Concepts, Data Sources, and Techniques

Handbook of Energy Modeling Methods

Short-Term Energy Forecasting Petroleum Product Price Module

Independent Statistics & Analysis

www.eia.gov

U.S. Department of Energy

Washington, DC 20585

Last Updated 2020

This report was prepared by the U.S. Energy Information Administration (EIA), the statistical and analytical agency within the U.S. Department of Energy. By law, EIA's data, analyses, and forecasts are independent of approval by any other officer or employee of the United States Government. The views in this report therefore should not be construed as representing those of the U.S. Department of Energy or other federal agencies.

1. Introduction

The *Short-Term Energy Outlook* (STEO) Petroleum Product Prices Module forecasts the following data series:

- U.S. average wholesale prices of the following fuels:
  - Motor gasoline
  - Diesel fuel
  - Jet fuel
  - Residual fuel oil
- U.S. average retail prices of the following fuels:
  - Regular-grade motor gasoline
  - On-highway diesel fuel
  - Heating oil
- Petroleum Administration for Defense District (PADD) level retail prices of regular-grade motor gasoline
- Refiners average imported cost of crude oil
- Refiners average acquisition cost of crude oil

For each month of the STEO's release, the Petroleum Products Prices Module uses linear models to forecast these prices over a 13-to-24 month time horizon. Every January, the STEO forecast is extended through December of the following year.

The Petroleum Product Prices Module is organized as a series of linear regression equations (see Section B.1) and accounting identities used to derive each price forecast. In the linear regression models,

functions of the various prices are the dependent variables, and the independent variablies typically include items such as economic indicators, product inventories, regional demand shares, refinery activity, weather forecasts, and related fuel and commodity prices, among other variables.

2. Data Sources

With the exception of the heating oil retail price, the monthly price and volume data used in the Petroleum Product Prices Module appear in three EIA publications: the Petroleum Supply Monthly (*PSM*), Petroleum Marketing Monthly (*PMM*), and Weekly Petroleum Status Report (*WPSR).* Data in the *PSM* and *PMM* are published with a two-month lag. For example, the *PMM* released at the end of March will contain data through January. For any given data series, if *WPSR* data are available, STEO uses those data to fill in the intervening months between the last *PMM* data point and the present. If WPSR data are not available, STEO uses modeled results to fill in the intervening months.

Key data inputs include the following:

- Monthly average U.S. wholesale prices (refiner price for resale) of motor gasoline, number 2 diesel fuel, and number 2 fuel oil, and the refiner price to end users for kerosene-type jet fuel, from the *PMM*.
- PADD-level gasoline and diesel fuel retail prices (including taxes) and daily Brent crude oil spot prices, from the *WPSR*. Monthly average gasoline and diesel retail prices are calculated as simple averages of the weekly prices.
- The national average heating oil retail price, obtained from the Bureau of Labor Statistics' monthly consumer price survey.
- Monthly volume data from surveys of primary suppliers such as refineries, pipelines, and bulk terminals, as well as regional monthly gasoline consumption and end-of-month inventories, reported in the *PSM*. These data are used for weighting regional-level prices to arrive at the U.S. average price.

3. Linear Regression Models

3.1. Common Independent Variables

Prices for refined petroleum products are driven by multiple factors, including the price of crude oil (the main input cost for production), supply and demand conditions for a given fuel, seasonal factors, and the cost of transporting fuels from production hubs to demand centers. Some of these variables, e.g. crude oil prices, are observable and readily available for inclusion in our models, while other variables are more difficult to directly observe. EIA uses proxy variables to account for unobservable factors. For example, to approximate supply and demand conditions for a fuel, EIA uses the level of inventory in a given month compared to typical levels for that time of year.

To capture the primary effects driving fuel prices, many of the linear regression models described below include the following types of independent variables:

a) Monthly change in the Brent crude oil price (which drives wholesale product prices)

b) Monthly change in the the wholesale product price (which drives retail product prices)
c) Difference between the current month's consumption of a given petroleum product and average consumption of that product during a previous multi-year period
d) Difference between the current month's inventories of a given petroleum product and average inventories of that product during a previous multi-year period
e) Monthly dummy variables, to capture seasonal effects
f) A trend variable
g) A one-month lagged dependent variable, to capture the effect of the previous month's dependent variable on the current month's dependent variable

The monthly price changes (a and b) reflect the passthrough of prices from one part of the supply chain to the next, which typically occurs with a short lag. Positive consumption differences from a monthly average (c) would be expected to put upward pressure on prices, while below-average consumption would be expected to put downward pressure on prices. Similarly, positive inventory differences from a monthly average (d) would be expected to put downward pressure on prices, while below-average inventories often push prices up. Trend variables (f) and lagged dependent variables (g) capture price trends driven by evolving market conditions.

Also, rather than modeling an absolute price, for all of the series in the Petroleum Product Prices Module, EIA models a price *margin,* which is the difference between two prices (e.g., the gap between a wholesale and a retail price). After EIA forecasts a price margin, EIA adds the margin to an actual price forecast to generate the forecast for the other price in the margin (e.g., to forecast a retail price, we add the retail price margin to the forecast wholesale price).

3.2 Motor Gasoline Prices

Gasoline is sold in two different formulations: conventional and reformulated. Each of these formulations is produced at refineries as a product called *blendstock for oxygenate blending* (BOB). BOB can be blended with ethanol at different ratios to produce finished motor gasoline that is sold at retail gasoline stations. However, it is most common for both of the two formulations to be blended at a ratio of 90% BOB and 10% ethanol, a blend called E10. For both finished conventional and reformulated gasoline, there are also three grades of gasoline based on octane content: regular, mid-grade, and premium.

The motor gasoline price model begins with a forecast of the U.S. average motor gasoline wholesale price. We use this U.S. average wholesale price as a proxy for the motor gasoline spot price in each region. We model retail regional motor gasoline prices, including taxes, as a function of the wholesale price and regional inventories. Retail prices of regular grade gasoline, all formulations, which cover over 80 percent of all motor gasoline sold in the United States, are forecast for the five PADDs and the national average. A national average retail price for all grades of motor gasoline is also forecast.

Gasoline price time series are *long-memory* series; that is, they contain unit roots (see Section B.2). Because of the unit roots, the price series are transformed to stationary series by calculating price spreads. This transformation allows the price series to be more meaningfully modeled using linear regression. The dependent variable in the motor gasoline wholesale price equation is the difference

between the average U.S. motor gasoline wholesale price and the Brent crude oil spot price (in cents per gallon). For regional retail prices, the dependent variable is the difference between the PADD regular gasoline retail price and the U.S. average motor gasoline wholesale price.

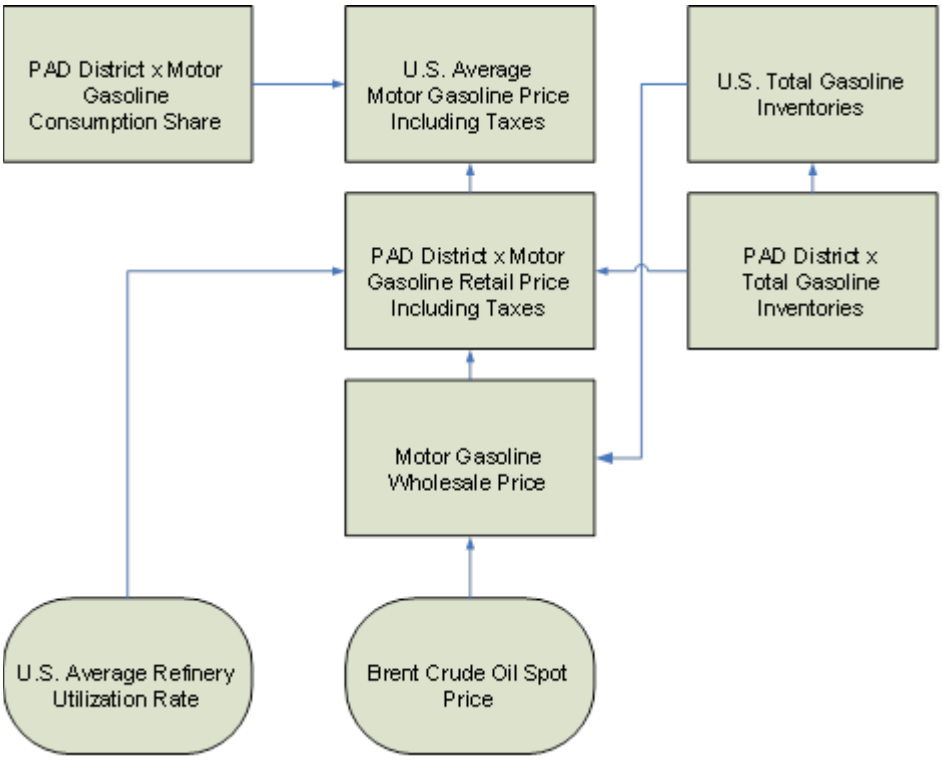
3.2.1. Motor gasoline wholesale prices

Because motor gasoline is a product of crude oil refining, EIA assumes the wholesale gasoline price is directly related to the marginal cost of crude oil. Because of the many different grades of crude oil, there are many measures of crude oil prices. STEO uses the Brent crude oil spot price in the regression model that forecasts the wholesale gasoline price. EIA research shows that motor gasoline wholesale prices are more strongly related to Brent crude oil prices than to West Texas Intermediate (WTI) crude oil prices (see EIA, What Drives U.S. Gasoline Prices?, October 30, 2014). EIA therefore uses the price difference between the average U.S. motor gasoline wholesale price and the average cost of Brent crude oil (converted from dollars per barrel to cents per gallon), referred to here as the *wholesale gasoline margin*, as the dependent variable for forecasting motor gasoline wholesale prices. The wholesale gasoline margin is estimated as a function of the following variables:

- The change in the spot price of Brent crude oil
- Inventory deviation from the previous four-year average
- A trend variable starting in 2003
- Monthly dummy variables
- A lagged dependent variable.

3.2.2. Motor gasoline retail prices including taxes

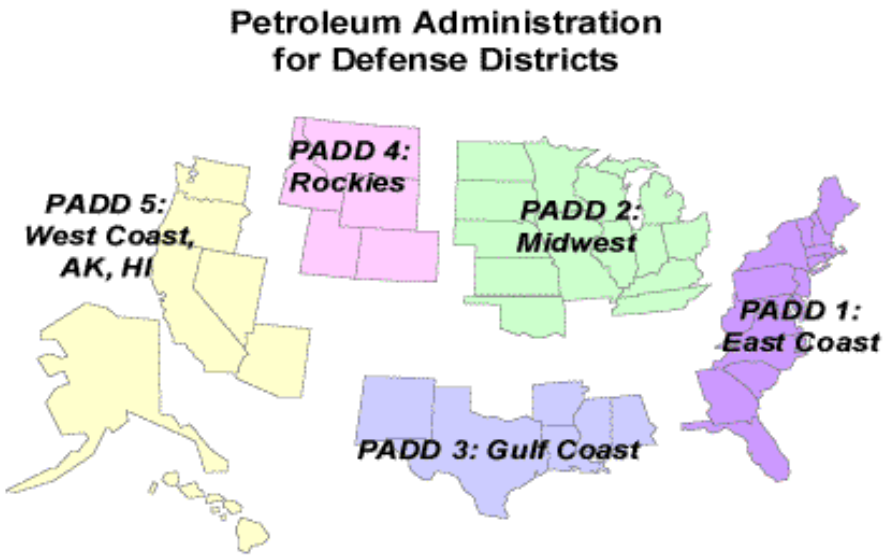
Figure 3.1. Regional Gasoline Retail Price Model



Regional motor gasoline retail prices for each PADD **x** are estimated as a function of the average U.S. motor gasoline wholesale price and regional gasoline inventories as illustrated in Figure 3.1. The model includes equations that estimate the regular-grade gasoline retail prices for the five PADDs (Figure 3.2). EIA uses the the retail margin, which is the difference between the average PADD-level motor gasoline retail price and the average U.S. gasoline wholesale price, as the dependent variable. In addition to capturing retail gasoline distribution costs and profits, the retail margins include all federal, state, and local taxes. For each PADD, the retail margin is estimated as a function of the following variables:

- The change in the gasoline wholesale price from the previous month
- The deviation in regional gasoline inventories from their prior four-year average,
- A trend variable beginning in 2003
- Monthly dummy variables
- A lagged dependent variable

Figure 3.2. Petroleum for Administration Defense Districts (PADDs)



3.2.3. Retail motor gasoline price, all grades

The regional motor gasoline retail prices are for regular-grade gasoline only, which makes up about 80% of the total motor gasoline market. Premium and mid-grade gasoline make up the remainder. The difference between the average price for all grades of gasoline and regular-grade gasoline has been fairly consistent, although growing slightly, over the past several years. EIA estimates the difference between the U.S. average price for all grades of gasoline and the U.S. average price of regular-grade gasoline (the *all-grade margin*) as a function of the following variables:

- The U.S. average regular-grade retail gasoline price level
- A trend variable beginning in 2009
- Monthly dummy variables
- A lagged dependent variable.

3.2.4. Motor gasoline consumption shares

Estimated regional gasoline consumption volumes are needed to weight regional regular-grade gasoline retail prices in the calculation of monthly average U.S. prices as well as quarterly and annual average regional and U.S. average prices. The regional consumption share is the total regional gasoline consumption divided by total U.S. gasoline consumption. EIA estimates the regional motor gasoline consumption shares from linear regression models with monthly dummy variables as the independent variables. Additional dummy variables may be included in individual models to control for observed outliers in the data series.

The error in the calculated average prices is expected to be much smaller than the error in the estimated consumption shares. For example, consider two regions, one with a price of $3.50 per gallon and the other with a price of $3.00 per gallon. Assume actual consumption is identical in the two regions but is

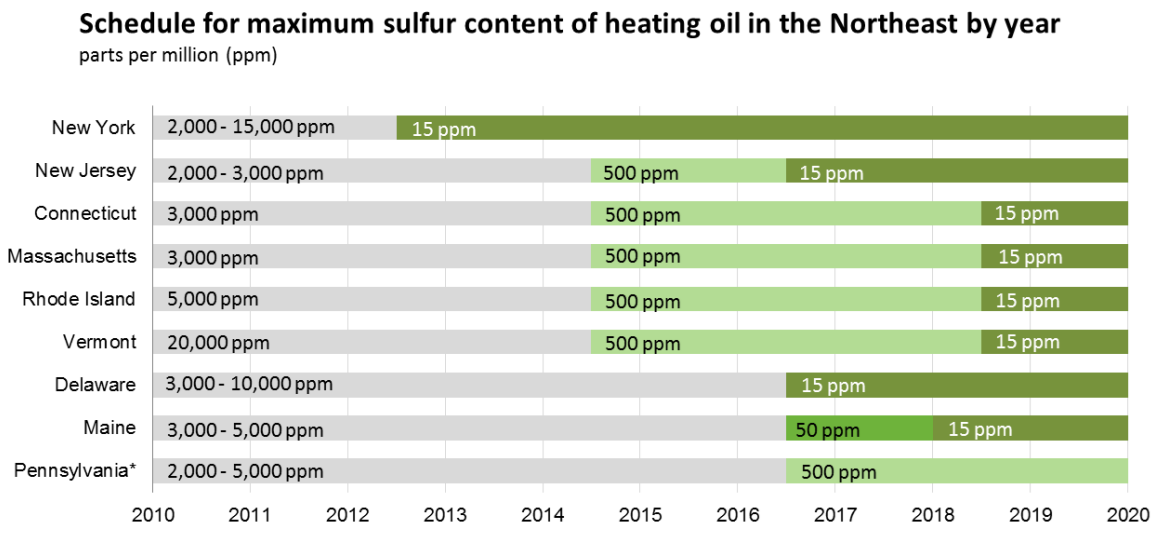
measured with 20 percent error in opposite directions: region A's measured consumption share is 60% and region B's is 40%, rather than 50% each. While the errors in the consumption shares are large, the error in the volume-weighted average price is much smaller: $3.30 per gallon when consumption is measured with error and $3.25 per gallon when consumption shares are correct (a 1.5% error).

3.3. Distillate Fuel Prices

Distillate fuel includes diesel fuels and fuel oils. Products known as No. 1, No. 2, and No. 4 diesel fuel are used in on-highway diesel engines, such as those in trucks and automobiles, as well as off-highway engines, such as those in railroad locomotives and agricultural machinery. Diesel fuel is primarily sold as ultra-low sulfur diesel (ULSD), which contains a maximum 15 parts per million (ppm) sulfur. ULSD accounted for about 96% of U.S. total distillate consumption in 2017.

Products known as No. 1, No. 2, and No. 4 fuel oils are used primarily for space heating and electric power generation. Distillate fuel is also distinguished by sulfur levels. Fuel oil, commonly referred to as heating oil, has different sulfur specifications depending on the state in which the fuel is sold (Figure 3.3). Although heating oil is used across the United States, its use is concentrated in the Northeast, where more than 80% of all U.S. heating oil was consumed in 2017.

Figure 3.3. Heating oil sulfur specifications in Northeast States.



Note:

Notes: Specifications change on July 1 of the years shown, with the exception of Maine's 15 ppm requirement, which changes on January 1, 2018.

* Philadelphia, Pennsylvania changed from 2,000 ppm to 15 ppm on July 1, 2015.

The petroleum product prices module includes average U.S. diesel fuel and heating oil wholesale prices (refiner prices for resale) and average U.S. diesel fuel and heating oil retail prices.

3.3.1. Diesel fuel oil wholesale price

Because distillate fuel is a product of crude oil refining, EIA assumes the wholesale price of distillate fuel is directly related to the marginal cost of crude oil. EIA selected the Brent crude oil spot price as the crude oil price that is most strongly related to distillate fuel wholesale prices.

The monthly price difference between the average U.S. wholesale diesel fuel price and the crude oil price (converted from dollars per barrel to cents per gallon), referred to as the *wholesale diesel margin*, is estimated as a function of the following variables:

- The monthly change in the spot price of Brent crude oil
- The difference between current-month distillate fuel consumption and the prior two-year average for the reference month
- The difference between beginning-of-month distillate fuel inventories and the prior four-year average inventories
- Monthly dummy variables
- A lagged dependent variable

3.3.2. Diesel fuel retail price including taxes

The retail margin, which is the difference between the U.S. average diesel fuel retail price and the diesel fuel wholesale price, captures retail diesel distribution costs and profits and includes all federal, state, and local taxes. It is estimated as a function of the following variables:

- The diesel fuel wholesale price change from the previous month
- A trend variable beginning in 2011
- Monthly dummy variables
- A lagged dependent variable

3.3.3. Heating oil wholesale price

Because heating oil and diesel are very similar fuels, their prices are affected by many of the same drivers, with heating oil prices generally following diesel prices, except for the effects of some market dynamics unique to heating oil. EIA therefore models the price difference between the U.S. average heating oil wholesale price and the U.S. average diesel fuel wholesale price (the *heating oil-diesel margin*). In general, we expect the heating oil wholesale price to be lower than the diesel wholesale price, because of the more relaxed heating oil product specifications, such as sulfur level. However, heating oil-diesel margins can turn positive during very cold winter months, primarily in the Northeast, where more than 80% of the households that use heating oil as their primary space-heating fuel are located.

The heating oil-diesel margin is estimated as a function of the following variables:

- The change in the diesel fuel wholesale price from the previous month
- The difference between beginning-of-month distillate fuel inventories and the prior four-year average inventories

- Heating degree day (HDD) deviations from normal in the Northeast census region (New England and Middle Atlantic census divisions), weighted by the number of households in each division that use heating oil as their primary space-heating fuel
- Monthly dummy variables
- A lagged dependent variable

3.3.4. Heating oil retail price

To estimate the U.S. average heating oil retail price, EIA models the difference between the retail and wholesale heating oil prices (the *heating oil retail margin*). In addition to capturing retail heating oil distribution costs and profits, this margin includes all federal, state, and local taxes. The heating oil retail margin is estimated as a function of the following variables:

- The change in the heating oil wholesale price from the previous month
- The difference between beginning-of-month distillate fuel inventories and the prior four-year average inventories
- U.S. Northeast HDD deviations from normal
- Monthly dummny variables
- A lagged dependent variable

3.4. Other Petroleum Product Prices

3.4.1. Jet fuel refiner price to end users

Jet fuel and diesel fuel are produced at similar distillation ranges within a refinery, and refineries have significant abilities to adjust production towards one fuel or the other. Because jet fuel and diesel are very similar fuels, their prices are affected by many of the same drivers, with jet fuel prices generally following diesel prices, except for the effects of some market dynamics unique to jet fuel. EIA therefore models the price difference between the U.S. average jet fuel wholesale price and the U.S. average diesel fuel wholesale price (the *jet fuel-diesel margin*). The world real gross domestic product is included as an explanatory variable to capture world economic growth as a proxy for global air travel and jet fuel demand.

The jet fuel-diesel margin is estimated as a function of the following variables:

- The change in the diesel fuel wholesale price from the previous month
- World real gross domestic product
- Monthly dummy variables
- A lagged dependent variable
- Residual fuel refiner price to end users

3.4.2. Residual fuel oil refiner price to end users

Residual fuel oil is a general classification for the heavier oils, known as No. 5 and No. 6 fuel oils, that remain after the distillate fuel oils and lighter hydrocarbons are distilled away in refinery operations. It

conforms to ASTM Specifications D 396 and D 975and Federal Specification VV-F-815C. No. 5 fuel oil, a residual fuel oil of medium viscosity, is also known as Navy Special and is defined in Military Specification MIL-F-859E, including Amendment 2 (NATO SymbolF-770). It is used in steam-powered vessels in government service and inshore power plants. No. 6 fuel oil includes Bunker C fuel oil and is used for the production of electric power, space heating, vessel bunkering, and various industrial purposes.

EIA estimates the difference between the average U.S. residual fuel price to end users and the cost of crude oil to U.S. refiners (the *residual fuel margin*). EIA includes total U.S. HDD in the model, because residual fuel use in the electric power sector typically increases during times of very cold weather, when natural gas that would otherwise be used for power generation is diverted to residential and commercial space heating use.

The residual fuel margin is estimated as a function of the following variables:

- The change in the cost of crude oil to U.S. refiners from the previous month
- U.S. average HDD
- A time trend beginning in 2003
- A time trend beginning in 2008
- Monthly dummy variables
- A lagged dependent variable