SHOPP’s EIA-877 Winter Heating Fuels Survey: Improving Data Quality

Presentation to
2016 State Heating Oil and Propane Program (SHOPP) Workshop
July 13, 2016 | Washington, D.C.

By
SHOPP Survey Team: David Dudley and Hallie Black
Overview

Survey Overview
- Purpose
- Who Uses SHOPP Data?
- Data Collection

Data Quality
- Individual Price Data
- Overall Price Data

Price Characteristics and Qualifiers
- Criteria for Identifying Correct Prices

Data Validation
- Data Issues
- Validation Process

Data Anomalies
- Outliers
- Countertrend Price Movements
- Patterning
- Human Error
- High-Impact Data

Ways to Improve Data Quality
- Weekly Data Collections
- Calls to Respondents
- Spot Errors in Reporting
Survey Overview

U.S. Energy Information Administration’s (EIA) State Heating Oil and Propane Program’s (SHOPP) EIA-877: Winter Heating Fuels Telephone Survey

Purpose

- Designed to collect data on state-level prices of residential propane and No. 2 heating oil during the heating season (October through March)

Who Uses Survey Data?

- Data used by state and federal governments, policymakers, industry analysts, press, and consumers of residential heating fuel
- Winter heating fuels price data continues to be one of the most requested datasets produced by EIA
Survey Overview – continued

Data collection

• EIA and states contact retail heating fuel outlets weekly via phone or email and obtain price data

• Data are due by noon EST on Tuesday, except when Monday is a holiday in which case data are due on Wednesday

• Data from some of the larger propane companies are provided directly to EIA, which may reduce the number of calls State Energy Offices (SEOs) have to make

• Prices entered into EIA’s Internet Data Collection system (IDC)

• Both EIA and SEOs have the responsibility to protect data at the respondent level
Data Quality

Data quality – a measure of the data’s ability to serve the survey’s purpose and end use

Individual Prices
- Accuracy – defined price characteristics
- Completeness – include/exclude qualifiers
- Consistency – same price category reported from week to week
- Appropriate Representation – true going-rate offered to the majority of residential customers that retail outlet services?

Overall Data
- Relevance – important to users
- Reliability – accessible and available when needed
Price Characteristics and Qualifiers
Price Characteristics and Qualifiers

Data survey collects Residential Propane and No. 2 Heating Oil prices

Residential customers – Individual customers or households using the fuel to heat their residences
  • “Residential Sales” do not include apartment buildings, multi-family dwellings, businesses, or institutions

Respondents should report data for the particular retail outlet being surveyed
  • Identify the outlet being targeted by stating which city the outlet is located
  • Location-specific information found within the IDC on “Main Form” screen
  • Prices can vary from outlet to outlet
Price Characteristics and Qualifiers - continued

Price based on a “leased” or “company-owned tank”
- Majority of U.S. heating fuel customers do not own their tanks
- Customer-owned tank prices tend to be less than the “market price”

Price should exclude
- Taxes
- Discounts
- Premiums paid for small or large volume purchases
- Surcharges for customers living outside the normal delivery area

Collect the market price
- “Charge”
- “Full-Credit”
- “Going Rate”
Price Characteristics and Qualifiers: Identifying appropriate representation of data

Price per gallon sometimes based on annual consumption (More common with propane companies)

Annual consumption rate for average residential customer

Identify a tier-level or product code that corresponds with that volume

Avoid asking for usage-per-season ranges that are round numbers (e.g., 1,000 gallons)

Make note within IDC
Data Validation
Excessive Data Issues can Strain Validation Process

During a heating season, on any given week, there are data issues.
- Inherent nature of the data
- Different types of issues
- Amount varies from week to week

Minimize the amount of data issues encountered
- Low Hanging Fruit – Data issues that are easily identifiable or require little effort to correct

Validation Process
- Identify the issues
- Determine the actual price
- Substantiate the information
Data Validation: Propane Validations 2015-2016

PROPANE DATA VALIDATIONS

- Propane Flagged
- Propane Adjusted

Weeks: Week #1, Week #2, Week #3, Week #4, Week #5, Week #6, Week #7, Week #8, Week #9, Week #10, Week #11, Week #12, Week #13, Week #14, Week #15, Week #16, Week #17, Week #18, Week #19, Week #20, Week #21, Week #22, Week #23, Week #24, Week #25, Week #26

Data validation: Propane Validations 2015-2016
Data Validation: No. 2 Distillate Validations 2015-2016

**HEATING OIL DATA VALIDATIONS**

- Heating Oil Flagged
- Heating Oil Adjusted

Weeks #1 to #26
Data Anomalies
Data Anomalies: Outliers

Outliers – Values that differ greatly from the majority set of prices observed within the standard deviation.

Minimum and Maximum values acknowledged:
• Not always flagged

Example #1:
$2.390, $2.589, $3.131, $3.139, $3.147, $3.158, $3.219, $3.628, $4.009

Example #2:

Normal Distribution of data
Data Anomalies: Countertrend Price Movements

Price that is moving contrary to the established majority set of prices in a state or shared market

Either direction (showing increase or decrease in price)

Cause of price movement not primary focus

Direction of movement is what is important during analysis

Example:

<table>
<thead>
<tr>
<th></th>
<th>Wk #1</th>
<th>Wk #2</th>
<th>Wk #3</th>
<th>Wk #4</th>
<th>Wk #5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company A</td>
<td>$1.890</td>
<td>$1.890</td>
<td>$1.930</td>
<td>$1.930</td>
<td>$1.930</td>
</tr>
<tr>
<td>Company B</td>
<td>$2.049</td>
<td>$2.079</td>
<td>$2.109</td>
<td>$2.109</td>
<td>$1.909</td>
</tr>
<tr>
<td>Company C</td>
<td>$1.787</td>
<td>$1.807</td>
<td>$1.857</td>
<td>$1.897</td>
<td>$1.937</td>
</tr>
</tbody>
</table>
Data Anomalies: Analyzing Patterns

We also look at respondent’s reporting histories to identify certain patterns (marketing behavior) exhibited over time.

Respondent could have a history of reporting:
• high or low prices
• prices near the mean or median

Individual respondent’s reporting histories
• History of even or odd pricing:
  • (e.g., $2.000, $1.800, $1.500) or (e.g., $2.479, $2.579, $2.629)
• History of reporting the same price for extended periods of time
• Clues that aid the validation process
Data Anomalies: Breaks in Pattern

Example #1: $2.378, $2.378, $2.478, $2.478, $2.478, $2.478, $2.478, $2.478, $2.678

Not always incorrect!

Should be questioned regardless
Possible causes:
response to changing market conditions, changing wholesale costs, or competitive strategy

Cash prices: Most pervasive data issue!

Example #2

Company A: $2.378, $2.378, $2.478, $2.478, $2.478, $2.478, $2.478, $2.378, $2.478
Company B: $2.378, $2.378, $2.478, $2.478, $2.478, $2.478, $2.478, $2.428, $2.478
Data Anomalies: Human Error

Transcription: when wrong key is pressed

Example #3a:
$2.378, $2.408, $2.408, $2.438, $2.478, $2.478, $2.478, $2.478, $2.278

Transposing numbers

Example #3b:
$2.378, $2.408, $2.408, $2.438, $2.478, $2.478, $2.478, $2.478, $2.784

Other causes for observed breaks in data patterns:
• Price for wrong product-type reported
  (e.g., kerosene, propane, heating oil)
• Request for wrong product code or tier-level
Data Anomalies: High Impact Data

The survey uses a weighted sample

Each retail outlet is assigned a weight and volume
• Representative of market-share

One respondent may have a different weight and volume assigned to them then another

Each price may have a different impact on state, PADD, and overall aggregates

We review individual prices that have a high impact on the state, PADD, and overall aggregates
Ways to Improve Data Quality
Ways to Improve Data Quality: Prepare before placing Call / Email

Before a call is made: **Review previously reported data**
- Utilize the IDC
- Company History report is helpful
- Simply go back a week or two to review the prices

Weekly Price Comparison
*The single most important action one can take to improve data quality!*

Example #1  $2.539, $2.579, $1.899

Look for consistencies or noticeable trends that may be emerging
Ways to Improve Data Quality: Take Time on Calls

**Take your time** when corresponding with respondents
- Don’t rush off the phone
- Calls can still be completed within a reasonable amount of time
- Same principle applies to emails

**Pay attention** to the data being reported
- Ask the respondent questions about the data being reported
- If collecting data via email, reply to respondent w/ well formulated questions

**Report** difficult respondents who are noncompliant
- Contact the SHOPP Operator
- We will work together to resolve issues
- Streamline a process for collecting future data
Ways to Improve Data Quality: Catch Errors as They Occur

Common errors made by respondents
  • Quote the wrong price from company’s price list
  • Quote the wrong product code or tier-level price

Catch errors as they occur

Verbalize the price difference by restating it in the form of a question
  • (e.g., $3.590? Did your residential propane really increase by $0.20 from last week to this week?)

  • (e.g., $3.490? That is a decrease of $0.10 from last week. This is a counter-trend movement compared with what we are seeing reported elsewhere within the state. Did your heating oil really decrease in price by $0.10?)

Asking questions at time of initial contact reduces the amount of respondent burden / disruption to their daily operations
For More Information: Data Publications

U.S. Energy Information Administration home page [www.eia.gov](http://www.eia.gov)

This Week in Petroleum (TWIP) [http://www.eia.gov/oog/info/twip/twip.asp](http://www.eia.gov/oog/info/twip/twip.asp)

Heating Oil and Propane Update (HOPU) [http://www.eia.gov/petroleum/heatingoilpropane/](http://www.eia.gov/petroleum/heatingoilpropane/)

Short-Term Energy Outlook (STEO) [http://www.eia.gov/forecasts/steo/](http://www.eia.gov/forecasts/steo/)


Department of Energy’s North East Home Heating Oil Reserve (NEHHOR) [http://www.energy.gov/fe/services/petroleum-reserves/heating-oil-reserve](http://www.energy.gov/fe/services/petroleum-reserves/heating-oil-reserve)
SHIPP’s EIA-877 Winter Heating Fuels Telephone Survey Contact Information

PBS Survey Manager
Marcela Rourk
(202) 586-4412
Marcela.Rourk@eia.gov

Survey Email Account
SHOPPSurvey
SHOPPSurvey@eia.gov

Survey Operator
David Dudley
(202) 586-9240

Survey Support
Hallie Black
(202) 586-7921