3080 Project Part 2: Regional Volatility Analysis

Nemmo Ciccone

2025-12-09

Introduction

Is there a statistically significant difference in the volatility (variance of weekly price changes) of regular gasoline between the Midwest and Central Atlantic regions?

This question was chosen to investigate the relative volatility of fuel costs in two distinct U.S. markets. Volatility is a critical economic metric representing risk. Higher volatility in gasoline prices introduces uncertainty for household budgets and business logistics. The Midwest and Central Atlantic regions differ significantly in their supply infrastructure: the Midwest relies on a network of regional refineries and pipelines that can be isolated during disruptions, whereas the Central Atlantic has access to the Colonial Pipeline and global imports. Assessing the difference in their price volatilities provides insight into regional market stability.

The data used for this analysis consists of weekly retail gasoline prices collected by the U.S. Energy Information Administration (EIA) from June 2000 to October 2025. To measure volatility, we transformed the raw price data into log first differences ($Log(Price_t) - Log(Price_{t-1})$). This transformation calculates the weekly change in prices, effectively removing long-term trends and stationarity issues, making it appropriate for comparing dispersion between the two regions.

Methods

To address the research question, we conduct the **Brown-Forsythe Levene's Test** (aka The Modified Levene Test) for equality of variances. This test was selected because it is a robust alternative to the standard F-test. While the F-test is highly sensitive to non-normality, the Brown-Forsythe test (which performs Levene's test using the median rather than the mean) is robust against heavy-tailed distributions, which are often present in financial time series data. By comparing the variances of the weekly price changes in the Midwest and Central Atlantic, we can determine if one region experiences significantly more volatile price swings than the other.

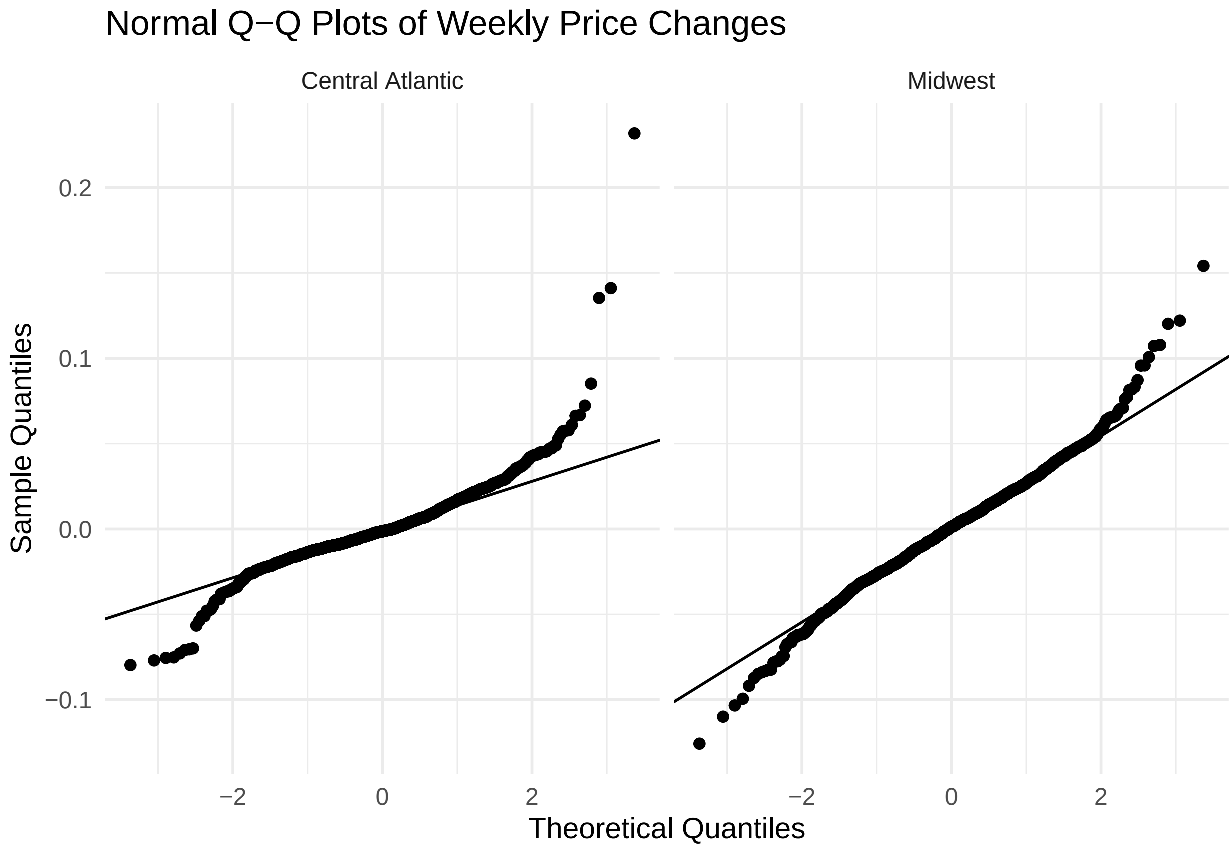
Assumptions

The validity of the Brown-Forsythe test relies on the following assumptions:

1. **Independence**: The samples from the two regions must be independent, and observations within each sample must be independent. We assume that the weekly price changes in the Midwest and Central Atlantic are independent for this analysis, although we acknowledge that time series data may exhibit some autocorrelation and time dependency.
2. **Normality***: The standard F-test requires strict normality. However, our assessment of the Normal Q-Q plots (below) reveal heavy tails, indicating the data is not normally distributed. The Brown-Forsythe test is specifically chosen because it does not require the data to be normally distributed and is robust to these deviations.

```r
# Import and clean data
reg_gas <- read_excel("C:/Users/19413/Downloads/eia_gas_data.xls", sheet = "Data 3", skip = 2)

# Select specific columns for Midwest and Central Atlantic
padd_cols <- c(
    "Date",
    "Weekly Central Atlantic (PADD 1B) Regular All Formulations Retail Gasoline Prices  (Dollars per Gal
    "Weekly Midwest Regular All Formulations Retail Gasoline Prices  (Dollars per Gallon)"
)
reg_gas_regions <- reg_gas[, padd_cols]
colnames(reg_gas_regions) <- c("Date", "Central Atlantic", "Midwest")

reg_gas_regions <- reg_gas_regions[-(1:511), ]

# Pivot to long format
reg_gas_long <- reg_gas_regions %>%
    pivot_longer(cols = -Date, names_to = "Region", values_to = "Price")

# Calculate log changes
reg_gas_logs <- reg_gas_long %>%
    group_by(Region) %>%
    mutate(
        Log_Price = log(Price),
        Log_Change = Log_Price - lag(Log_Price)
    ) %>%
    ungroup() %>%
    na.omit()
```

```r
# Q-Q Plots for Normality Check
ggplot(reg_gas_logs, aes(sample = Log_Change)) +
    stat_qq() +
    stat_qq_line() +
    facet_wrap(~Region) +
    labs(
        title = "Normal Q-Q Plots of Weekly Price Changes",
        x = "Theoretical Quantiles",
        y = "Sample Quantiles"
    ) +
    theme_minimal()
```



The Q-Q plots above display the distribution of weekly price changes for the Central Atlantic and Midwest regions against a theoretical normal distribution. The data points for both regions deviate significantly from the reference line in the tails, indicating heavy tails (leptokurtosis). This confirms that the data is not normally distributed and justifies the use of the robust Brown-Forsythe test instead of the standard F-test.

Statistical Hypotheses

We test the following hypotheses regarding the variances of weekly price changes (σ^2):

- **Null Hypothesis (H_0):** $\sigma^2_{Midwest} = \sigma^2_{CentralAtlantic}$ (The volatility is the same in both regions).
- **Alternative Hypothesis (H_A):** $\sigma^2_{Midwest} \neq \sigma^2_{CentralAtlantic}$ (The volatility differs between the regions).

Results

The code below conducts the Brown-Forsythe Levene's Test to compare the variances of the Midwest and Central Atlantic regions.

```
# Conduct Brown-Forsythe Levene's Test (center = median)
leveneTest(Log_Change ~ Region, data = reg_gas_logs, center = median)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value    Pr(>F)
## group    1  204.26 < 2.2e-16 ***
```

```
##       2648
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretation of Results: The output above displays the F-value and the p-value (Pr(>F)) for the test. As the p-value is less than 0.05, we reject the null hypothesis. Rejection of the null hypothesis confirms that the variances are statistically different, supporting the observation that the Midwest experiences higher volatility.

Conclusions

The results of the Brown-Forsythe test provide statistical evidence to conclude that there is a significant difference in the volatility of regular gasoline prices between the Midwest and Central Atlantic regions. Specifically, the Midwest experiences higher volatility (greater variance in weekly price changes) compared to the Central Atlantic.

In practical terms, this means that consumers and businesses in the Midwest face less predictable gas prices with larger week-to-week price changes than those in the Central Atlantic. This instability may be driven by the Midwest's reliance on regional refining capacity which can be more susceptible to local outages, compared to the Central Atlantic's access to broader supply chains.

Limitations and Future Work: A key limitation of this analysis is the assumption of independence. Since the data is a time series, the price changes likely exhibit autocorrelation (volatility clustering), where periods of high volatility are followed by high volatility. This violates the strict independence assumption and may affect the precision of the p-value. Future work could reduce this limitation by employing methods, such as GARCH models, which explicitly model volatility clustering.

Generalization: These conclusions are generalizable to the specific regions and time period studied (2000-2025). While they highlight a historical difference in volatility, they may not predict future volatility if there are major shifts in energy policy, infrastructure, or global oil market dynamics that fundamentally change regional supply chains.

References

Baghestani, Hamid, and Jörg Bley. "Do Directional Predictions of US Gasoline Prices Reveal Asymmetries?" Journal of Economics and Finance, vol. 44, 2020, pp. 348–60.

"East Coast and Gulf Coast Transportation Fuels Markets." U.S. Energy Information Administration, 3 Feb. 2016, https://www.eia.gov/analysis/transportationfuels/padd1n3/.

"Gas Prices." AAA Gas Prices, American Automobile Association, https://gasprices.aaa.com/. Accessed 9 Dec. 2025.

"Gasoline and Diesel Fuel Update." U.S. Energy Information Administration, https://www.eia.gov/petroleum/gasdiesel/. Accessed 9 Dec. 2025.

"Gasoline and Diesel Fuel Update: Processing Methods." U.S. Energy Information Administration, https://www.eia.gov/petroleum/gasdiesel/gas_proc-methods.php. Accessed 9 Dec. 2025.

Glen, Stephanie. "Brown-Forsythe Test: Definition." Statistics How To, https://www.statisticshowto.com/brown-forsythe-test/. Accessed 9 Dec. 2025.

"Midwest Refinery Outages Increase Regional Gasoline Prices." Today in Energy, U.S. Energy Information Administration, 22 Aug. 2024, https://www.eia.gov/todayinenergy/detail.php?id=62904.

Tappata, Mariano. "Rockets and Feathers: Understanding Asymmetric Pricing." RAND Journal of Economics, vol. 40, no. 4, 2009, pp. 673–87.

"Why California Usually Pays More at the Pump for Gasoline." Today in Energy, U.S. Energy Information Administration, 5 May 2025, https://www.eia.gov/todayinenergy/detail.php?id=65184.

Project Repository

The code and data for this project are available at: https://github.com/nemmo-ciccone/Regional-Gas-Price-Disparities

Generative AI Documentation

This document was prepared with the assistance of Google Gemini 2.5. The AI was used to help with syntax for creating R code chunks, data manipulation, and debugging. All code was reviewed and modified as necessary by the author. The AI did not contribute to the analysis, interpretation of results, or writing of the document.

Nemmo Ciccone, December 9th, 2025