

3080_final_project

Nemmo Ciccone

2025-11-02

R Markdown

Introduction

Gasoline prices are a vital economic indicator and a significant component of consumer budgets. While national average prices are widely reported, they obscure the substantial price differences and volatility experienced by consumers at a regional level. These regional variations are driven by distinct local supply chains, state and local taxes, and environmental regulations mandating specific fuel blends (e.g., reformulated vs. conventional gasoline).

This project investigates the statistical properties of U.S. gasoline prices on a regional basis. The analysis will use weekly price data from the U.S. Energy Information Administration (EIA) from 2003 to the present. The research aims to answer two primary questions:

Center: Is there a statistically significant difference in the overall mean weekly price of regular gasoline between distinct U.S. regions (e.g., the West Coast vs. the Gulf Coast)?

Spread: Is there a statistically significant difference in the price volatility (i.e., the standard deviation of weekly price changes) between U.S. regions (e.g., the Midwest vs. the East Coast)?

By answering these two statistical questions, this project aims to provide a quantitative foundation for exploring the broader empirical drivers of regional price disparities. Should statistically significant differences in mean price or volatility be found, the analysis will conclude by discussing the potential real-world implications, such as differing supply chain logistics, varying state tax burdens, regional differences in economic welfare, or distinct consumer search behaviors across these markets.

Data Summary

(a) & (b) Data Collection The data represents a sample, not a population. It is sourced from the U.S. Energy Information Administration's (EIA) Form EIA-878, "Motor Gasoline Price Survey." Every Monday, the EIA collects retail price data as of 8:00 a.m. from a representative sample of approximately 800 retail gasoline outlets across the United States.

This sample includes major national retailers, small chains, and independent owners. Data is collected via a multi-mode strategy, including telephone, email, text, fax, and web surveys. The final published figures are volume-weighted average prices for each region.

(c) Data Modifications The raw data will be modified in two ways for this analysis: Data Selection: The analysis will use the “Regular All Areas All Formulations” price series. This series is the most representative of the consumer market and is the most appropriate for making “apples-to-apples” comparisons between regions that have different fuel formulation requirements. Transformation (for Volatility): To analyze price volatility (spread) independent of long-term trends, a second dataset will be created by calculating the first difference ($Price_t - Price_{t-1}$) of the weekly price data. This transforms the dataset from “price” to “week-over-week change.”

(d) Potential Issues Non-Stationarity: The raw price data is non-stationary, meaning its mean and variance change over time (e.g., prices from 2003 are not comparable to prices from 2023). This trend could skew a simple calculation of variance, which is why the first-difference transformation is necessary to analyze volatility.

Missing Data: The dataset may contain NA or blank values for certain weeks or regions, which will need to be identified and omitted.

Methodology Changes: The EIA updated its sampling and estimation methodology in May 2018. This could create a structural break in the data, though its impact on comparing parameters between regions (which are sampled concurrently) is likely minimal.

(e) Appropriateness The EIA data is ideal for this research. Its weekly frequency, long historical time-span, and clear regional segmentation (PADDs) provide robust, high-integrity data to calculate and compare the long-term parameters of center (mean) and spread (variance)

Exploratory analysis

```
library(readxl)

# List all sheets in the Excel file
excel_sheets("eia_gas_data.xls")

## [1] "Contents"   "Data 1"      "Data 2"      "Data 3"      "Data 4"      "Data 5"
## [7] "Data 6"      "Data 7"      "Data 8"      "Data 9"      "Data 10"     "Data 11"
## [13] "Data 12"

# Read a specific sheet, we will use "Data 3", which is Regular All Areas All Formulations
# first 2 rows are headers to skip
reg_gas<-read_excel("eia_gas_data.xls", sheet = "Data 3", skip = 2)
colnames(reg_gas)

## [1] "Date"
## [2] "Weekly U.S. Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [3] "Weekly East Coast Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [4] "Weekly New England (PADD 1A) Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [5] "Weekly Central Atlantic (PADD 1B) Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [6] "Weekly Lower Atlantic (PADD 1C) Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [7] "Weekly Midwest Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [8] "Weekly Gulf Coast Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [9] "Weekly Rocky Mountain Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [10] "Weekly West Coast Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
```

```

## [11] "Weekly California Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [12] "Weekly Colorado Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [13] "Weekly Florida Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [14] "Weekly Massachusetts Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [15] "Weekly Minnesota Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [16] "Weekly New York Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [17] "Weekly Ohio Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [18] "Weekly Texas Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [19] "Weekly Washington Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [20] "Weekly Boston, MA Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [21] "Weekly Chicago Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [22] "Weekly Cleveland, OH Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [23] "Weekly Denver Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [24] "Weekly Houston Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [25] "Weekly Los Angeles Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [26] "Weekly Miami, FL Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [27] "Weekly New York City Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [28] "Weekly San Francisco Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"
## [29] "Weekly Seattle, WA Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)"

# Only keep the PADD Regions columns and Date
# Keep Date + PADD zone columns
padd_cols <- c(
  "Date",
  "Weekly New England (PADD 1A) Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)",
  "Weekly Central Atlantic (PADD 1B) Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)",
  "Weekly Lower Atlantic (PADD 1C) Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)",
  "Weekly Midwest Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)",
  "Weekly Gulf Coast Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)",
  "Weekly Rocky Mountain Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)",
  "Weekly West Coast Regular All Formulations Retail Gasoline Prices (Dollars per Gallon")
)

reg_gas_regions <- reg_gas[, padd_cols]

# Check new columns
colnames(reg_gas_regions) <- c('Date', 'New England', 'Central Atlantic', 'Lower Atlantic', 'Midwest',
                                'Gulf Coast', 'Rocky Mountain', 'West Coast')

head(reg_gas_regions)

## # A tibble: 6 x 8
##   Date           'New England' 'Central Atlantic' 'Lower Atlantic' 'Midwest'
##   <dttm>          <dbl>            <dbl>            <dbl>            <dbl>
## 1 1990-08-20 00:00:00      NA              NA              NA              NA
## 2 1990-08-27 00:00:00      NA              NA              NA              NA
## 3 1990-09-03 00:00:00      NA              NA              NA              NA
## 4 1990-09-10 00:00:00      NA              NA              NA              NA
## 5 1990-09-17 00:00:00      NA              NA              NA              NA
## 6 1990-09-24 00:00:00      NA              NA              NA              NA
## # i 3 more variables: 'Gulf Coast' <dbl>, 'Rocky Mountain' <dbl>,
## #   'West Coast' <dbl>

```

```

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

## # A tibble: 6 x 8
##   Date           'New England' 'Central Atlantic' 'Lower Atlantic' Midwest
##   <dttm>          <dbl>            <dbl>            <dbl>        <dbl>
## 1 2000-06-05 00:00:00      1.59            1.56            1.46        1.65
## 2 2000-06-12 00:00:00      1.62            1.59            1.49        1.80
## 3 2000-06-19 00:00:00      1.64            1.62            1.54        1.87
## 4 2000-06-26 00:00:00      1.66            1.64            1.54        1.79
## 5 2000-07-03 00:00:00      1.67            1.65            1.54        1.68
## 6 2000-07-10 00:00:00      1.65            1.65            1.53        1.59
## # i 3 more variables: 'Gulf Coast' <dbl>, 'Rocky Mountain' <dbl>,
## #   'West Coast' <dbl>

# show all missing values in the dataset
sapply(reg_gas_regions, function(x) sum(is.na(x)))

```

```

##       Date     New England Central Atlantic Lower Atlantic
##       0             0              0              0
## Midwest     Gulf Coast    Rocky Mountain West Coast
##       0             0              0              0

```

References

- Baghestani, Hamid, and Bley, Jorg (2020). *Do Directional Predictions of US Gasoline Prices Reveal Asymmetries?* Journal of Economics and Finance, 44, 348–360.
- Tappata, Mariano. (2009). *Rockets and Feathers: Understanding Asymmetric Pricing*. RAND Journal of Economics, 40(4), 673–687.
- EIA Gasoline Data
- EIA Methods
- AAA Gas Prices
- EIA Today in Energy