# Time Series Modeling and Forecasting of US Consumer Price Index Using WTI Crude Oil Price

**Yuxi Liu[1,a,*]**

[1]*College of Arts and Science, New York University, New York, United States*
*a. yl7777@nyu.edu*
*\*corresponding author*

*Abstract:* This study aims to forecast U.S. inflation by using crude oil prices as a key variable. WTI crude oil price and Consumer Price Index Year-over-Year (CPI YoY) data from 2000 to 2022 are extracted to construct a time series model. The empirical analysis relies on Autoregressive Integrated Moving Average (ARIMA) model and regression to formulate forecasts based on historical patterns and error dynamics. The study finds out that compared to the automatically generated ARIMA model, using fitted values of the WTI time series model to predict CPI YoY through a multi-variable model is better. According to the prediction, the CPI YoY forecast for 2023 reveals a decreasing tendency in 2023, implying a slower rate of increase in U.S. inflation. The result provides insights into economic conditions, helps decision-making, and mitigates the potential risks associated with price fluctuations. Overall, the research contributes to a deeper understanding of the dynamics between energy markets and inflationary pressures in the United States.

*Keywords:* U.S. Inflation, CPI forecast, WTI Crude Oil Price, time series analysis, ARIMA model

## 1. Introduction

### 1.1. Research Background and Significance

Crude oil price plays a vital role in oil-importing countries like the United States, particularly in the transportation and manufacturing sectors [1]. For the U.S., the breakeven inflation rate is closely linked to oil prices for most of the period. An oil price increase can lead to inflation and higher interest rates [2]. In specific, the causality-in-variance between WTI crude oil price and inflation is significant in both directions, implying the predictive power of crude oil price for CPI [3]. Therefore, it is plausible to predict U.S. inflation based on crude oil prices.

Inflation prediction has been a major interest because it sheds light on decision makings. Given the significant link between unemployment and inflation, governments, central banks, and policymakers rely on the prediction to enact monetary and fiscal policies, set interest rates, and manage stability [4]. In the context of the Russia-Ukraine conflict, according to The Federal Reserve, the rising commodity prices and oil prices imposed a downward pressure on global economic activity and upward pressure on inflation. In response to the geopolitical risk, the global GDP level has reduced by approximately 1.5 percent and global inflation has increased by 1.3 percentage points in

2022 [5]. For consumers, inflation eroded their purchasing power. Together, the lower consumer sentiment and tighter financial conditions harmed the good-producing industries.

Forecasting U.S. inflation holds great significance for various stakeholders especially for the post-pandemic period. With an accurate prediction, policymakers can implement appropriate monetary policies, adjust interest rates, and maintain economic stability. This study realizes the U.S. inflation forecast by using WTI crude oil price as a key indicator.

## 1.2. Literature Review

A considerable amount of research is done to study inflation. Zakaria explains that alterations in oil prices can impact inflation via direct as well as indirect mechanisms. The direct effect is reflected on the demand side. If oil has a significant share in the consumer basket, increased oil prices exert upward pressure on inflation. The indirect effect is reflected on the supply side. Elevated oil prices will cause higher producer prices, which contribute to inflation. Additionally, high oil prices discourage purchasing power, which potentially leads to higher wages and cost-push inflation [6].

Mensi investigates the nonlinear relationships between two variables in twelve major global economies through a nonparametric quantile causality approach. The results highlight the importance of macroeconomic factors in driving and amplifying the parallel movements between oil prices and inflation [7]. Salisu augments the Phillips curve through both symmetric and asymmetric shifts in oil price and employs Westerlund and Narayan (WN) estimator to build a multi-predictor model which captures endogeneity, persistence, and heteroscedasticity effects [8]. Chou studies the inflation in Taiwan and finds that international oil prices exert an enduring influence on inflation, displaying a pronounced long-term pass-through effect [9]. Katircioglu examines the correlation between oil prices and OECD countries' macroeconomic indicators using an econometric method. The findings indicate a long-term relationship between oil prices, GDP, inflation, and unemployment. Specifically, the results verify that oil prices have negative impacts on these major economic measures in OECD countries [10]. The discussions about considering macroeconomic factors are essential to construe, model, and forecast inflation.

Previous empirical research has predominantly focused on investigating the predictability of inflation through oil prices. However, few studies take advantage of their correlation to make predictions. This research fills the gap in existing research by using the time series ARIMA model to predict U.S. inflation using WTI crude oil prices.

## 1.3. Research Contents and Framework

To examine and use the predictive power of WTI crude oil price on U.S. inflation, the research is structured into two sections. The first section studies WTI and CPI YoY time series independently. It aims to fit WTI and CPI YoY time series into ARIMA models.

In the second section, the study uses the WTI time series as a key indicator to build regression models for the CPI YoY time series. Two models are considered in this part: the automatic ARIMA model and the dynamic regression model. While the automatic ARIMA model uses the original WTI time series as an independent variable to generate a model for CPI YoY, the dynamic regression model uses fitted values given by the WTI time series ARIMA model in the first part.

At last, the study forecasts CPI YoY values in 2023. It compares the prediction values to real values from January to June and determines the reliability of the forecast. As a result, all real values fall into the prediction intervals. The predicted CPI YoY values from July to December give insights into U.S. inflation that benefits policymakers, investors, and consumers.

## 2.    Methodology

### 2.1.    Data Selection

This research focuses on studying the average price of West Texas Intermediate (WTI) crude oil and the year-over-year percentage change in the Consumer Price Index (CPI YoY) for all urban consumers in the U.S. city average. Instead of using CPI, CPI YoY can better capture the change in volatility of inflation over the one-year period. The CPI YoY formula is:

$$CPI\ YoY = \left(\left(\frac{CPI_t}{CPI_{t-12}}\right) - 1\right) \times 100 \tag{1}$$

$CPI_t$ represents the CPI for month t, and $CPI_{t-12}$ represents the CPI for the same month in the previous year. Given the CPI YoY prediction, the CPI can be calculated by the same formula.

### 2.2.    Data Preprocessing

To model time series data, the data structure needs to be examined. Stationarity is the underlying assumption before fitting it into an ARIMA model. If the series is non-stationary, differencing can be applied to transform it into a stationary series, resulting in constant mean and variance [11]. The time series stationarity in this study is assessed by conducting the Augmented Dickey-Fuller (ADF) test.

### 2.3.    ARIMA Models

The Autoregressive Integrated Moving Average (ARIMA) model presents a simulation and forecasting tool that offers several key advantages, including its ability to effectively handle nonlinear data patterns, its utilization of advanced statistical processes, and its practicality in risk analysis [11].

### 2.3.1. ARIMA Model for Single Time Series Variable

The ARIMA model, specified as ARIMA (p, d, q), is composed of three components: the autoregressive order (p), the degree of differencing applied to the time series (d), and the moving average order (q). For example, the first difference WTI series is of the form:

$$\Delta WTI_t = WTI_t - WTI_{t-1} \tag{2}$$

Its corresponding ARIMA (p,1,q) model is of the form:

$$WTI_t = \beta_0 + \emptyset_1 WTI_{t-1} + \emptyset_2 WTI_{t-2} + \cdots + \emptyset_p WTI_{t-p} + \varepsilon_t + \theta_t \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{3}$$

$WTI_t$ is the WTI series of first-order difference, and $\emptyset$, $\beta$, and $\theta$ are coefficients that will be estimated for the model to ensure stationarity.

### 2.3.2. ARIMA Model for Multi-Time Series Variable

In the context of multiple variables, the ARIMA model processes the dependent variable by analyzing its past values. The autoregressive (AR) term in the model accounts for the connection between the current value of the dependent variable and its past values, using lagged values of the dependent variable as predictors. The integrated (I) component reflects the differencing operation performed on the dependent variable to achieve stationarity. Lastly, the moving average (MA) component considers the relationship between the current value of the dependent variable and its historical forecast errors.

### 2.3.3. Dynamic Regression Model with ARIMA errors

The dynamic regression model combines the concepts of ARIMA and regression analysis. It allows for predicting the future values of the dependent variable considering the time-dependent relationship with the independent variable. The regression model is of the form:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \cdots + \beta_k x_{k,t} + \eta_t, \tag{4}$$

$$(1 - \emptyset_1 B)(1 - B)\eta_t = (1 + \theta_1)\varepsilon_t, \tag{5}$$

$y_t$ is modeled as a linear combination of $k$ predictor variables with the assumption that the error series $\eta_t$ follows an ARIMA (1,1,0) structure, where $\varepsilon_t$ is a white noise series [12].

## 3. Results

### 3.1. Descriptive Statistics

This study uses data sourced from the Federal Reserve Economic Data (FRED) website (https://fred.stlouisfed.org/). For research purposes, the average monthly WTI crude oil price and monthly CPI YoY from January 2000 to December 2022 are used to forecast CPI YoY by using R software. Table 1 shows the statistics of WTI, CPI YoY, and CPI.

Table 1: Summary statistics.

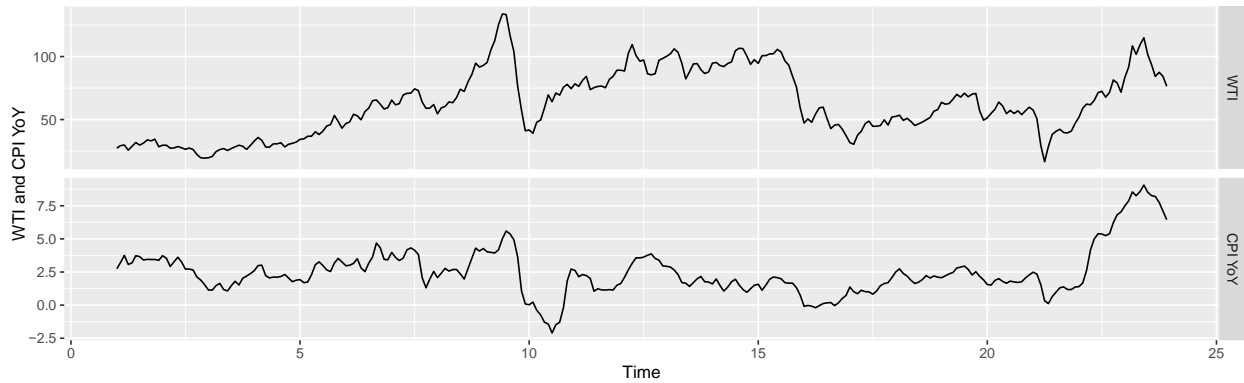| Statistic | WTI | CPI YoY | CPI |
|---|---|---|---|
| No. of observations | 276 | 276 | 276 |
| Minimum | 16.55 | -2.10 | 168.8 |
| 1st Quartile | 41.08 | 1.51 | 197.4 |
| Median | 59.27 | 2.16 | 225.8 |
| Mean | 62.49 | 2.50 | 223.0 |
| 3rd Quartile | 84.25 | 3.26 | 244.0 |
| Maximum | 133.88 | 9.06 | 298.0 |
| Range | 117.33 | 7.55 | 129.2 |
| Standard deviation | 26.05 | 1.80 | 31.40 |
| Skewness | 0.3298 | 1.1723 | 0.1945 |
| Kurtosis | 2.2200 | 5.6077 | 2.4005 |

Figure 1: Graph of WTI and CPI YoY time series from 2000-2022.

## 3.2. Data Transformation

To fit the WTI and CPI YoY time series into an ARIMA model, the study first examines stationarity using the ADF test. Table 2 lists the ADF test results for the two time series before and after the first difference. Since the p-value for both time series is less than the critical value after taking the first difference, WTI and CPI YoY first difference time series are stationary.

Table 2: ADF test for WTI and CPI YoY before and after first difference.

| Variable | Dickey-Fuller | Lag Order | P-value | Result |
|---|---|---|---|---|
| WTI | -2.3266 | 6 | 0.4384 | Non-Stationary |
| CPI YoY | -2.954 | 6 | 0.1741 | Non-Stationary |
| WTI first diff. | -6.7148 | 6 | 0.01 | Stationary |
| CPI YoY first diff. | -5.779 | 6 | 0.01 | Stationary |

## 3.3. Fitting into the ARIMA model

After finding stationary time series for WTI and CPI YoY, it is safe to identify ARIMA models and estimate the parameters for two series.

### 3.3.1. Single variable series ARIMA model

Table 3 shows the results obtained from automatically fitted ARIMA models for two series using R. Table 4 shows the outcomes from the Ljung-box test of residuals.

Table 3: Results from ARIMA models for WTI and CPI YoY time series.

| Data | Model | Log-likelihood | AIC | AICc | BIC |
|---|---|---|---|---|---|
| WTI | ARIMA(1,1,0) | -856.97 | 1717.95 | 1717.99 | 1725.18 |
| CPI | ARIMA(2,1,2) | -102.67 | 219.34 | 219.76 | 244.66 |
| YoY | (1,0,0)[12] with drift | | | | |

Table 4: Residual test for ARIMA (1,1,0) and ARIMA (2,1,2)(1,0,0) [12] with drift.

| WTI ARIMA Model | CPI YoY ARIMA Model |
|---|---|
| Q* = 25.711, df = 23, p-value = 0.3147 | Q* = 62.083, df = 19, p-value = 1.804e-06 |

<div align="center">Table 4:(continued).</div>

| Degrees of freedom: 1. Lag order: 24 | Degrees of freedom: 6. Lag order: 24 |
|---|---|

The p-value for ARIMA (1,1,0) is greater than 0.05, and the residuals exhibit minimal autocorrelation, indicating a resemblance to white noise. This implies that the model fits well with the WTI time series. Based on the estimation of ARIMA (1,1,0) parameters, the model can be expressed as follows:

$$\Delta WTI_t = 0.3494\Delta WTI_{t-1} + \varepsilon_t \tag{6}$$

However, for ARIMA (2,1,2)(1,0,0) [12] with drift, the p-value falls below the critical value, indicating that the residuals exhibit significant autocorrelation. The model is inappropriate for the CPI YoY time series.

### 3.3.2. Two variable series ARIMA model

By considering the correlation between WTI crude oil price and inflation, CPI YoY can be better forecasted by combining two time series to better capture the lags. Table 5 shows the results of ARIMA and dynamic regression models automatically fitted by R using two series. While the ARIMA model uses the original data from the two series, the dynamic regression model uses the fitted values from ARIMA (1,1,0) for the WTI series as the independent variable.

<div align="center">Table 5: Results from ARIMA models for CPI YoY by combining with WTI series.</div>

| Method | Model | Log-likelihood | AIC | AICc | BIC |
|---|---|---|---|---|---|
| ARIMA | ARIMA(0,1,2) (1,0,2)[12] errors | -58.58 | 131.16 | 131.58 | 156.48 |
| Dynamic Regression | ARIMA(3,1,2) (2,0,0)[12] errors | -76.62 | 173.25 | 174.08 | 209.41 |

<div align="center">Table 6: Residual test for ARIMA (0,1,2)(1,0,2) [12] errors and ARIMA (3,1,2)(2,0,0) [12] errors.</div>

| Automatically Fitted ARIMA Model | Dynamic Regression ARIMA Model |
|---|---|
| Q* = 19.417, df = 19, p-value = 0.4304 | Q* = 24.329, df = 17, p-value = 0.1108 |
| Degrees of freedom: 5. Lag order: 24 | Degrees of freedom: 7. Lag order: 24 |

The information criteria shows ARIMA (0,1,2)(1,0,2) [12] errors is a better fit compared to ARIMA (3,1,2)(2,0,0) [12] errors. From Table 5, the automatically fitted ARIMA model has bigger log-likelihood, and smaller AIC, AICc, and BIC. Moreover, from Table 6, the residual result shows that the p-value is bigger, suggesting that the residual distribution is not significantly different from white noise. The estimated parameters for ARIMA (0,1,2)(1,0,2) [12] errors are listed in Table 7. The graph of fitted values for the CPI YoY forecast ideal model and real values are shown in Figure 2.

<div align="center">Table 7: Estimated parameters for ARIMA (0,1,2)(1,0,2) [12] errors.</div>

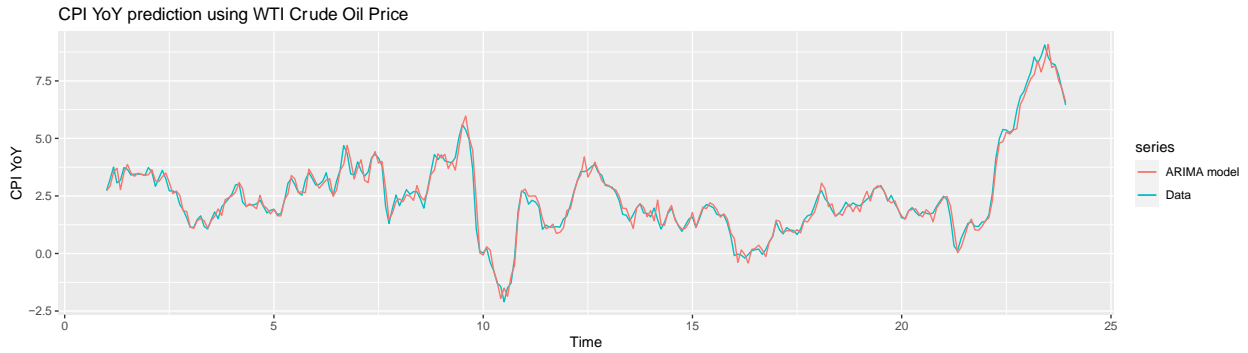| | ma1 | ma2 | sar1 | sma1 | sma2 | xreg |
|---|---|---|---|---|---|---|
| Coefficients | 0.5722 | 0.1142 | 0.4106 | -1.3457 | 0.4477 | 0.0068 |
| s.e. | 0.0617 | 0.0592 | 0.3763 | 0.3618 | 0.3122 | 0.0028 |

Figure 2: Real CPI YoY values and fitted values from ARIMA (0,1,2)(1,0,2)[12] errors.

## 3.4. Forecast

Finally, the optimal model for CPI YoY can predict future values using the known monthly average WTI crude oil price, updated until June 2023. Table 8 shows the monthly forecast for CPI YoY along with the real CPI YoY values recorded from January 2023 to June 2023. Figure 3 shows the CPI YoY prediction plot, which indicates a decreasing trend in 2023.

Table 8: Future forecast of CPI YoY from ARIMA (0,1,2)(1,0,2) [12] errors model.

| Date | Real Value | Point Forecast | Lo 80 | Hi 80 | Lo 95 | Hi 95 |
|------|-----------|----------------|-------|-------|-------|-------|
| Jan 2023 | 6.41015 | 5.94530 | 5.56865 | 6.32194 | 5.36926 | 6.52133 |
| Feb 2023 | 6.03561 | 5.49134 | 4.78955 | 6.19314 | 4.41804 | 6.56465 |
| Mar 2023 | 4.98497 | 4.59702 | 3.65044 | 5.54357 | 3.14936 | 6.04465 |
| Apr 2023 | 4.93032 | 4.27052 | 3.13059 | 5.41045 | 2.52714 | 6.01390 |
| May 2023 | 4.04761 | 3.61442 | 2.30947 | 4.91938 | 1.61866 | 5.61019 |
| Jun 2023 | 2.96918 | 2.66705 | 1.21571 | 4.11839 | 0.44742 | 4.88668 |
| Jul 2023 | N/A | 2.79644 | 1.21219 | 4.38069 | 0.37353 | 5.21934 |
| Aug 2023 | N/A | 2.87697 | 1.17013 | 4.58382 | 0.26658 | 5.48737 |
| Sep 2023 | N/A | 2.80829 | 0.98708 | 4.62949 | 0.02300 | 5.59358 |
| Oct 2023 | N/A | 2.61378 | 0.68499 | 4.54258 | -0.33605 | 5.56362 |
| Nov 2023 | N/A | 2.57210 | 0.54141 | 4.60279 | -0.53358 | 5.67778 |
| Dec 2023 | N/A | 2.72807 | 0.60035 | 4.85578 | -0.52600 | 5.98213 |
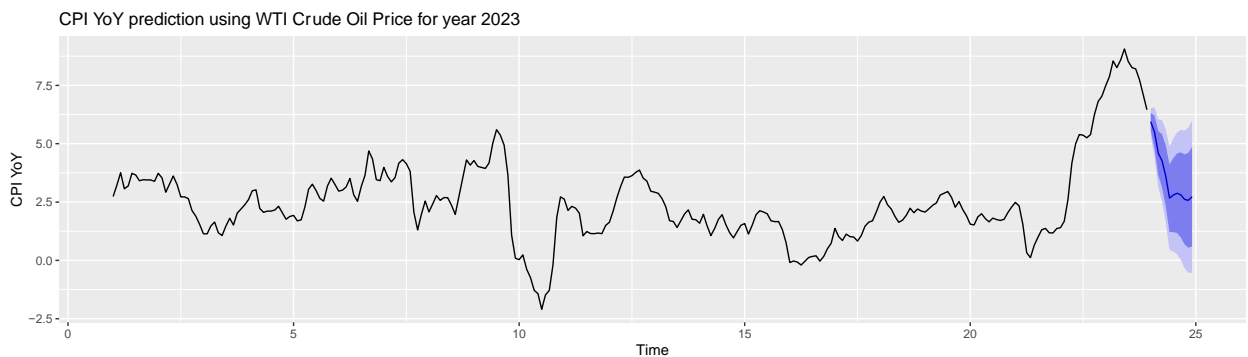


Figure 3: Forecast CPI YoY values from ARIMA (0,1,2)(1,0,2) [12] errors.

# 4. Discussion

Based on the results, the ARIMA (0,1,2)(1,0,2) [12] errors model is powerful in capturing the inner relationship between WTI crude oil price and CPI YoY. The oil price coefficient is positive and has small standard errors, indicating a positive relationship between the two variables. The positive coefficient implies a direct impact or correlation between the exogenous variable and the dependent variable.

The study also finds out that the automatic ARIMA model performs better compared to the dynamic regression model in fitting CPI YoY values. The comparative advantage can be credited to the model's simplicity and ability to capture time series patterns. By employing algorithmic techniques to determine the most suitable values for the ARIMA parameters (p, d, q), the model automatically detects and incorporates seasonality in the data. This is particularly useful when analyzing time series data with periodic patterns. Moreover, it demonstrates the flexibility in exogenous variables. The automatic ARIMA model enables the incorporation of additional factors that improve forecasting accuracy and capture additional explanatory power.

# 5. Conclusion

This study uses monthly data that encompasses a period starting from January 2000 to December 2022. The result verifies the predictive power of WTI crude oil price on U.S. inflation. The ARIMA model and regression with the ARIMA errors model are engaged to forecast CPI YoY. The ideal model is selected based on how well it explains the stochastic variance in the data. By using the inner correlation between WTI crude oil price and inflation, ARIMA (0,1,2)(1,0,2) [12] errors is selected as an ideal model. Compared to the automatically generated ARIMA model that predicts based on time series patterns in CPI YoY historical data, the ideal model that combines WTI crude oil price performs better in accuracy.

Using ARIMA (0,1,2)(1,0,2) [12] errors, the study gives the inflation forecast result for 2023. The model prediction for CPI YoY value from January 2023 to June 2023 demonstrates a remarkable predictive effect. The real values fall into prediction intervals and the CPI YoY point forecast values show the same deflationary tendency. It provides a reference and basis to satisfy the needs of government, investors, and consumers to formulate policies and make financial decisions, which helps the macroeconomic operation and contributes to a good economic order.

However, the proposed model only considers WTI crude oil price as a single endogenous variable. While using WTI crude oil price as a predictor demonstrates a certain level of predictive power, exploring alternative models such as vector autoregression (VAR) or machine learning algorithms may provide more accurate and robust inflation forecasts. Future research can improve the model and forecast accuracy by incorporating other relevant factors, such as interest rates, employment data, and global economic indicators.

# References

[1] Nademi, A., & Nademi, Y. (2018). Forecasting crude oil prices by a semiparametric Markov switching model: OPEC, WTI, and Brent cases. Energy Economics, 74, 757–766. https://doi.org/10.1016/j.eneco.2018.06.020
[2] Alturki, S., & Olson, E. (2022). Oil sentiment and the U.S. inflation premium. Energy Economics, 114, 106317. https://doi.org/10.1016/j.eneco.2022.106317
[3] Mensi, W., Rehman, M. U., Hammoudeh, S., Vo, X. V., & Kim, W. J. (2023). How macroeconomic factors drive the linkages between inflation and oil markets in global economies? A multiscale analysis. International Economics, 173, 212–232. https://doi.org/10.1016/j.inteco.2022.12.003
[4] Jean-Louis, C. O. M. B. E. S., & Lesuisse, P. (2022). Inflation and unemployment, new insights during the EMU accession. International Economics. https://doi.org/10.1016/j.inteco.2022.09.004

[5]  Caldara, D., Conlisk, S., Iacoviello, M., & Penn, M. (2022). The Effect of the War in Ukraine on Global Activity and Inflation. FEDS Notes, 2022(3141). https://doi.org/10.17016/2380-7172.3141

[6]  Zakaria, M., Khiam, S., & Mahmood, H. (2021). Influence of oil prices on inflation in South Asia: Some new evidence. Resources Policy, 71, 102014. https://doi.org/10.1016/j.resourpol.2021.102014

[7]  Wang, Y. S., & Chueh, Y. L. (2013). Dynamic transmission effects between the interest rate, the US dollar, and gold and crude oil prices. Economic Modelling, 30, 792–798. https://doi.org/10.1016/j.econmod.2012.09.052

[8]  Salisu, A. A., & Isah, K. O. (2018). Predicting US inflation: Evidence from a new approach. Economic Modelling, 71, 134–158. https://doi.org/10.1016/j.econmod.2017.12.008

[9]  Chou, K. W., & Tseng, Y. H. (2011). Pass-through of oil prices to CPI inflation in Taiwan. International Research Journal of Finance and Economics, 69(69), 73-83.

[10]  Katircioglu, S. T., Sertoglu, K., Candemir, M., & Mercan, M. (2015). Oil price movements and macroeconomic performance: Evidence from twenty-six OECD countries. Renewable and Sustainable Energy Reviews, 44, 257–270. https://doi.org/10.1016/j.rser.2014.12.016

[11]  Lu, Y., & AbouRizk, S. M. (2009). Automated Box–Jenkins forecasting modelling. Automation in Construction, 18(5), 547–558. https://doi.org/10.1016/j.autcon.2008.11.007

[12]  Rob J Hyndman and George Athanasopoulos. (2021). Chapter 10 Dynamic regression models | Forecasting: Principles and Practice (3rd ed). In otexts.com. Otexts. https://otexts.com/fpp3/dynamic.html