Monthly Diesel Price Movement and Volatility in Metro Manila: An ARIMA-GARCH Approach

Lanz Railey A. Fermin, Rafael Ma. Vicente D. Tagulao, Cyd Nicolas A. Santos

Abstract

Fuel prices are critical to economic activity, impacting transportation costs, consumer prices, and market stability. In the first half of 2025, Philippine diesel prices cumulatively increased by over ₱11, demonstrating the domestic market's vulnerability to global crude oil inelasticity and supply fluctuations. Consequently, understanding and forecasting fuel price behavior is essential for informed decision-making across all sectors. Drawing inspiration from ARIMA-GARCH models applied to oil markets in Pakistan and the United States, this study aims to identify an effective ARIMA-GARCH model for characterizing the monthly average diesel price in Metro Manila. The dataset, containing observations from January 2000 to November 2021, was sourced from the Department of Energy. After model fitting and refinement, an $\mathrm{ARIMA}(1,1,0)$-$\mathrm{GARCH}(1,1)$ model with an AIC of 3.5560 was selected. The mean model suggests that month-to-month diesel price changes modestly depend on the previous month's change, while the GARCH coefficients indicate high persistence ($\alpha_1 + \beta_1 \approx 0.999$), implying that significant price fluctuations tend to prolong before stabilizing.

1 Introduction

1.1 Background of the Study

Fuel is a major driver of economic activity in the Philippines, with its effects directly felt in the transportation sector. Price increases raise the cost of moving goods and people, contributing to inflation and prompting shifts in labor allocation. Beyond transport, fuel costs also affect electricity generation, emissions, and the like [14]. Globally, crude oil's inelastic nature makes fuel prices highly sensitive to geopolitical events and the decisions of major producers [10]. Given this volatility, the ability to accurately forecast fuel prices becomes essential. Reliable price models can support decision-making among consumers, regulators, and industry actors. Studies in Pakistani [6] and American [12] markets demonstrate that ARIMA-GARCH models effectively capture both trend and conditional volatility in oil prices. The insights derived from such models are vital for strengthening economic resilience and guiding effective policy interventions amid global price fluctuations.

1.2 Statement of the Problem

The main research question of this study is whether the ARIMA-GARCH model is sufficient to effectively capture the underlying dynamics of average monthly diesel prices in Metro Manila, Philippines. This study seeks to assess the limitations of simpler time series models in capturing the correlations that GARCH models are designed to account for. In addition, this study aims to extract information from the structure and parameters of the identified model to determine broader trends in fuel prices. To address this question, the study presents an analysis of historical data on average monthly diesel prices in Metro Manila, Philippines.

Submission for MATH 271.2 Advanced Time Series and Forecasting, AY 2025-2026 Intersession

1.3 Scope and Limitations

This study will primarily focus on modeling volatility using the ARIMA-GARCH framework, placing no emphasis on seasonal dynamics and excluding more complex GARCH variants. The dataset used in this analysis will be limited data from the year 2000 onward, ensuring relevance to the official price monitoring scheme by the Oil Industry Management Bureau of the Department of Energy (DOE) [3]. By narrowing the scope, the study aims to produce contextually meaningful results while avoiding overfitting and unnecessary model complexity.

2 Methods

2.1 Preliminaries

In this section, we recall and add some key definitions and properties that would serve as foundations for the analysis of the time series data and the identification of the best-fitting model done in the succeeding sections.

Definition 1 (ARIMA-GARCH Model)**.** A time series $\{X_t\}$ that follows an ARIMA model with GARCH innovations, denoted by $\mathrm{ARIMA}(p,d,q)$-$\mathrm{GARCH}(m,s)$ model for (X_t, σ_t):

$$\begin{cases} X_t & = \sum_{i=1}^{p} \phi_i X_{t-i} + Y_t + \sum_{j=1}^{q} \theta_j Y_{t-j} \\ Y_t & = \sigma_t \varepsilon_t \\ \sigma_t^2 & = \alpha_0 + \sum_{i=1}^{m} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2 \end{cases} \tag{1}$$

where $\varepsilon_t \overset{\text{iid}}{\sim} (0,1)$, and coefficients satisfy stationarity, causality, and invertibility conditions. For definitions of stationarity and both the ARIMA and GARCH models, refer to Appendix A.

Definition 2 (Weighted Ljung-Box Test)**.** This variant of the original Ljung-Box Test defined in Definition 8 in Appendix A, proposed by Fisher and Gallagher, uses a weighted statistic that gives weight on earlier lags.

This is usually used for diagnostics on standardized residuals since their distributions differ from the raw data itself, and so the classical Q-statistic may no longer be valid [4]. Its statistic is given by

$$\tilde{Q}_{\mathrm{W}} = n(n+2) \sum_{h=1}^{m} \frac{m-h+1}{m} \frac{\hat{\rho}_h^2}{n-h}. \tag{2}$$

The test selects three certain lags to be checked for autocorrelation, based on the degrees of freedom $p+q$: one for minimal checking ($h_1 = 1$), another for moderate depth ($h_2 = [2\times(p+q)+(p+q)-1]$), and the last one for extended coverage ($h_3 = [4 \times (p+q) + (p+q) - 1]$).

For definitions of other statistical tests in determining stationarity and autocorrelation of time series data, refer to Appendix A.

2.2 Data Description

This study utilizes a monthly dataset of average diesel prices in Philippine Pesos per liter, sourced from the DOE [3]. This dataset, last updated in November 2021, is based on the DOE's regular weekly publications of retail prices from selected gas stations across Metro Manila. This analysis will primarily focus on diesel, excluding other fuel types like gasoline, kerosene, and LPG.

As shown in Figure 1, the time series plot of the data does not exhibit clear seasonal patterns but has an apparent upward trend. The visual inspection suggests that the data may not be stationary. A prominent spike is evident in 2008, aligning with the onset of the Global Financial Crisis. In other years, some clusters with low variance can be observed, as indicated by smaller fluctuations in price.

2.3 Relevant Assumptions

In this study, we restrict the lag length to values less than 5 ($h < 5$), in line with the principle of parsimony, which emphasizes model simplicity without compromising explanatory power. Simpler models are generally more robust and less vulnerable to overfitting, provided they offer an adequate fit to the data. All statistical and hypothesis tests are performed at a 95% confidence level ($\alpha = 0.05$).

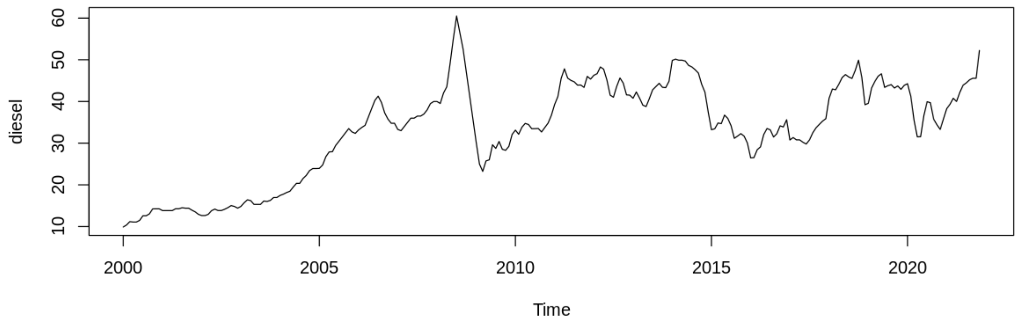

Figure 1: Monthly average diesel prices for Metro Manila, as reported by the Department of Energy. Data points include prices from January 2000 to November 2021.

2.4 Initial Diagnostics

The entirety of data exploration and modeling was performed using Google Colab and the following libraries: `TSA`, `tseries`, `itsmr`, `forecast`, `lmtest`, and `rugarch`. Before the time series modeling can be performed, it is necessary that the diesel prices data be stationary and autocorrelated. These are assessed using the ADF and Ljung-Box tests, respectively.

The ADF test was performed using the `adf.test` function, and returned a Dickey-Fuller statistic of -2.8088 and a p-value of 0.2353. Hence, there is no sufficient evidence to reject the null hypothesis; thus, the time series on the original diesel prices is nonstationary and requires transformation. To address this, the series was transformed using first-order differencing. After differencing, the ADF test now returns a Dickey-Fuller statistic of -6.0224 and a p-value of 0.01. This time, there is sufficient evidence to say that the differenced time series is already stationary.

Meanwhile, employing the Ljung-Box test was implemented via `Box.test` with the `Ljung-Box` type. It yielded a Q-statistic value of 58.004 with 1 degree of freedom, and a p-value less than 2.62×10^{-14}. This result indicates the rejection of the null hypothesis and concludes that there exists an autocorrelated relationship between the data points of the differenced diesel prices. These two tests, then, imply that the current data is now appropriate for fitting linear time series and volatility models.

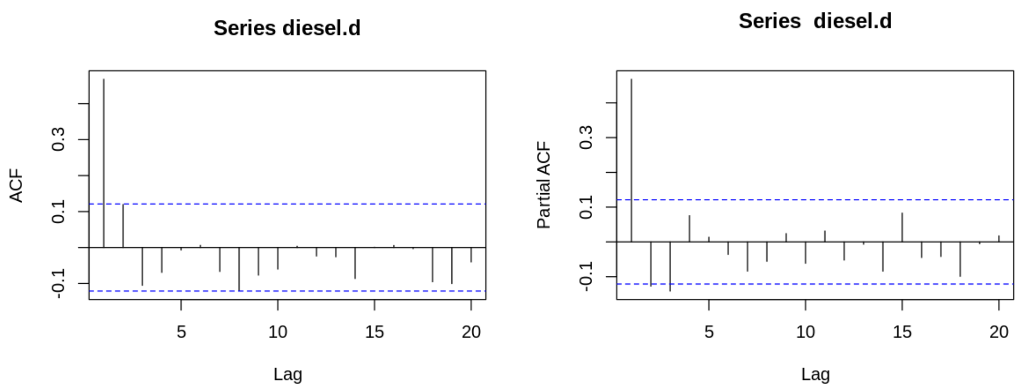

Figure 2: ACF and PACF correlograms for the differenced diesel prices.

2.5 Mean Model Fitting

The identification of prospective ARIMA models, which would subsequently be evaluated for goodness-of-fit with the diesel prices data, started with the observation of possible lag orders through inspection of the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots as displayed in Figure 2.

First, as we employed one differencing to make our time series data stationary, we define that the order $d = 1$. Moreover, the PACF plot reveals that there are significant spikes in lags 1, 2, and 3; hence, we include these as possible orders in the $\mathrm{AR}(p)$ models (i.e. $p = 0, 1, 2, 3$). Finally, the ACF plot shows that the possible orders q for the MA(q) models are $q = 0, 1$.

Using the `Arima` function in `R`, and evaluating the AIC for each of the combinations for the candidate models for the ARIMA model shows that the $\mathrm{ARIMA}(3, 1, 0)$ process returns the lowest AIC with a value of 1025.78.

2.6 Volatility Model Fitting

Further analyzing the fitted ARIMA$(3, 1, 0)$ model to the original time series shows that the residuals show an ARCH effect. Running a Ljung-Box test on the residuals of the mean model results in a $Q-$statistic of 0.031073 with a $p-$value of 0.8601. Hence, there is no statistically significant evidence of autocorrelation in the residuals (see Appendix B.1 for the pertinent ACF and PACF plots).

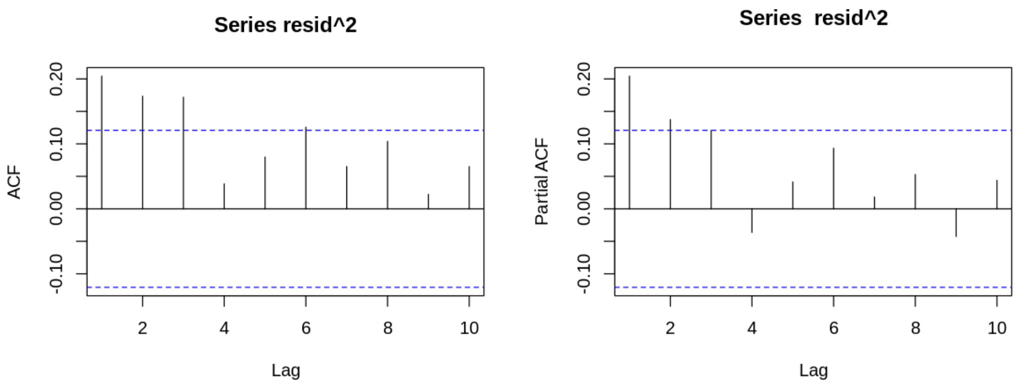

Figure 3: ACF and PACF correlograms for the squared residuals of the mean model.

However, Figure 3 shows some spikes in lags $1, 2,$ and 3. This indicates the evidence of the ARCH effects, suggesting that the variance of the residuals is not constant over time and exhibits autocorrelation. Thus, we fit another ARCH or GARCH model to capture this.

The PACF of the squared residuals show that possible orders for m are $0, 1,$ and 2. However, Tsay suggests that for financial time series, only lower order GARCH models are used in practical applications [11]. Moreover, [13] also encountered that in modeling pricing options, both of ACF and PACF plots showed intersections at lags 1, 2, and 3. As a result, they restricted the orders of their GARCH model until 2. Hence, for this model, we also limit the orders m, s to $0 \leq m, s \leq 2$. Performing a grid search using `Arima` on the squared residuals from the mean model showed that the optimal orders are $m = 1$ and $s = 1$, with the lowest AIC value of 1641.48.

2.7 Model Refinement and Residual Diagnostics

Previously, we have determined that an ARIMA$(3, 1, 0)-$GARCH$(1, 1)$ model seem to be the most sufficient to model the diesel prices in Metro Manila. However, further refinements on the model will be made should the `ugarchfit` report show that some coefficients of the current model are not significant.

Afterwards, diagnostic checks are performed on both the standardized residuals and the standardized squared residuals to assess whether the model has adequately captured the overall behavior and volatility of the time series. This involves verifying the absence of autocorrelation in both.

3 Results and Discussion

3.1 Initial Obtained Model

From the initial model fitting performed in Section 2, the current best-fit model was found to be $\mathrm{ARIMA}(3,1,0)$-$\mathrm{GARCH}(1,1)$. However, the significance levels of these coefficients, as outlined in the GARCH model report created by the function `ugarch`, showed indicators of potential model refinement. For instance, the `omega` term, which denotes the long-run average level of volatility, was also found to be not statistically significant. Additionally, the `ar2` and `ar3` terms were found to have p-values greater than 0.05. This meant that an improvement for the current model would include its reduction to a simpler autoregressive model of order 1.

As such, we consider the $\mathrm{ARIMA}(1,1,0)$-$\mathrm{GARCH}(1,1)$ model as a refinement and fit it with the current data. Equation 3 writes out the model equation, in accordance with the definition:

$$\begin{cases} \nabla X_t = 0.4844 X_{t-1} + Y_t \\ Y_t = \sigma_t \varepsilon_t \\ \sigma_t^2 = 0.1242 Y_{t-1}^2 + 0.8748 \sigma_{t-1}^2 \end{cases} \tag{3}$$

where $\varepsilon_t \overset{\text{iid}}{\sim} (0,1)$ and the coefficients satisfy the necessary conditions.

Table 1 compares the estimates for the models before and after refinement, as well as their evaluation metrics. It can be noted that the refined model did not have any coefficients that are not statistically significant, except for the `omega`. Along with the lower value on AIC and other information criteria, these indicate that, in fact, an improvement to the model has been made.

Table 1: Comparison of model parameters, their p-values, and information criteria scores of initial and refined model.

Parameter	**Initial Model** $\mathrm{ARIMA}(3,1,0)$-$\mathrm{GARCH}(1,1)$		**Refined Model** $\mathrm{ARIMA}(1,1,0)$-$\mathrm{GARCH}(1,1)$	
	Estimate	$\Pr(>\lvert t\rvert)$	Estimate	$\Pr(>\lvert t\rvert)$
ar1	0.525813	0.000000	0.484394	0.00000
ar2	-0.033217	0.684096	-	-
ar3	-0.122755	0.079683	-	-
omega	0.024723	0.194096	0.022722	0.19167
alpha1	0.119906	0.000001	0.124245	0.00000
beta1	0.879094	0.000000	0.874755	0.00000
Information Criteria	**Score**		**Score**	
Akaike IC	3.6632		3.5560	
Hannan-Quinn IC	3.6961		3.6779	
Bayes IC	3.7450		3.7204	

Further interpretation of the same report generated by `ugarchfit` confirms this claim. Evaluation of the standardized residuals and squared residuals through the Weighted Ljung-Box Test, as in Definition 2, revealed that no significant serial correlation was observed on all automatically selected lags based on the model's degrees of freedom.

Table 2 outlines these findings for both standardized residuals and standardized squared residuals. With these residual analyses, it can be concluded that no significant ARCH effect is present. Moreover, this also suggests that the refined ARIMA-GARCH model has a well-specified mean model and captures volatility clustering relatively well.

Table 2: Weighted Ljung-Box test results on standardized residuals and standardized squared residuals of the refined model.

Std. Residuals (df = 1)			**Std. Squared Residuals (df = 2)**		
Lag	Statistic	p-value	Lag	Statistic	p-value
1	1.449	0.22866	1	0.019	0.0890
2	1.490	0.44255	5	4.748	0.1744
5	5.693	0.06041	9	6.089	0.2880

3.2 Model Interpretation

The mean model $\mathrm{ARIMA}(1, 1, 0)$ implies that the first difference of diesel prices, that is, the month-to-month change in diesel pump price follows a first-order autoregressive process. This means that the change modestly depends on the previous month's movement. For instance, if the diesel price increased last month, it can be inferred that there would be a tendency, albeit not guaranteed, for the price to continue increasing this month at a reduced strength. This seeming momentum in price changes can be attributed to real-life factors and economic realities that affect diesel prices such as persistent supply-demand conditions, imposition of various tax policies like tariffs and excise tax collections, and more importantly, the ongoing trends in global oil prices.

On the other hand, the volatility model $\mathrm{GARCH}(1, 1)$ component captures the conditional variance of the diesel price changes. The estimated parameters, represented by the α_1 and β_1, show that both recent price shocks and lagged volatility contribute to current volatility. Notably, the sum of these coefficients is approximately 0.999, which is very close to 1. This indicates a high degree of volatility persistence or "stickiness", a common feature in financial and energy markets.

In practical terms, this implies that when a significant fluctuation in diesel prices occurs, the elevated level of variability is likely to persist for an extended period before stabilizing. This behavior is consistent with periods of economic or geopolitical stress, during which fuel price volatility remains elevated due to continued market uncertainty. A short note relating the Philippine context on diesel prices can be found in Appendix B.2.

The findings of this short project are parallel to prior research that also attempted to model and forecast fuel prices. In [8], it was found that using GARCH for American crude oil prices suggests that modeling the series as a martingale process is more appropriate than using a simple random walk model. They added that using a martingale process to do a one-month forward forecast has been found to create more reliable interval forecasts for oil prices; hence, making it also a more reliable predictor.

Meanwhile, [9] identified the Integrated GARCH (IGARCH) model as the most effective framework for capturing volatility in crude oil prices due to its robustness and minimal forecast error metrics. Their conclusions are particularly relevant to this project for two reasons. First, diesel prices tend to move in a similar direction with crude oil prices, as diesel is a direct derivative of crude. Second, the near-unit sum of the estimated GARCH parameters in the refined model suggests a volatility structure that shares similar properties with IGARCH, in which shocks to volatility exhibit long memory and slow mean reversion.

3.3 Long-term Forecasting

To conclude this exploration on diesel price, we observe the refined model's long-term behavior by performing a five-month-ahead forecast, as shown in Figure 4. The projected values, shown as the orange line, display a general downward trend. It can be said that diesel prices in the months after November 2021 are expected to gradually decrease and stabilize until early 2022.

The yellow shaded region illustrates the 1-sigma (one standard deviation) unconditional confidence bands. As the forecast extends further, the uncertainty increases, as evidenced by the widening of the yellow bands. This is a typical characteristic of time series forecasts, where predictive confidence diminishes over longer horizons.

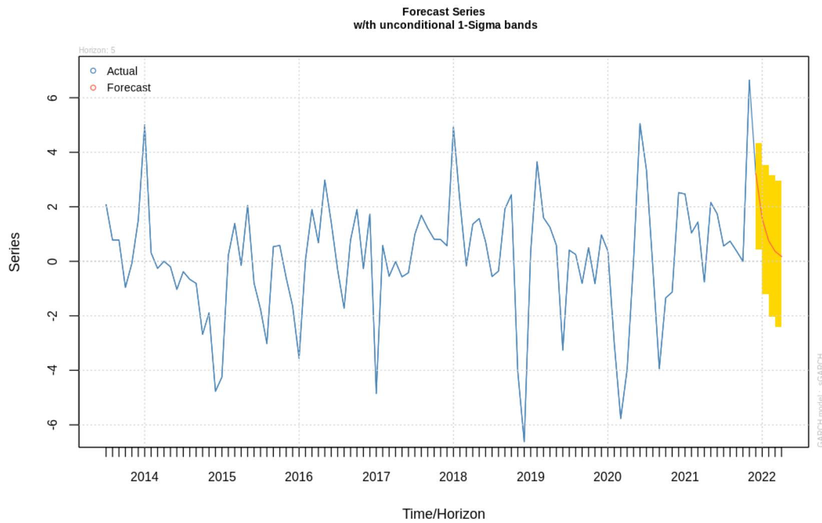

Figure 4: A five-step-ahead forecast using the refined model $\mathrm{ARIMA}(1,1,0)$-$\mathrm{GARCH}(1,1)$.

References

[1] H. Akaike. Fitting autoregreesive models for prediction. In *Selected Papers of Hirotugu Akaike*, pages 131–135. Springer, 1969.

[2] D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a):427–431, 1979.

[3] G. Domingo and T. Ty. What Fuel Prices Can Tell Us: DOE Price Monitoring Data in Competition Enforcement and Consumer Welfare. Discussion paper, Philippine Competition Commission, 2021.

[4] T. J. Fisher and C. M. Gallagher. New Weighted Portmanteau Statistics for Time Series Goodness of Fit Testing. *Journal of the American Statistical Association*, 107(498):777–787, June 2012.

[5] D. N. Flores. Fuel prices up by over P1 per liter, ending two-week price cuts, July 14 2025.

[6] A. Ghaffar, U. Aslam, and Z. Zardari. Gauging the Fuel Price Volatility using GARCH Models. *International Journal of Emerging Business and Economic Trends*, 2(2):115–121, 2023.

[7] G. M. Ljung and G. E. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

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

[9] F. W. Ng'ang'a. *Modelling and forecasting of crude oil price volatility: Comparative analysis of volatility models*. PhD thesis, Strathmore University, 2021.

[10] R. Selmi, S. Hammoudeh, and M. E. Wohar. What drives most jumps in global crude oil prices? *World Economy*, 46(3):598–618, 2022.

[11] R. S. Tsay. *An Introduction to Analysis of Financial Data with R*. Wiley Series in Probability and Statistics. Wiley-Blackwell, Hoboken, NJ, Oct. 2012.

[12] Y. Xiang. Using ARIMA-GARCH model to analyze fluctuation law of international oil price. *Mathematical Problems in Engineering*, pages 1–7, 2022.

[13] A. Yasmin, R. Riaman, and S. Firman. Optimization Modeling of Investment Portfolios Using The Mean-VaR Method with Target Return and ARIMA-GARCH. *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, 10:147–165, 03 2025.

[14] Zero Carbon Analytics. Increasing gas imports will raise electricity prices in the Philippines - Zero Carbon Analytics, July 8 2025.

A Mathematical Preliminaries

Definition 3 (Stationarity)**.** Let $\{X_t\}$ be a time series process. Then $\{X_t\}$ is weakly stationary if it satisfies the following conditions: (1) $\mu_X(t)$ is independent of t, and (2) $\gamma_X(t+h,t)$ is independent of t for all h.

Definition 4 (Autocorrelation)**.** The process $\{X_t\}$ is said to have autocorrelation if there is a statistically significant linear relationship between its current value and one or more of its lagged values. That is, if $\operatorname{cov}(X_t, X_{t-k}) \neq 0$ for some $k \neq 0$.

Definition 5 (ARIMA)**.** A time series process $\{X_t\}$ is said to be an integrated ARMA model $\mathrm{ARIMA}(p,d,q)$ if $\nabla^d X_t = (1-B)^d X_t$ is a causal $\mathrm{ARMA}(p,q)$ process. The model is usually written as

$$\phi(B)(1-B)^d X_t = \delta + \theta(B) Z_t, \tag{4}$$

where d is a nonnegative integer, $\delta = \mu(1-\phi_1-\cdots-\phi_p)$, and $Z_t \sim \mathrm{WN}(0,\sigma^2)$.

Definition 6 (GARCH)**.** A time series process $\{Y_t\}$ is a generalized autoregressive conditionally heteroscedastic model of order (m,s), denoted as GARCH(m,s) if it is of the form

$$\begin{cases} Y_t = \sigma_t \epsilon_t \\ \sigma_t^2 = \alpha_0 + \displaystyle\sum_{i=1}^{m} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2 \end{cases} \tag{5}$$

where $\epsilon_t \overset{\text{iid}}{\sim} (0,1)$, $\alpha_0 > 0$, $\alpha_i, \beta_j \geq 0$, and $\sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) < 1$.

Definition 7 (ADF Test)**.** The Augmented Dickey-Fuller (ADF) test is a common statistical test used to check whether a given time series process is stationary or not through the existence of a unit root in an autoregressive polynomial [2].

Consider the AR polynomial $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$.

1. $H_0 : \phi_1 = 1$
   $H_1 : \phi_1 < 1$
2. Test statistic: $\hat{\tau}_\mu := \hat{\phi}_1^* \,/\, \widehat{SE}(\hat{\phi}_1^*)$
3. Reject H_0 if $\hat{\tau}_\mu < -2.86$. It follows that the process is stationary.

Definition 8 (Ljung-Box Test)**.** The Ljung-Box test is a portmanteau test that checks whether the autocorrelations of a given time series process at multiple lags are different from zero. The same test is also used for residual diagnostics [7].

1. $H_0 : \rho_1 = \rho_2 = \cdots = \rho_m = 0$
   $H_1 : \rho_i \neq 0$ for some $i \in \{1, 2, \ldots, m\}$.
2. Test statistic: $Q = n(n+2) \sum_{h=1}^{m} \frac{\hat{\rho}_h^2}{n-h}$
3. Reject H_0 if $Q > \chi^2(m, 1-\alpha)$. It follows that the process is autocorrelated.

Definition 9 (AIC)**.** The Akaike Information Criterion (AIC) is a metric used to evaluate the relative quality of a statistical model for a given time series data. Given a set of prospective models, the model with the lowest AIC value is deemed the best [1]. With $\hat{\sigma}_k^2$ as the maximum likelihood estimator of the variance in a normal regression model, and k as the number of parameters in the model, the AIC is defined as

$$\mathrm{AIC} = \log \hat{\sigma}_k^2 + \frac{n+2k}{n}, \tag{6}$$

B Supplementary Tables and Figures

B.1 Mean Model Residual ACF and PACF Plots

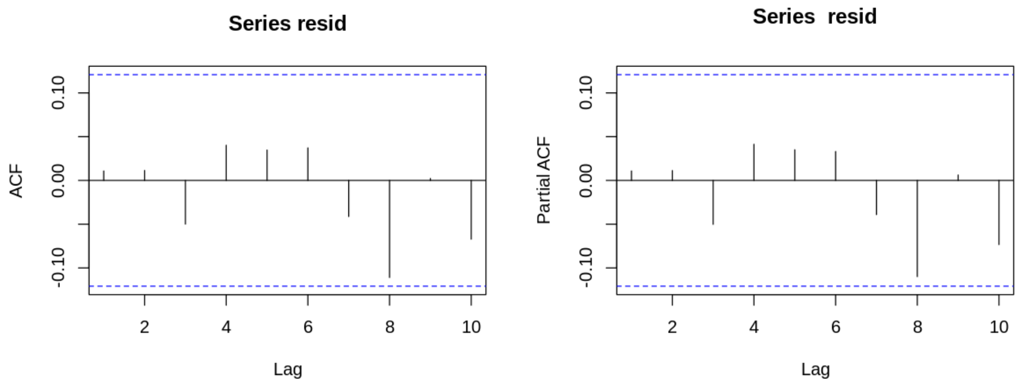

Figure 5: ACF and PACF Plots fot the residuals after fitting the mean model.

B.2 A Remark on Diesel Volatility

The volatility spikes, and clustering, on diesel prices can be empirically observable in several periods within the dataset. For example, during the 2008 Global Financial Crisis, global crude oil prices surged to record highs before crashing abruptly. This led to sharp and persistent fluctuations in domestic diesel prices. The enactment of the TRAIN Law in the Philippines during 2018 also prompted sudden hikes in fuel prices, primarily due to the uptick of excise taxes imposed by the government. Finally, in 2020, the COVID-19 pandemic led to a dramatic collapse in oil demand worldwide, causing significant drops in fuel prices, followed by unstable price fluctuations as public mobility changed over time. Figure 6 visualizes these spikes, with lines indicating the events.

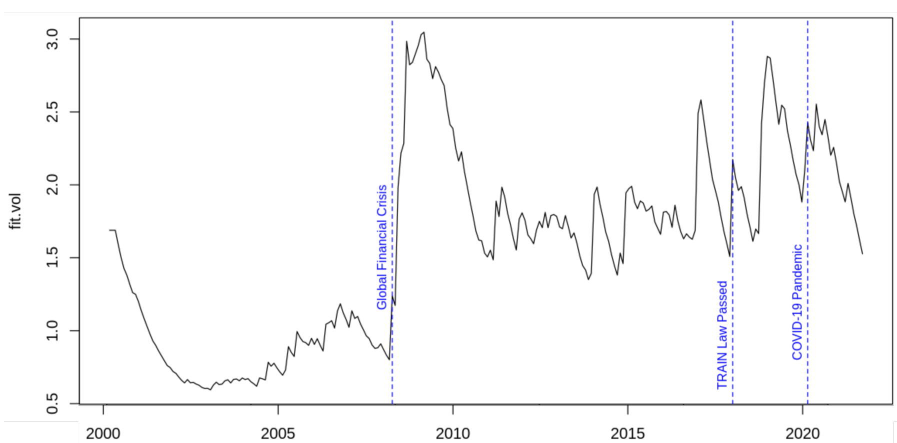

Figure 6: The volatility plot of diesel price changes, as captured by the $\mathrm{ARIMA}(1,1,0)$ and $\mathrm{GARCH}(1,1)$ model. The label in each of the blue lines indicate important contexts throughout the years.