail price of gasoline is highly volatile, much more than those of other consumption goods. Especially in a car-dependent society like the US, changes in gasoline prices are heavily influencing the affluence of many individuals. Therefore, gasoline price forecasts are relevant for, but not limited to, consumers. Baumeister et al. (2017) list several other areas of the economy where gasoline prices play an important role, e.g. home prices, inflation expectations, vehicle purchase decisions, and the revenue of taxes on gasoline.

This puts the forecasting of gasoline prices in the spotlight as a means to facilitate consumption, central bank and economic modeling decisions. Nonetheless, forecasting gasoline prices is no easy task due to low predictability and dependency on many national and global conditions and variables.

Therefore, this paper attempts to forecast the inflation rate of the US retail price of gasoline for the month of October 2014. In the process, several time series models are evaluated in regard to their forecasting ability.

The remainder of this paper is organized as follows. Section 2 provides a descriptive analysis of the dataset and outlines our identification strategy for both univariate and bivariate models. Building on this framework, Section 3 details the model selection process, justifying the specific candidate specifications chosen for our analysis. Finally, in Section 4 we conduct the forecast based on the two models we consider best-performing. Section 5 concludes.

2 Methodology and Data Diagnostics

Before specifying our models, we explore the time-series properties of `gaspriceinflation`. Visual inspection of Figure 1 suggests that the series is mean-reverting, a characteristic typical of commodity inflation rates. This is statistically confirmed by the Augmented Dickey-Fuller (ADF) tests in Table 1, which reject the null hypothesis of a unit root for all variables, indicating they are integrated of order zero, $I(0)$.

Table 1: Descriptive Statistics, Lagged Correlations, and Unit Root Tests

Variable	Mean	Std. Dev.	Corr (y_t, x_{t-1})	ADF (t-stat)
`gaspriceinflation`	0.51	5.13	1.00	-5.78***
`spotpricegasinflation`	1.04	9.84	0.66	-5.29***
`rac_wtichange`	0.89	7.90	0.58	-6.27***
`brentpriceinflation`	0.89	7.90	0.58	-6.27***
`brentspotpriceinflation`	0.98	8.73	0.54	-6.11***
`WTIspotpriceinflation`	0.90	8.18	0.54	-9.09***
`uscpichange`	0.23	0.26	0.30	-3.77***
`proxycrudeinventorieschange`	0.09	1.08	0.11	-4.28***
`cfnai`	-0.10	0.72	0.09	-3.15**
`consumpchange`	0.12	3.43	0.09	-4.15***
`rea`	-1.06	23.80	0.06	-3.28**
`crudeproductionchange`	0.10	1.08	-0.03	-9.37***

Note: ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.
The ADF test null hypothesis is the presence of a unit root.

Furthermore, Figure 1 reveals a structural shift in volatility around the year 1999. The series transitions from a stable period to a high-variance phase, with the standard deviation (σ) increasing from 2.92% in the pre-1999 sample to 6.35% in the subsequent period.

To identify the appropriate univariate lag structure, we examine the Autocorrelation (ACF) and Partial Autocorrelation (PACF) functions in Figure 2. The ACF shows a slow decay with significant annual spikes, suggesting seasonality and an underlying autoregressive process. While a pure AR(p) model would imply a PACF that truncates after p lags, the continued persistence in Figure 2b suggests that a hybrid ARMA(p, q) is more adequate.

We explicitly exclude integrated (ARIMA) models due to the stationarity confirmed by the ADF test ($d = 0$). Similarly, pure MA(q) models are disregarded as baseline candidates based on the diagnostic evidence in Figures 2a and 2b. The ACF exhibits significant persistence with a slow decay, while the PACF displays prominent spikes at early lags, most notably at lags 1 and 2 before largely cutting off. This combination strongly suggests that the underlying dynamics are governed by an autoregressive process. A pure MA specification would fail to capture the clear autoregressive momentum identified in the partial correlations, therefore, we consider both ARMA and AR processes.

For the multivariate extensions, we identify potential predictors by computing the correlation between the retail `gaspriceinflation` and one-period lagged exogenous variables (see Table 1). Brent crude oil inflation (`brentpriceinflation`) is identified as the primary candidate for the bivariate models.

This choice is motivated by the high observed correlation (0.58) and the direct economic transmission mechanism where crude oil serves as the primary input cost for retail gasoline production. While other oil-related variables—specifically `rac_wtichange`—could have served as suitable candidates, our selection of `brentpriceinflation` is motivated by its status as a more liquid and widely traded global benchmark. Given that the contemporaneous correlation between Brent and WTI in our sample is near-perfect ($r = 0.9999$), the two series are statistically interchangeable; therefore,

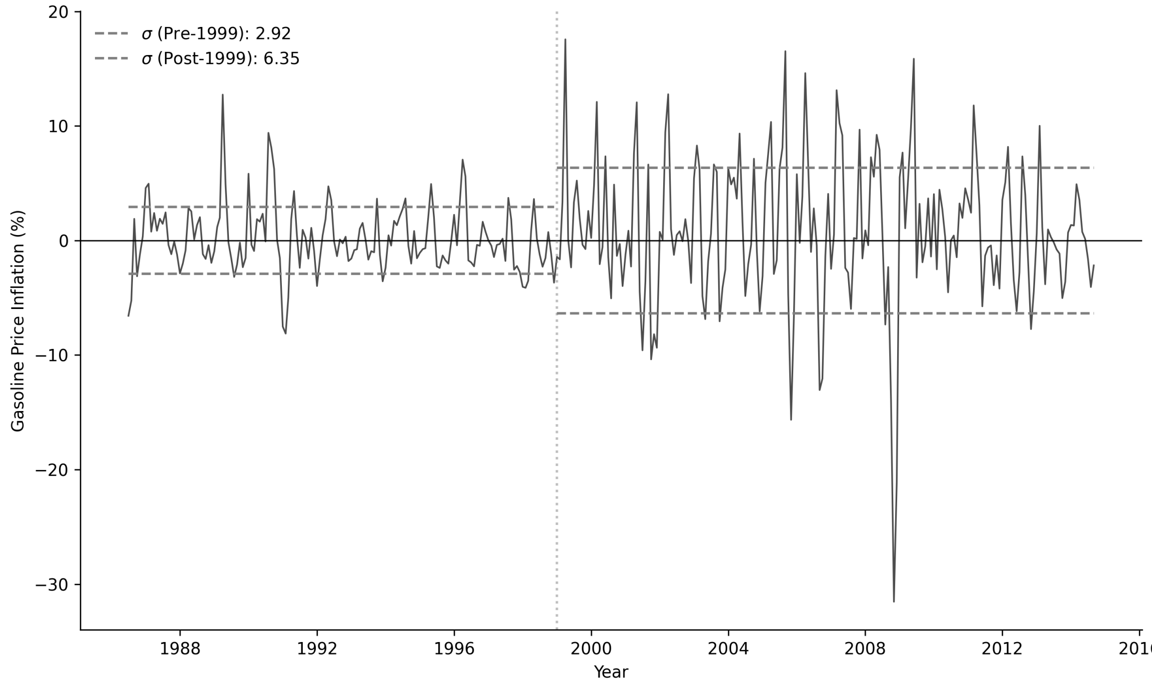
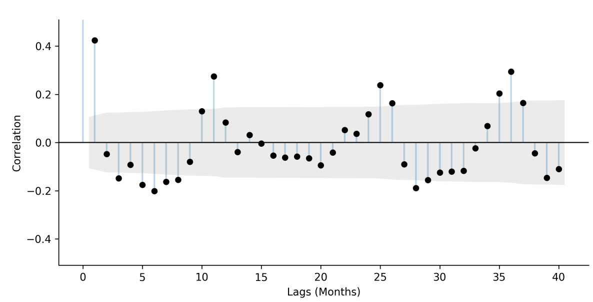
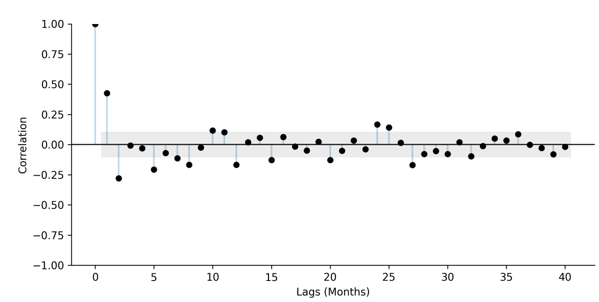


Figure 1: Gasoline Price Inflation (July 1986 – September 2014).



(a) ACF of Gasoline Prices



(b) PACF of Gasoline Prices

Figure 2: Autocorrelation and Partial Autocorrelation Functions

selecting the globally dominant benchmark is the simple choice.

Finally, we must decide on the optimal model specification among the candidates. We utilize two complementary approaches: in-sample Information Criteria (AIC and BIC) and out-of-sample predictive performance measured by the Mean Squared Prediction Error (MSPE). By evaluating both, we ensure that our chosen model is not only well-fitted to historical data but also robust in generating the October 2014 forecast.

2.1 Model Selection Strategy

2.1.1 Univariate Specification

To identify the optimal specifications for forecasting, we utilize a two-stage grid search strategy. For the univariate case, we evaluate AR(p) models for all $p \in \{1, \ldots, 20\}$ and ARMA(p, q) models where $p, q \in \{1, \ldots, 10\}$.

The selection of primary candidates is based on two complementary approaches:

1. **Information Criteria:** We employ the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) to assess in-sample fit. While the AIC tends to favor more complex models to capture additional dynamics, the BIC imposes a stricter penalty on the number of parameters ($k \ln(n)$) to prioritize model parsimony.
2. **Predictive Performance:** We evaluate out-of-sample forecasting performance by minimizing the Mean Squared Prediction Error (MSPE).

If both approaches converge on the same specification, it lends additional credibility to the model's robustness. Based on these criteria, we select four primary univariate candidates to carry forward for final evaluation.

2.1.2 Bivariate Specification

To incorporate the influence of crude oil markets on retail gasoline prices, we extend our framework to include exogenous predictors via the ARIMAX(p, d, q) framework. The general specification is given by:

$$\Delta^d y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \Delta^d y_{t-i} + \sum_{k=1}^{r} \beta_k x_{t-k} + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-k} \tag{1}$$

where y_t is the retail gasoline price inflation and x_{t-j} represents lagged Brent crude oil inflation. Based on our empirical diagnostics, we simplify this general form for our initial grid search. First, since the ADF tests confirmed all variables are stationary ($I(0)$), we set the order of integration $d = 0$. While we primarily prioritize the more parsimonious ARX(p, r) model where $q = 0$, we also evaluate ARIMAX($p, 0, q$) specifications to determine if a moving average component captures additional short-term error dynamics.

Models in both the univariate and the bivariate case include a constant α_0. This is important, as the mean of our variable is not exactly 0.

The primary bivariate candidate, the ARX(p, r) model, is defined as:

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j=1}^{r} \beta_j x_{t-j} + \epsilon_t \tag{2}$$

The distributed-lag structure for the exogenous component (r) is motivated by the "transmission mechanism", where global oil price shocks may take several months to filter through the refining and distribution chain to reach the retail pump. By treating Brent inflation as a strictly exogenous regressor we assume that the feedback from local retail markets to global oil indices are negligible.

In the same manner as in the univariate case, we will be fitting a grid of models for both ARX and ARIMAX specifications, evaluating them by the information criteria and then finally using MSPE to determine on an ultimate model to carry forward.

2.2 Evaluation Framework: MSPE and Diebold-Mariano

To evaluate out-of-sample forecasting performance, we reserve the final 36 observations of the sample (roughly 10% of the data) as a hold-out period. We generate recursive one-step-ahead forecasts and utilize the Mean Squared Prediction Error (MSPE) as our primary metric for comparison.

Furthermore, to assess whether the differences in MSPE between candidate models are statistically significant, we employ the Diebold-Mariano (DM) test. The DM test evaluates the null hypothesis that two competing models have equal predictive accuracy based on a loss function (in our case, the squared error). This allows us to distinguish between numerical superiority and true statistical improvement in a high-volatility environment.

3 Forecasting Performance

3.1 Univariate Forecasting

The results of the univariate grid search are presented in Table 2. As expected, we observe a divergence between the two information criteria: the AIC selects the relatively complex AR(15) and ARMA$(8, 9)$, while the BIC favors the more parsimonious AR(2) and ARMA$(1, 3)$.

To evaluate out-of-sample performance, we conduct recursive one-step-ahead forecasting for the final 36 observations. The results are visualized in Figure 3, with detailed MSPE values for all candidate models provided in Table 4.

The comparison of MSPE values confirms that the parsimonious models outperformed the complex specifications. The ARMA$(1, 3)$ emerged as the top-performing univariate model, while the ARMA$(8, 9)$ yielded the highest error.

However, a lower MSPE does not automatically imply a statistically significant improvement. To determine if the ARMA$(1, 3)$ genuinely provides a better forecast than its competitors, we conduct a Diebold-Mariano (DM) test. The full test results are provided in (Table 5).

The test results suggest that the differences in predictive accuracy are not statistically significant at the 5% level. Specifically, the high volatility in gasoline inflation makes it difficult to statistically distinguish between the candidate models based on their MSPE.

Table 2: Information Criteria for AR and ARMA Model Specifications

Panel A: Top 5 AR Models by AIC					
Metric	AR(15)	AR(12)	AR(16)	AR(13)	AR(17)
AIC	1955.95	1956.16	1956.73	1958.03	1958.63
Panel B: Top 5 AR Models by BIC					
Metric	AR(2)	AR(5)	AR(8)	AR(3)	AR(6)
BIC	1998.27	2000.73	2003.12	2004.07	2004.93
Panel C: Top 5 ARMA Models by AIC					
Metric	ARMA(8,9)	ARMA(6,10)	ARMA(7,9)	ARMA(8,10)	ARMA(7,10)
AIC	1949.31	1949.44	1950.76	1950.81	1952.27
Panel D: Top 5 ARMA Models by BIC					
Metric	ARMA(1,3)	ARMA(2,2)	ARMA(2,0)	ARMA(2,3)	ARMA(1,4)
BIC	1994.01	1997.42	1998.27	1998.28	1998.90

Note: Panels A and B rank pure autoregressive models based on AIC and BIC. Panels C and D rank models from the $ARMA(p, q)$ search space.

Table 3: Information Criteria for Bivariate ARX and ARIMAX Model Specifications

Panel A: Top 5 ARX Models by AIC					
Metric	ARX(9,17)	ARX(9,11)	ARX(9,14)	ARX(10,11)	ARX(9,15)
AIC	1739.54	1739.73	1740.13	1740.17	1740.25
Panel B: Top 5 ARX Models by BIC					
Metric	ARX(9,4)	ARX(10,4)	ARX(9,1)	ARX(9,3)	ARX(9,5)
BIC	1799.23	1804.59	1804.71	1804.78	1804.88
Panel C: Top 5 ARIMAX Models by AIC					
Metric	ARIMAX(10,0,7)	ARIMAX(10,0,8)	ARIMAX(10,0,10)	ARIMAX(7,0,4)	ARIMAX(6,6,4)
AIC	1779.82	1781.71	1782.15	1784.89	1784.98
Panel D: Top 5 ARIMAX Models by BIC					
Metric	ARIMAX(2,5,1)	ARIMAX(2,5,3)	ARIMAX(2,5,2)	ARIMAX(2,5,4)	ARIMAX(9,0,4)
BIC	1839.02	1841.33	1843.68	1844.52	1846.25

Note: Panels A and B rank ARX(p, r) models. Panels C and D rank ARIMAX(p, q, r) models. All specifications use `brentpriceinflation` as the exogenous regressor.

Table 4: Rolling MSPE Results Comparison

Univariate Results

Model	MSPE
ARMA(1,3)	10.15
AR(2)	11.05
AR(15)	11.46
ARMA(8,9)	12.26

Bivariate Results

Model	MSPE
ARIMAX(2,5,1)	7.56
ARX(9,4)	9.54
ARX(9, 17)	9.63
ARIMAX(10,0,7)	9.89

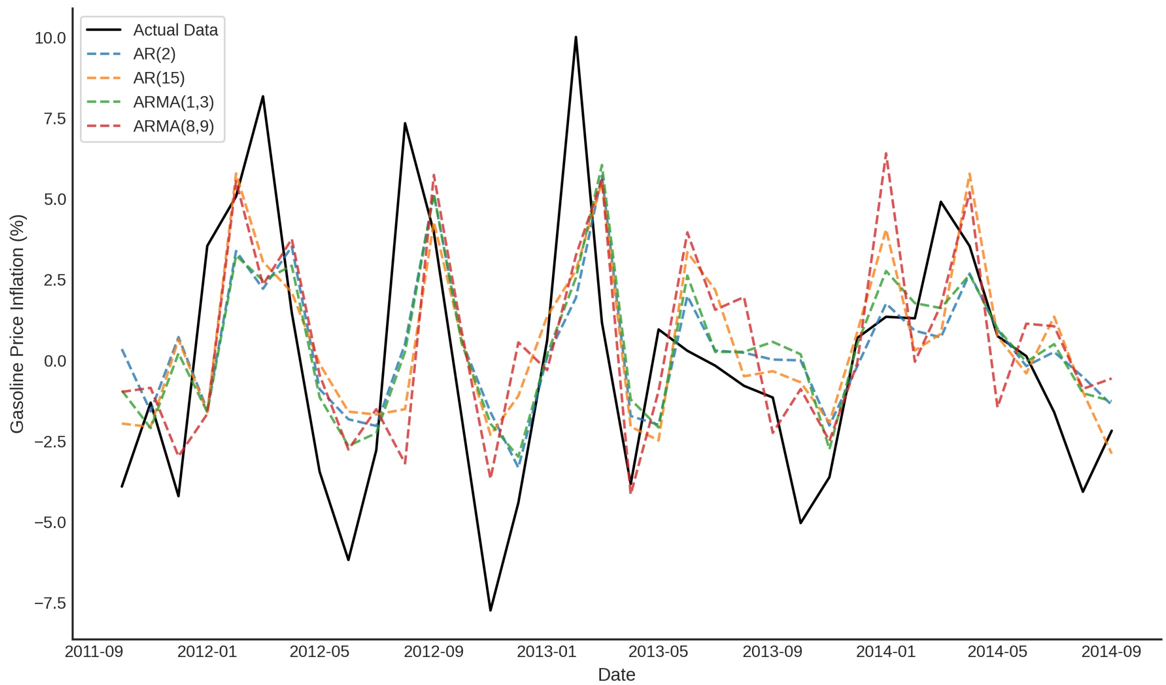


Figure 3: Recursive One-Step-Ahead Forecasts (Hold-out Period: 2011M10–2014M09)

Despite the lack of statistical significance in the DM tests, we select the $ARMA(1,3)$ as our preferred univariate benchmark. This decision is motivated by three factors. First, the $ARMA(1,3)$ yielded the lowest numerical MSPE (10.15), providing the closest point-estimates to the actual series over the hold-out period. Second, following the principle of parsimony, the model effectively captures the short-term dependencies of gasoline inflation with only four parameters, avoiding the overfitting risks inherent in the AIC-selected $AR(15)$. Finally, even though $ARMA(1,3)$ is not statistically proved better than the second best ($AR(2)$) model, it is still statistically more likely that this model is better as it proved statistically better at the 10% level.

3.2 Bivariate Forecasting

Now, we include an additional variable - the Brent crude oil inflation rate - in our models. This yields AR(p,r) and ARMA(p,q,r) models, where r indicates the number of lags, with which the covariate comes into the prediction.

As Table 3 reveals, the AIC favours an ARX(9,17), while the BIC favours an ARX(9,4). When only considering ARIMAX, AIC chooses ARIMAX(10,0,7) and BIC chooses ARIMAX(2,5,1). The differences in the choices of the AIC, although it essentially chooses an ARX for both selections, is due to the values of p and r being capped at 10 for the ARX-model and at 20 for the ARIMAX [1].

In the next step, we compute the rolling MSPE results, analogous to how we did it for the univariate case. The lowest MSPE is assigned to an ARIMAX$(2,5,3)$. We now conduct a Diebold-Mariano Test, in which we compare the four models picked by AIC and BIC to our univariate benchmark,

[1] Although the models are called ARIMAX here, we set the order of integration to 0 for any of these models

the ARMA(1,3)-model. As Table 6 shows, including the Brent crude oil inflation rate in the model does not significantly enhance forecasting performance compared to the benchmark.

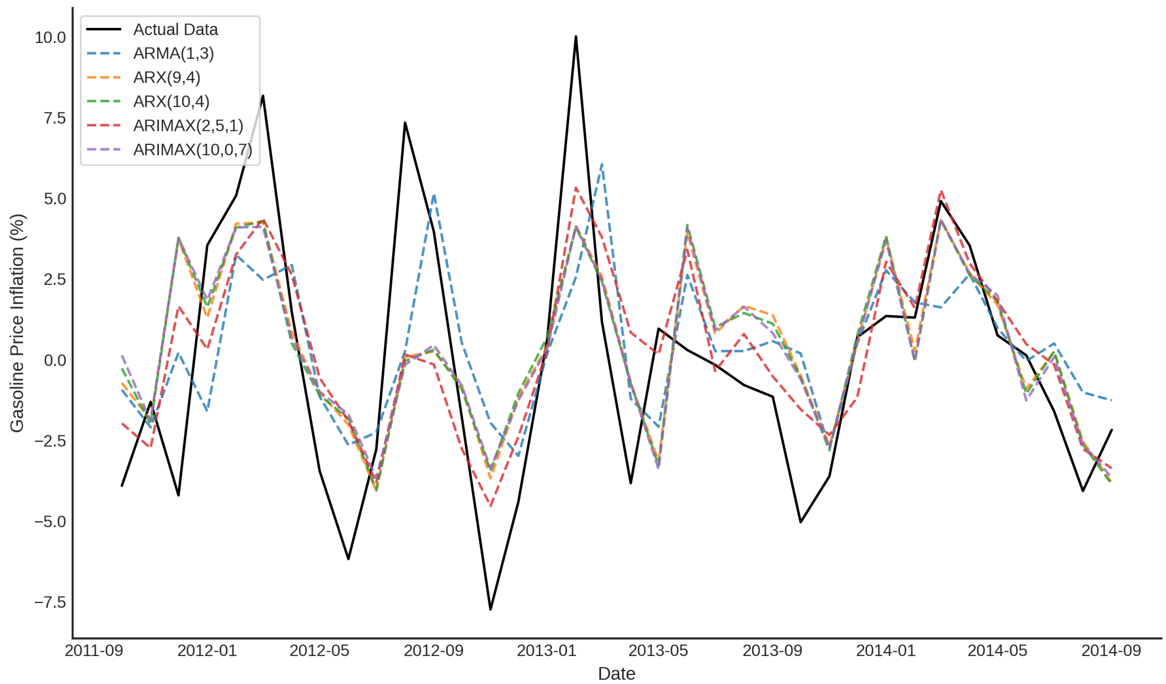


Figure 4: Recursive One-Step-Ahead Forecasts (Hold-out Period: 2011M10–2014M09)

Table 5: Diebold-Mariano Test for candidate univariate models (Benchmark: ARMA(1, 3))

Comparison	DM Statistic	p-value
vs. $AR(2)$	-1.71	0.088
vs. $AR(15)$	-1.29	0.197
vs. $ARMA(8,9)$	-0.99	0.321

4 Results

Of all models that we tested, the lowest MSPE is assigned to an ARIMAX$(2,5,1)$. Although it is not significantly better at forecasting than the ARMA$(1,3)$ according to the Diebold-Mariano Test (p-value: 0.119), it is still the best-performing in terms of statistics. Therefore, we choose the ARIMAX$(2,5,1)$ for our forecast. That said, we are also conducting the forecast with the ARMA(1,3) to discern whether the forecasting results of both models are similar.

Our main forecast, based on the ARIMA(2,5,1), predicts an inflation rate for the US retail price of gasoline for October 2014 of -2.67 percentage points. In contrast, the ARMA(1,3) forecasts an inflation rate of 0.77 percentage points. The differences between these two forecasts highlight how notoriously difficult gasoline price forecasts are and how important specification choices are.

Table 6: Diebold-Mariano Test for candidate bivariate models (Benchmark: ARMA$(1,3)$)

Comparison	DM Statistic	p-value
vs. $ARIMAX(2,5,1)$	1.56	0.119
vs. $ARX(9,4)$	0.32	0.749
vs. $ARX(9,17)$	0.26	0.796
vs. $ARIMAX(10,0,7)$	0.14	0.893

5 Conclusion

Gasoline prices play an important role in many areas of the economy, far beyond simple consumption choices. That said, the inflation rate of gasoline is highly volatile and notoriously difficult to predict. Nonetheless, this paper attempted to forecast the inflation rate of the US retail price of gasoline for October 2014, based on gasoline inflation data from July 1986 to September 2014. Employing that data, we assessed the forecasting performance of various time series models.

We decided that an ARIMA(2,5,1) including a one-lagged time series on the inflation rate of Brent crude oil as an additional variable is the best choice for the forecast of the gasoline inflation rate. Our model predicted an inflation rate of -2.67 percentage points for October 2014.

References

Baumeister, Christiane et al. (Mar. 2017). "Inside the Crystal Ball: New Approaches to Predicting the Gasoline Price at the Pump". In: *Journal of Applied Econometrics* 32.2, pp. 275–295. DOI: 10.1002/jae.2510. URL: https://ideas.repec.org/a/wly/japmet/v32y2017i2p275-295.html.

Bernard, Corentin (Dec. 2024). "How Much Do People Spend on Gas Each Month?" In: URL: https://www.jdpower.com/cars/shopping-guides/how-much-do-people-spend-on-gas-each-month.

Vaidyanathan, Shruti (May 2021). "Analysis: Gasoline Costs Consume Nearly 20% of Some Household Budgets". In: URL: https://www.aceee.org/blog-post/2021/05/analysis-gasoline-costs-consume-nearly-20-some-household-budgets.

A Statement of generative AI usage

During the completion of this assignment, ChatGPT and Gemini assisted us with writing Python code and debugging LaTeXissues.