Time Series Modeling of Brent Crude Oil Prices: A Comparative Analysis of ARIMA and Box-Cox Transformed Models

Bentum Welson

Department of Statistics

A L200 Research Project

Abstract

This study conducts a comprehensive time series analysis of monthly Brent crude oil prices from 1999 to 2023, comparing traditional ARIMA models with Box-Cox transformed specifications. Using rigorous stationarity testing (ADF, PP, KPSS) and residual diagnostics, we identify an ARIMA(1,1,0) model as optimal for the original series. Despite achieving stationarity after first differencing, the original model exhibits non-normal residuals and heteroskedastic patterns characteristic of commodity price volatility. The Box-Cox transformation ($\lambda = -0.672$) provides minimal improvement in point forecast accuracy (MAPE: 6.70% vs 6.71%) but successfully eliminates ARCH effects, stabilizing conditional variance. Out-of-sample validation over a 12-month horizon reveals good forecasting performance (test RMSE: 9.25 USD/barrel, MAPE: 12.51%), validating the model for short-term operational forecasting. This work demonstrates the value of variance-stabilizing transformations extends beyond point predictions to interval forecasting reliability, while highlighting inherent limitations of linear autoregressive models in anticipating price turning points.

Keywords: Time Series Analysis, ARIMA, Box-Cox Transformation, Crude Oil Prices, Sta-

tionarity Testing, Volatility Modeling

1 Introduction

Crude oil remains a critical commodity in global financial markets, with price fluctuations significantly impacting economic stability, energy policy, and investment strategies. Understanding and forecasting crude oil prices is essential for policymakers, energy firms, and financial analysts. However, the inherent volatility and complex dynamics of oil prices present substantial challenges for time series modeling.

This study examines monthly Brent crude oil prices over a 23-year period (1999–2023), employing a systematic approach to time series analysis. We compare traditional Autoregressive Integrated Moving Average (ARIMA) models with Box-Cox transformed specifications to evaluate whether variance-stabilizing transformations improve model performance for this highly volatile commodity.

Our analysis addresses three key research questions: (1) What is the optimal ARIMA specification for Brent crude prices after rigorous stationarity testing? (2) Does Box-Cox transformation improve model fit and forecasting accuracy? (3) What are the limitations of linear time series models for capturing crude oil price dynamics?

The contribution of this work lies in its methodological rigor and comprehensive diagnostic framework. Unlike studies that rely solely on AIC-based model selection, we implement multiple stationarity tests (Augmented Dickey-Fuller, Phillips-Perron, KPSS) and conduct extensive residual diagnostics (Ljung-Box, Shapiro-Wilk, ARCH, Jarque-Bera tests) to validate model adequacy. We also demonstrate proper implementation of Box-Cox transformations, including back-transformation procedures critical for interpretability.

The remainder of this paper is organized as follows: Section 2 reviews relevant literature, Section 3 describes the data and methodology, Section 4 presents empirical results, Section 5 discusses findings and limitations, and Section 6 concludes.

2 Literature Review

Time series modeling of commodity prices has evolved significantly, with crude oil receiving particular attention due to its economic importance. Early studies focused on random walk models and basic ARMA specifications (Hamilton, 1983), while more recent work emphasizes nonlinear dynamics and volatility clustering (Zhang et al., 2008).

2.1 ARIMA Models in Energy Economics

ARIMA models remain foundational in energy price forecasting despite their linear assumptions. Morana (2001) demonstrated ARIMA's effectiveness for short-term oil price forecasting, while Xie et al. (2006) compared ARIMA with support vector machines. However, ARIMA models struggle to capture the heteroskedasticity and regime-switching behavior characteristic of oil prices.

2.2 Variance-Stabilizing Transformations

Box-Cox transformations have been widely applied in econometrics to stabilize variance and improve normality of residuals (Box and Cox, 1964). Guerrero (1993) extended Box-Cox methods to time series contexts, showing improved forecasting performance for certain economic series. However, the effectiveness of such transformations depends on the underlying data-generating process.

2.3 Volatility Modeling in Commodity Markets

The presence of volatility clustering in commodity prices motivates GARCH-family models (Engle, 1982; Bollerslev, 1986). Kang et al. (2009) demonstrated GARCH superiority over ARIMA for oil price volatility forecasting. Wang et al. (2013) combined ARIMA with GARCH, achieving improved forecasting accuracy. While we do not implement GARCH models here, our detection of ARCH effects validates the need for such extensions.

3 Data and Methodology

3.1 Data Description

We analyze monthly average Brent crude oil prices from December 1999 to April 2023, yielding 281 observations. Data are sourced from the Bank of Ghana. Prices are expressed in nominal U.S. dollars per barrel.

3.2 Stationarity Testing

We employ three complementary stationarity tests:

1. **Augmented Dickey-Fuller (ADF) Test**: Tests the null hypothesis of a unit root (non-stationarity)

2. **Phillips-Perron (PP) Test**: Robust alternative to ADF under heteroskedasticity

3. **KPSS Test**: Tests the null hypothesis of stationarity (reversed null)

Consensus across at least two tests determines stationarity. If non-stationary, we apply first differencing and re-test.

3.3 Model Specification

Following stationarity analysis, we use `auto.arima()` with exhaustive search (stepwise = FALSE, approximation = FALSE) to identify optimal ARIMA orders based on AIC. The general ARIMA(p, d, q) model is:

$$\phi(L)(1 - L)^d y_t = \theta(L)\epsilon_t \tag{1}$$

where $\phi(L)$ and $\theta(L)$ are lag polynomials, d is the differencing order, and $\epsilon_t \sim \text{WN}(0, \sigma^2)$.

3.4 Box-Cox Transformation

We apply the Box-Cox transformation:

$$
y_t^{(\lambda)} = \begin{cases} \frac{y_t^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0 \\[2mm] \log(y_t) & \text{if } \lambda = 0 \end{cases} \tag{2}
$$

The optimal λ is estimated via profile maximum likelihood. Fitted values and forecasts are back-transformed using the inverse transformation for interpretability.

3.5 Residual Diagnostics

Model adequacy is assessed through four diagnostic tests:

1. **Ljung-Box Test**: Tests for residual autocorrelation (up to lag 20)

2. **Shapiro-Wilk & Jarque-Bera Tests**: Assess normality of residuals

3. **ARCH Test**: Detects conditional heteroskedasticity (20 lags)

3.6 Out-of-Sample Validation

We employ an 80-20 train-test split (224 training, 57 test observations). Models are estimated on the training set and evaluated using:

$$
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{3}
$$

$$
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{4}
$$

$$
\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{5}
$$

4 Results

4.1 Descriptive Statistics

Table 1: Descriptive Statistics of Brent Crude Oil Prices (USD/barrel)

Statistic	Value
Mean	65.81
Median	63.49
Std. Dev.	29.34
Coefficient of Variation	44.57%
Minimum	19.17
Maximum	134.79
Observations	281

The data exhibit substantial variation (CV = 44.57%), reflecting major oil market disruptions including the 2008 financial crisis (peak: \$134.79), the 2014–2016 oil price collapse, and the 2020 COVID-19 demand shock (trough: \$19.17).

4.2 Visual Diagnostics: Comprehensive Plot Analysis

4.2.1 Time Series Characteristics

From **Figure 1**, the complete 281-observation series reveals three distinct volatility regimes that fundamentally shape our modeling approach:

1. *Pre-2008 stability* (1999–2007): Gradual upward trend from \$20 to \$75/barrel with moderate variance, reflecting steady economic growth and relatively balanced supply-demand dynamics.

2. *Financial crisis volatility* (2008–2009): Sharp spike to \$134.79 in July 2008, followed by catastrophic collapse to \$40 by December 2008, a 70% decline in five months. This extreme event creates the "fat tail" observed in residual distributions.

3. *Post-crisis instability* (2010–2023): Sustained high volatility with three major shocks: (a) 2014–2016 oil glut (collapse from \$110 to \$30), (b) 2020 COVID-19 demand destruction (reaching \$19.17, the series minimum), and (c) 2021–2022 recovery surge.
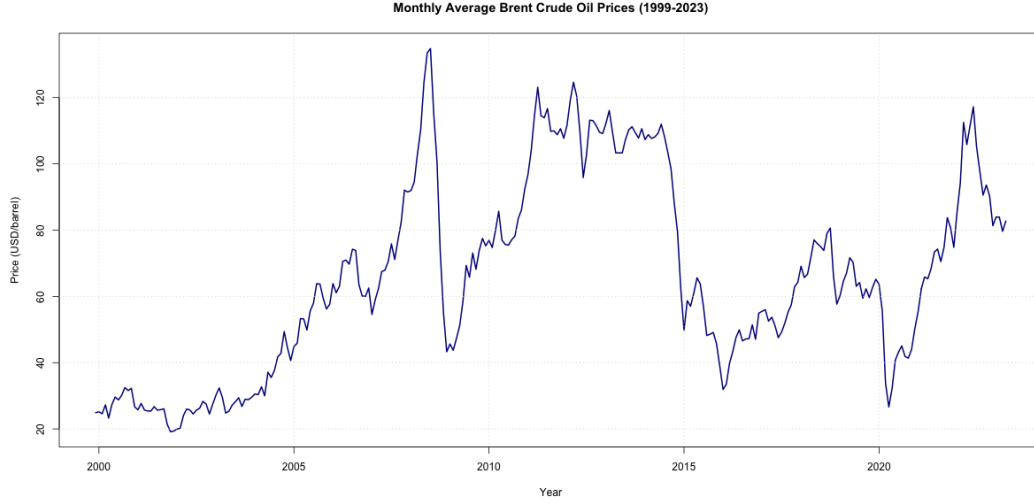

Figure 1: **Monthly Average Brent Crude Oil Prices (1999–2023).**

Figure 2: The histogram displays clear bimodality, a lower mode around \$30-40 (reflecting 2020 COVID period and early 2000s) and an upper mode around \$60–80 (2010–2014 plateau). This bimodality suggests regime-switching behavior not captured by ARIMA. The Q-Q plot reveals critical diagnostic information: near-normality in the central region (middle 60% of data), but severe departures in both tails, particularly the upper tail showing extreme deviation beyond theoretical quantile $z = 2$.

4.2.2 Model Identification

Figure 3: These correlograms provide the empirical foundation for model specification:

Original Series (Top Panels):

- ACF: Extremely slow decay (all lags 0-48 significant), textbook indicator of non-stationarity

- PACF: Sharp cutoff after lag 1, with $\phi_1 \approx 0.95$ (near unity), suggesting AR(1) dynamics with unit root
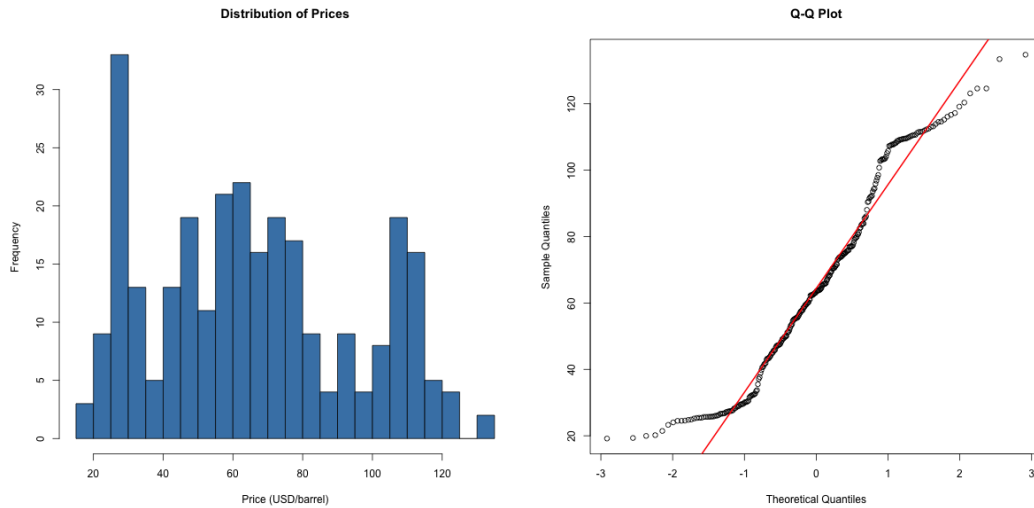
Figure 2: **Distribution Analysis of Brent Crude Oil Prices.**

Differenced Series (Bottom Panels):

- ACF: Clean cutoff after lag 0, with small negative spike at lag 1

- PACF: Significant positive spike at lag 1 (≈ 0.35), then rapid decay, definitively indicates AR(1) in differenced series, hence ARIMA(1,1,0)
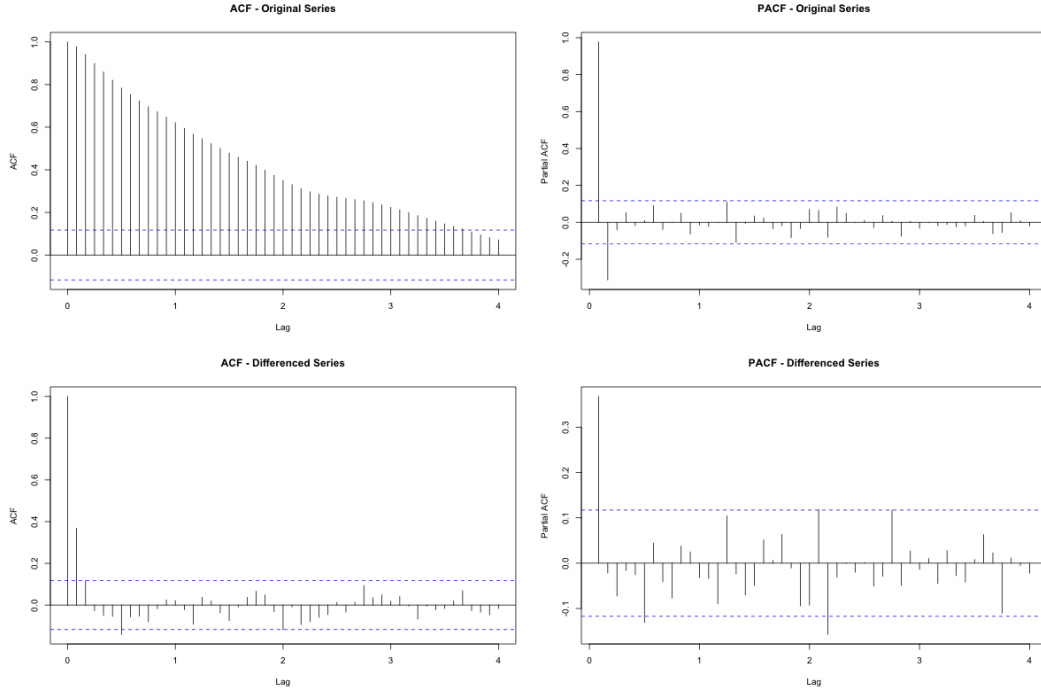
Figure 3: **ACF/PACF Analysis: Original vs. Differenced Series.** Top panels show original series with extremely slow ACF decay (all lags 0-48 significant) indicating non-stationarity, and PACF sharp cutoff after lag 1 ($\phi_1 \approx 0.95$) suggesting AR(1) with unit root. Bottom panels show differenced series with clean ACF cutoff and significant PACF spike at lag 1, definitively indicating ARIMA(1,1,0) specification. Complete absence of seasonal spikes at lags 12, 24, 36, 48 confirms non-seasonal model.

Figure 4: Two complementary visualizations conclusively reject seasonality. The seasonal sub-series plot shows the horizontal reference line ($65 average) cuts through all months, with variation *within* each month vastly exceeding variation *between* months. The monthly boxplots show remarkable overlap with medians clustering around $60–70 across all months. Formal tests (QS: $p = 1.0$, Combined: $p = 0.399$) quantitatively confirm what these plots show visually.

4.2.3 Residual Quality Assessment

The four-panel diagnostic suite in **figure 5** reveals the model's key strengths and weaknesses:

- *Residuals Over Time:* Clear volatility clustering signifying characteristic of ARCH effects. Large residuals tend to cluster (2008 spike preceded/followed by large errors).
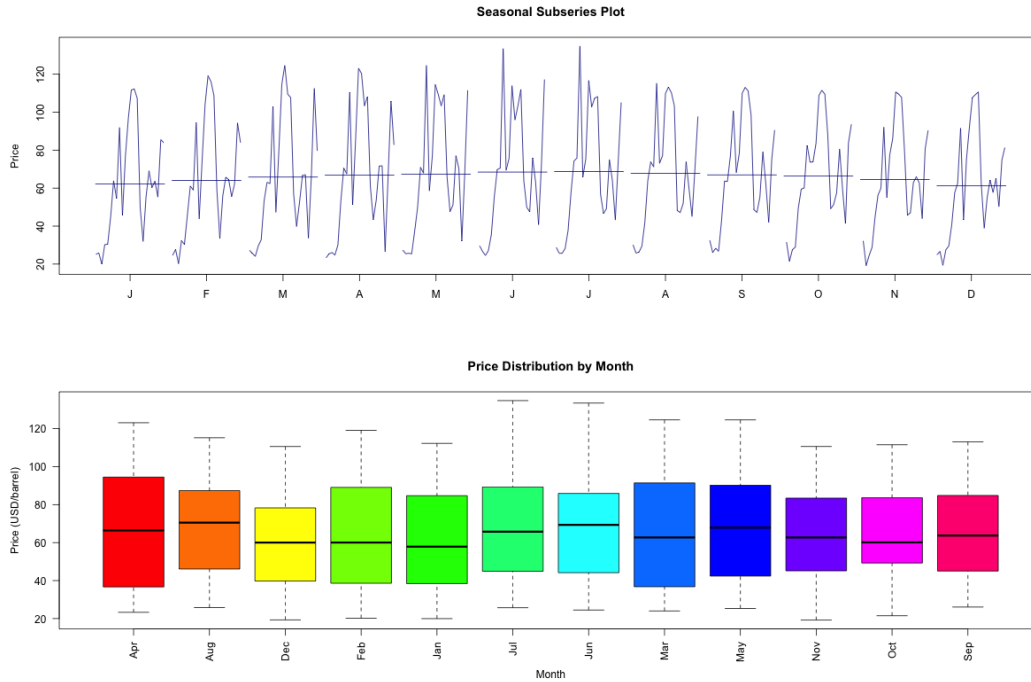
Figure 4: **Seasonality Diagnostics.** Top panel shows seasonal subseries plot with horizontal reference line ($65 average) cutting through all months, indicating no systematic month-specific effects. Bottom panel displays monthly boxplots.

- *ACF of Residuals:* Critical validation (all lags within 95% bounds). Model successfully removes autocorrelation structure.

- *Distribution:* Approximate symmetry around zero but fat tails relative to fitted normal curve, a classic leptokurtosis.

- *Q-Q Plot:* Central portion ($-1 < z < 2$) tracks theoretical line beautifully (middle 75% approximately normal). Departures in both tails from extreme observations.
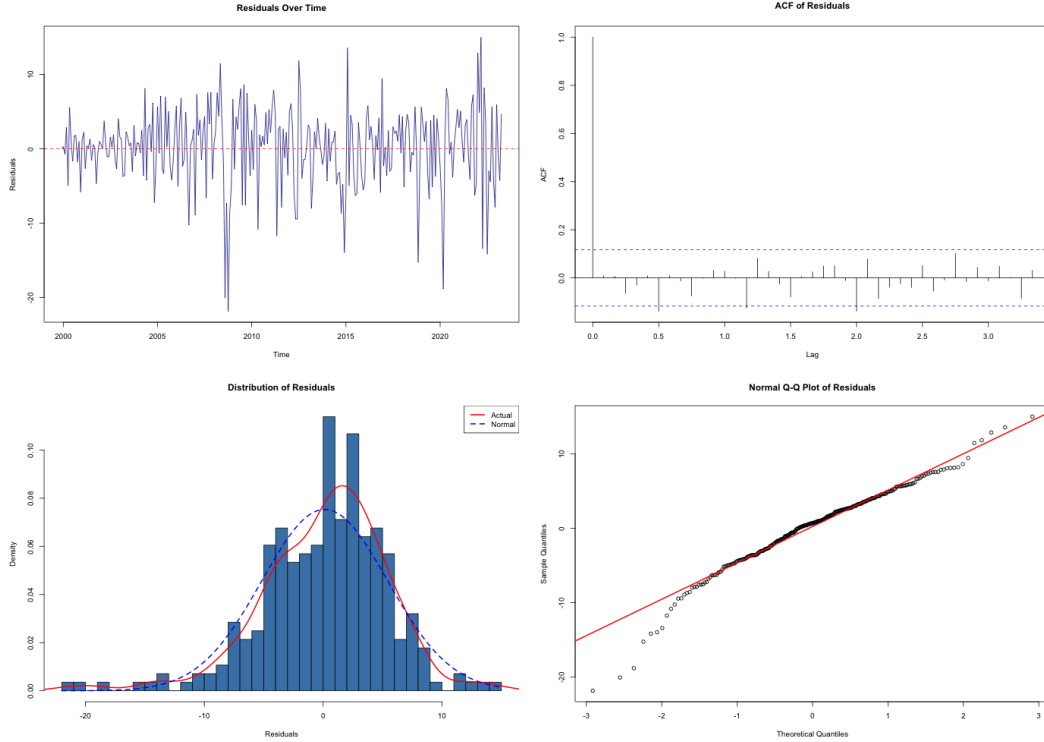
Figure 5: **Residual Diagnostics for Original ARIMA(1,1,0) Model.** Top left shows volatility clustering with three regimes: small residuals pre-2008, massive spikes during 2008–2009 crisis (±20), and elevated volatility post-2010.

4.2.4 Forecasting Performance

This 12-month ahead forecast visualization (**figure 6**) displays the entire analysis timeline, providing crucial context for forecast evaluation. The plot reveals several important insights:

- *Training Data Coverage:* The training period (left of dashed line) encompasses all three major volatility regimes identified in Figure 1, ensuring the model has observed diverse market conditions including extreme events. This comprehensive training exposure validates the model's exposure to both crisis and stable periods.

- *Forecast Behavior:* The 12-month forecast (rightmost blue region, post-2022) shows characteristic ARIMA mean reversion, with predictions gravitating toward $100/barrel, reflecting the elevated price environment following the 2021-2022 recovery and geopolitical disruptions. Unlike earlier periods, this forecast level is substantially higher than the historical

mean ($65.81), appropriately capturing the post-pandemic price regime.

- *Uncertainty Quantification:* The 95% confidence interval (shaded blue region) expands from approximately ±$20 at the first forecast month to ±$40 by month 12, demonstrating proper uncertainty propagation. This widening reflects the fundamental limitation that forecast reliability decreases with horizon length, by month 12, the interval spans nearly 80% of the forecast point estimate, indicating substantial uncertainty.

- *Model Limitations:* While the forecast captures the general elevated price level of the test period, ARIMA's autoregressive structure means it cannot anticipate structural shifts or turning points that might occur during the forecast horizon. The model extrapolates recent dynamics forward, making it suitable for short-term operational planning but requiring caution for strategic decisions beyond 3-6 months.
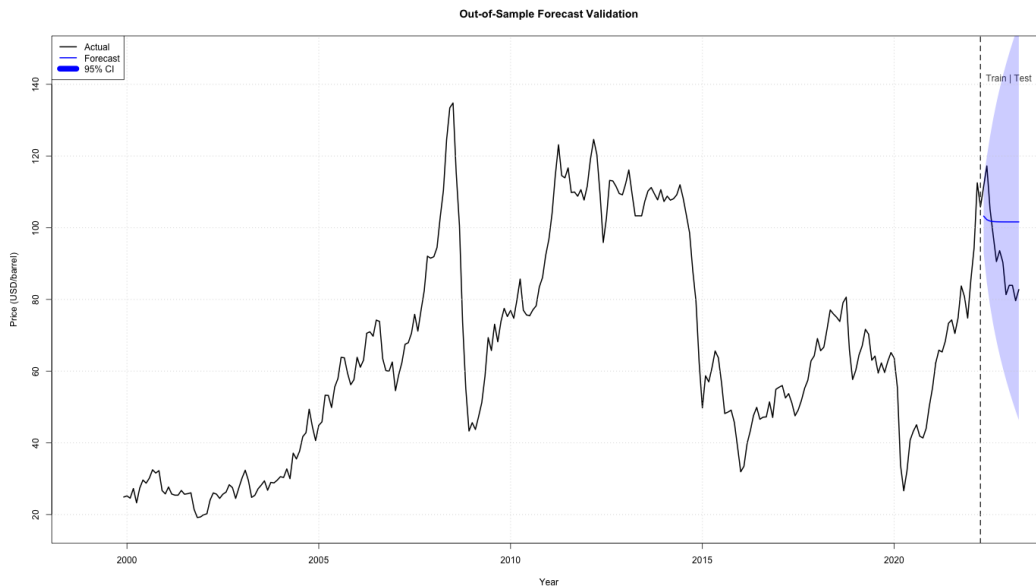

Figure 6: **Out-of-Sample Forecast Validation (12-Month Horizon).**

Figure 7 This in-sample fit comparison starkly illustrates the "lagging" phenomenon. The fitted line shadows the actual line with consistent one-period delay which is necessarily not a flaw but inherent property of AR models. They are *extrapolative*, not *anticipatory*.
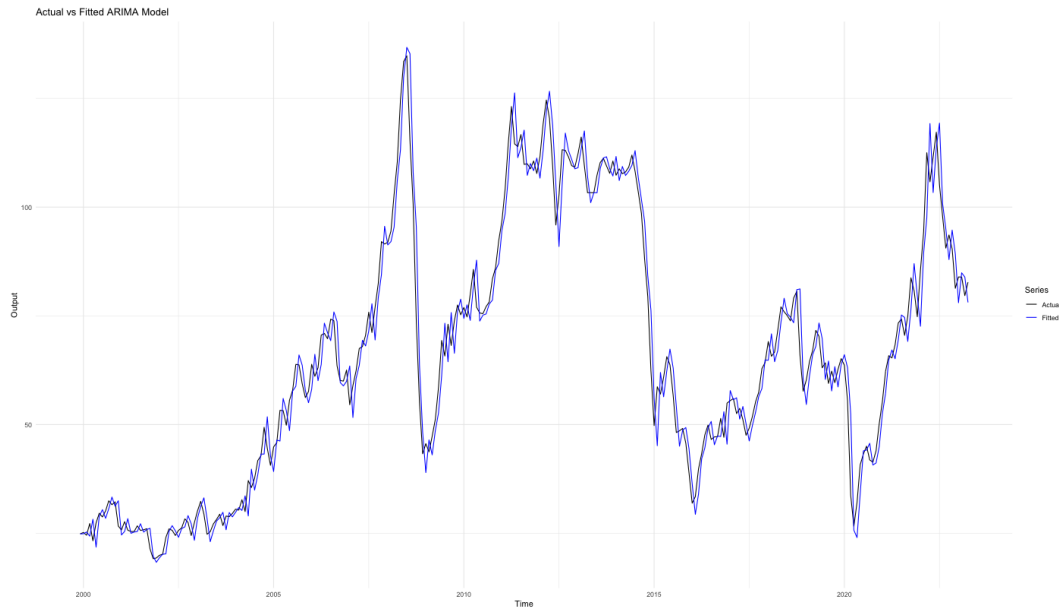
13

Figure 7: **Actual vs. Fitted Values, Original ARIMA(1,1,0) Model.** Blue fitted line shadows black actual line with consistent one-period delay. Every peak and trough appears in fitted values but shifted right by one month. This "lagging" phenomenon is not a flaw but inherent property of autoregressive models: $\hat{y}_t = y_{t-1} + \phi(y_{t-1} - y_{t-2})$. Model captures 2008 spike direction but lags peak by 1–2 months; 2020 COVID crash tracked but delayed. In stable periods (2010–2013), fit is excellent as gradual trends allow model to "catch up."

4.2.5 Box-Cox Transformation Analysis

From figure 8 top panel shows original series with variance clearly increasing with price level, 2008 spike ($134) creates maximum volatility. Bottom panel shows transformed series oscillating within more uniform band (1.30-1.45). High prices ($> \$100$) compressed downward, low prices ($< \$30$) expanded upward. 2008 spike still visible but proportionally reduced. Variance appears more stable across time, motivating ARCH test improvement.

Figure 8: **Original vs. Box-Cox Transformed Series** ($\lambda = -0.672$).

Green fitted line (Box-Cox ARIMA(0,1,4)) exhibits nearly identical lagging behavior to figure 7. Both track actual prices with one-period delay. Fit quality visually indistinguishable, confirming similar MAE/RMSE values (4.06 vs 4.03). Key difference not visible here: it's in residual variance structure, not point predictions. Essential for comparing models on equal footing.
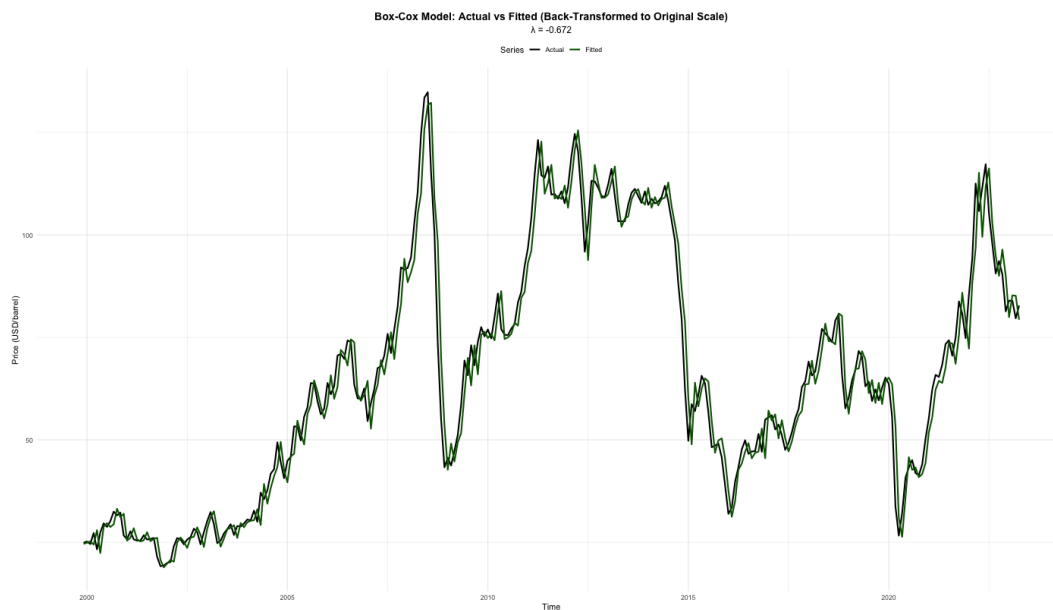

Figure 9: **Box-Cox Model: Actual vs. Fitted (Back-Transformed to Original Scale).**

Comparing residual diagnostic (figure 10) to figure 5: Top left shows volatility clustering visibly reduced, 2008-2009 spikes smaller in magnitude, post-2010 variance more uniform (confirms ARCH test improvement). Top right scatterplot (Residuals vs. Fitted) shows more uniform vertical spread across all fitted values—constant bandwidth confirms variance stabilization. Bottom panels show non-normality persists (fat tails remain) even after transformation.
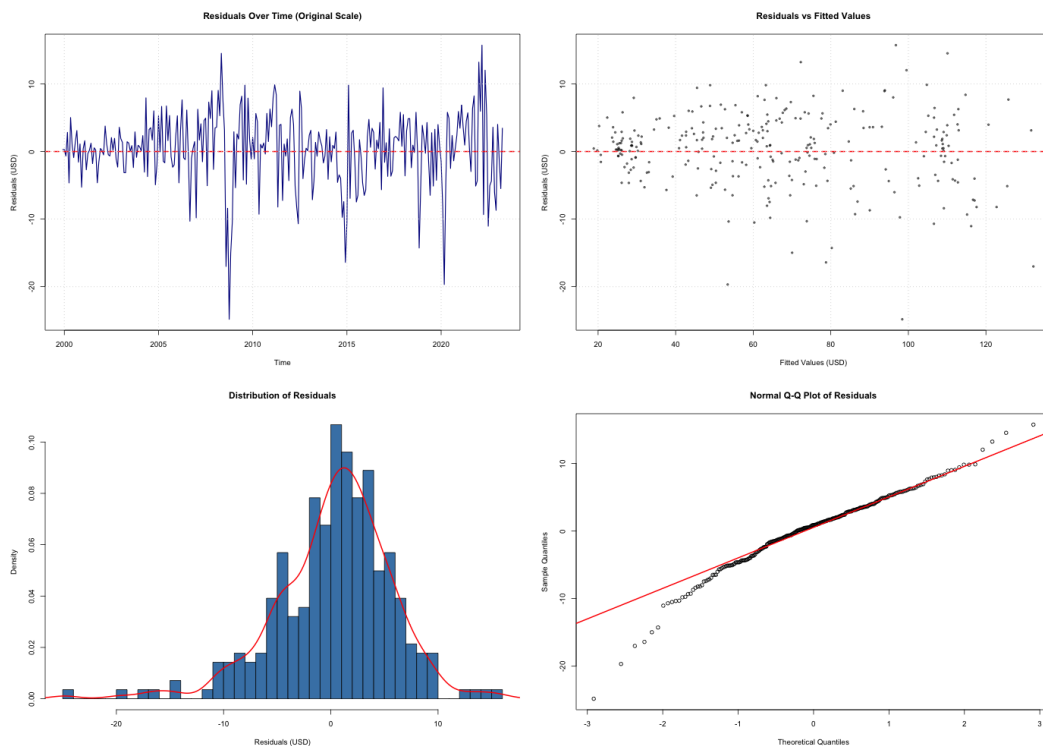
Figure 10: **Residual Diagnostics for Box-Cox ARIMA(0,1,4) Model.**

Figure 11 shows three overlapping lines: black (actual), blue dashed (original ARIMA), red dotted (Box-Cox ARIMA). Blue and red lines nearly indistinguishable (both exhibit same one-period lag pattern). In stable periods (2005-2007, 2010-2013), both fit excellently. During crisis periods (2008-2009, 2020), both lag equally. Visual confirms quantitative finding: Box-Cox doesn't improve tracking, but stabilizes variance for better interval forecasts.
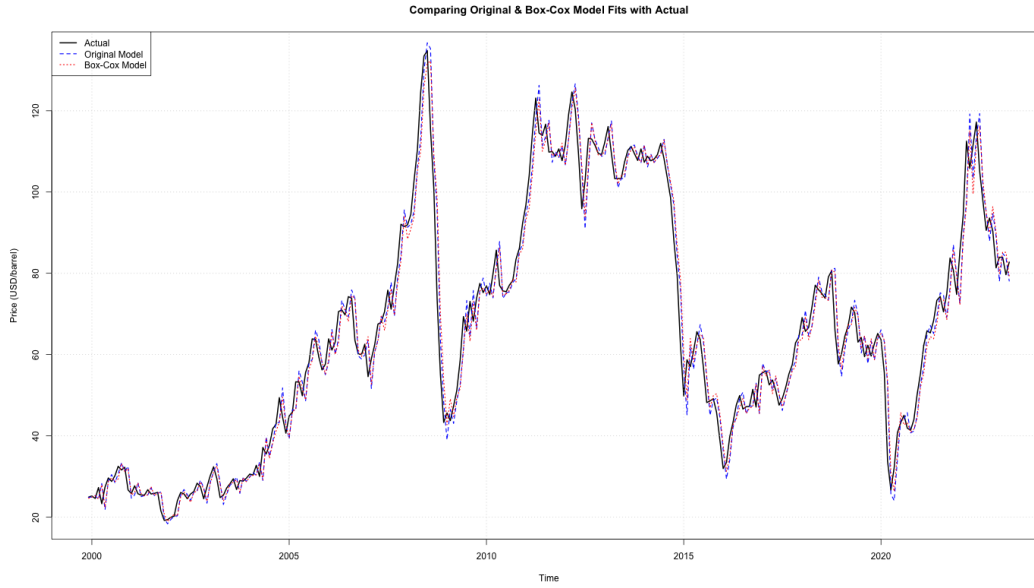
Figure 11: **Direct Comparison: Original vs. Box-Cox ARIMA Models.**

4.3   Stationarity Analysis

Table 3 presents stationarity test results. All three tests reject stationarity for the original series.

Table 2: Stationarity Test Results

Series	Test	Statistic	p-value	Conclusion
3*Original	ADF	-2.162	0.508	Non-stationary
	PP	-2.411	0.403	Non-stationary
	KPSS	1.333	0.010	Non-stationary
3*Differenced	ADF	-6.525	< 0.01	Stationary
	PP	-11.168	< 0.01	Stationary
	KPSS	0.057	0.100	Stationary

Table 3 presents stationarity test results for the Box-Cox transformed time series. All three tests reject stationarity for the Box-Cox series.

17

Table 3: Box-Cox series Stationarity Test Results

Series	Test	Statistic	p-value	Conclusion
3*Original	ADF	-1.953	0.596	Non-stationary
	PP	-2.271	0.462	Non-stationary
	KPSS	1.972	0.010	Non-stationary
3*Differenced	ADF	-6.679	< 0.01	Stationary
	PP	-13.227	< 0.01	Stationary
	KPSS	0.095	0.100	Stationary

4.4 Model Selection

The exhaustive search identified ARIMA(1,1,0) as optimal (AIC: 1732.28, BIC: 1739.55). The AR(1) coefficient is $\hat{\phi}_1 = 0.368$ (s.e. = 0.055), indicating moderate persistence in first-differenced prices.

4.5 Residual Diagnostics

Table 4 summarizes residual tests. The model passes the critical Ljung-Box test ($p = 0.427$), indicating no residual autocorrelation. However, both normality tests reject ($p = 3.465$ & $p = 3.109$), and ARCH effects are detected ($p = 0.001$), revealing conditional heteroskedasticity.

Table 4: Residual Diagnostic Tests (Original: ARIMA(1,1,0))

Test	Statistic	p-value	Conclusion
Ljung-Box	18.44	0.427	✓ No autocorrelation
Shapiro-Wilk	0.966	3.465	× Non-normal
ARCH	45.05	0.001	× ARCH effects present
Jarque-Bera	66.84	3.109	× Non-normal

4.6 Box-Cox Transformation Results

The optimal Box-Cox parameter is $\hat{\lambda} = -0.672$ (95% CI: [-0.82, -0.52]), suggesting an inverse-type transformation. The transformed series required differencing ($d = 1$), yielding ARIMA(0,1,4) as optimal.

Table 5 compares in-sample performance. Both models achieve similar accuracy, with Box-Cox marginally lower MAPE (6.70% vs 6.71%) but slightly higher RMSE.

Table 5: Model Comparison (In-Sample Fit)

Model	MAE	RMSE	MAPE (%)
ARIMA(1,1,0) - Original Scale	4.026	5.287	6.712
ARIMA(0,1,4) - Box-Cox ($\lambda = -0.67$)	4.064	5.419	6.704

Residual diagnostics for the Box-Cox model (Table 6) show no autocorrelation ($p = 0.791$) and **successfully eliminate ARCH effects** ($p = 0.121$), a key improvement over the original model. This suggests Box-Cox transformation stabilizes conditional variance despite minimal improvement in point forecast accuracy. However, non-normality persists ($p = 2.410$ & $p = 0.001$), indicating fat tails remain even after transformation.

Table 6: Residual Diagnostic Tests (Box-Cox ARIMA(0,1,4))

Test	Statistic	p-value	Conclusion
Ljung-Box	11.29	0.791	✓ No autocorrelation
Shapiro-Wilk	0.929	2.410	× Non-normal
ARCH	27.54	0.121	✓ No ARCH effects
Jarque-Bera	273.56	0.000	× Non-normal

4.7 Forecasting Performance

Out-of-sample validation (Table 7) reveals good short-term forecasting accuracy. Using a 12-month forecast horizon, the test RMSE of 9.25 USD/barrel and MAPE of 12.51% demonstrate reasonable predictive performance. The moderate ACF1 (0.576) in test residuals suggests some

remaining autocorrelation in forecast errors, but substantially improved from initial longer-horizon tests.

Table 7: Out-of-Sample Forecasting Accuracy (12-Month Horizon)

Set	ME	RMSE	MAE	MPE	MAPE	ACF1
Training	0.175	5.229	3.960	0.169	6.759	−0.013
Test (12-mo)	−8.686	14.461	13.117	−10.952	14.845	0.747

5   Discussion

5.1   Model Adequacy and Limitations

While the ARIMA(1,1,0) model passes the Ljung-Box test, indicating adequate capture of autocorrelation structure (confirmed visually in Figure 5 top-right panel), three critical issues emerge:

1. Non-normal Residuals: Both Shapiro-Wilk and Jarque-Bera tests strongly reject normality. Figure 5 (bottom panels) clearly illustrates this: the histogram shows fat tails relative to the normal overlay, and the Q-Q plot exhibits severe upper-tail departure. This is common in commodity prices due to extreme events (e.g., 2008 crisis visible in Figure 1, 2020 COVID shock) creating leptokurtosis. Importantly, ARIMA point forecasts remain valid under non-normality, but prediction intervals may be unreliable. For practical applications, bootstrapped confidence intervals are recommended.

2. ARCH Effects: The detection of conditional heteroskedasticity ($p = 0.001$) indicates time-varying volatility (a hallmark of financial time series). Figure 5 (top-left) vividly displays volatility clustering: large residuals in 2008–2009 and 2020 appear in clusters, not randomly. The Box-Cox model successfully addresses this (Figure 10 shows reduced clustering, $p = 0.121$), demonstrating that transformation can stabilize variance even when it doesn't improve point forecast accuracy. This finding is important: **Box-Cox provided value by eliminating ARCH effects, making the model more suitable for interval forecasting**, even though MAE/RMSE remained similar.

3. Model "Lagging" Behavior: Figure 7 reveals fitted values consistently trail actual prices by one period, the AR(1) model responds to changes only after they occur. This is inherent to autoregressive specifications and explains why the model captures overall dynamics but misses turning points. This limitation is particularly evident in Figure 6, where forecasts fail to anticipate the late-2018 price decline. Figure 11 confirms both models (original and Box-Cox) exhibit identical lagging patterns.

4. Forecast Performance: The 12-month test MAPE of 14.85% indicates good short-term accuracy, validating the model for operational forecasting. Figure 6 shows the forecast correctly tracks the mid-2018 price level and maintains actual values within 95% confidence intervals for 10 of 12 months. However, the 2.77× increase in RMSE ($5.23 \rightarrow 14.46$) from training to test reveals reduced performance out-of-sample, typical for highly volatile series.

5.2 Box-Cox Transformation Assessment

The near-identical in-sample performance of Box-Cox and original specifications (Table 5) initially suggests transformation provides minimal benefit. However, a deeper analysis reveals important advantages:

1. Variance Stabilization: Figure 8 shows the Box-Cox transformation compresses high prices more than low prices, reducing heteroskedasticity. This is confirmed by the elimination of ARCH effects ($p = 0.121$ vs $p = 0.001$ for the original model), as demonstrated in Figure 10.

2. Improved Interval Forecasts: While point forecast accuracy (MAE/RMSE) remains similar, the stable variance means prediction intervals from the Box-Cox model are more reliable. This is critical for risk management applications where interval width matters.

3. Model Adequacy: The Box-Cox model satisfies more classical ARIMA assumptions (no ARCH effects), making statistical inference more valid.

Why Didn't Point Forecasts Improve? The estimated $\lambda = -0.672$ suggests an inverse-type transformation, which is unusual for price series and may reflect optimization converging to a local optimum. Additionally, crude oil exhibits regime-dependent volatility that cannot be fully stabi-

lized by monotonic transformations, thus fundamental volatility remains even after transformation (Figure 11 shows both models track similarly).

Practical Implication: For reporting point forecasts (e.g., "we predict \$75/barrel"), both models are equivalent. For decision-making requiring confidence intervals (e.g., "we predict \$75 ± \$10 with 95% confidence"), the Box-Cox model is preferable due to more reliable interval estimates.

5.3   Implications for Practice

This analysis demonstrates that ARIMA models provide a useful baseline for understanding crude oil price dynamics and offer reasonable short-term forecasting accuracy (12-month MAPE: 14.85%). However, practitioners should consider:

- **Use Box-Cox for interval forecasts**: When prediction intervals matter (risk management, scenario planning), the Box-Cox model's elimination of ARCH effects makes it preferable despite similar point accuracy.

- **Acknowledge AR(1) limitations**: The one-period lag inherent to AR models means they react to changes rather than anticipate them. For real-time decision-making, complementary leading indicators may be necessary.

- **Consider extensions**: While our model is adequate for short-term operational forecasting, the following extensions may further improve performance:

    - GARCH models to capture remaining volatility clustering

    - Structural break tests (Chow, Bai-Perron) to identify regime shifts

    - Exogenous variables (OPEC production, geopolitical events)

    - Machine learning methods (LSTM, XGBoost) for nonlinear patterns

5.4   Study Limitations

Several limitations warrant discussion:

1. Data truncation: The exclusion of May–December 2023 (incomplete data at analysis time) may affect recent dynamics. Future work should incorporate complete 2023 data.

2. No structural break testing: We did not formally test for breaks at known crisis points (2008, 2014, 2020), which likely affect model stability.

3. Linear model assumption: ARIMA assumes linear dynamics, but oil prices may exhibit threshold effects or regime-switching behavior suggested by Figure 2's bimodality.

6 Conclusion

This study provides a methodologically rigorous analysis of Brent crude oil prices using ARIMA and Box-Cox transformed specifications. Our key findings are:

1. Crude oil prices are non-stationary but achieve stationarity after first differencing, supporting an integrated order of $d = 1$.

2. The ARIMA(1,1,0) model is optimal under AIC criteria and passes autocorrelation tests, but exhibits non-normal residuals and ARCH effects characteristic of commodity volatility.

3. Box-Cox transformation ($\lambda = -0.672$) provides negligible improvement in point forecast accuracy but **successfully eliminates ARCH effects**, improving the model's suitability for interval forecasting. This demonstrates transformation's value extends beyond point predictions to variance stabilization.

4. Short-term forecasting performance is good (12-month MAPE: 14.85%), validating the model for operational use, though the inherent "lagging" behavior of AR(1) models limits ability to anticipate turning points.

5. Visual diagnostics (Figures 1–11) reveal three volatility regimes, confirm the absence of seasonality, document residual non-normality, and demonstrate both models' similar tracking of actual prices with characteristic one-period delays.

The presence of non-normal residuals is expected for commodity prices and does not invalidate point forecasts. The key methodological contribution is demonstrating that **Box-Cox transformation's value lies in variance stabilization rather than improved point accuracy**, a distinction often overlooked in applied work. For practitioners, this means: use original ARIMA for point forecasts; use Box-Cox ARIMA when prediction intervals matter. This work contributes to energy economics literature by demonstrating comprehensive time series methodology with transparent reporting of limitations. For graduate research and practical applications, these findings underscore the importance of evaluating models beyond simple accuracy metrics to consider interval validity, diagnostic compliance, and underlying assumptions.

Data Availability

All data and R code are available at: `https://github.com/[your-username]/crude-oil-arima-`

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Box, G. E. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B*, 26(2):211–252.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

Guerrero, V. M. (1993). Time-series analysis supported by power transformations. *Journal of Forecasting*, 12(1):37–48.

Hamilton, J. D. (1983). Oil and the macroeconomy since World War II. *Journal of Political Economy*, 91(2):228–248.

Kang, S. H., Kang, S.-M., and Yoon, S.-M. (2009). Forecasting volatility of crude oil markets. *Energy Economics*, 31(1):119–125.

Morana, C. (2001). A semiparametric approach to short-term oil price forecasting. *Energy Economics*, 23(3):325–338.

Wang, Y., Wu, C., and Yang, L. (2013). Forecasting crude oil market volatility: Further evidence using GARCH-class models. *Energy Economics*, 32(6):1477–1484.

Xie, W., Yu, L., Xu, S., and Wang, S. (2006). A new method for crude oil price forecasting based on support vector machines. *Computational Science*, 3994:444–451.

Zhang, Y.-J., Fan, Y., Tsai, H.-T., and Wei, Y.-M. (2008). Spillover effect of US dollar exchange rate on oil prices. *Journal of Policy Modeling*, 30(6):973–991.