

# **Use of Multimode Data Collection and Dissemination Methods in Quick Turnaround Surveys**

Macro International Inc.  
**Benita O'Colmain**

Energy Information Administration  
**Paula Weir**

# Overview

- Survey background and description
- Uses for survey data
- Data collection methods
- Integration and processing of data
- Data dissemination methods
- Management tools
- Future directions

2

Overview of what will be covered in the presentation. The first part will start with some background and a brief description of the attributes of the surveys and the uses for the survey data. Then I'll describe the data collection methods and how the data is integrated and processed followed by the ways in which the data is disseminated. Second part of the presentation will cover the tools that we use to track the survey progress in real-time and to track the survey performance over time. Lastly I'll mention a few of the things we're currently working on as enhancements to the data collection.

## Survey Background

- Weekly EIA-878 gasoline and EIA-888 diesel surveys
- Gasoline -1990 Iraqi invasion of Kuwait
- Diesel Fuel -1993 EPA regulations requiring use of low sulfur diesel

3

EIA's weekly motor gasoline survey was initiated in August 1990 when the Iraqi invasion of Kuwait and the resulting rise in gasoline prices led to a need for frequent monitoring of motor gasoline prices at the national level. The weekly diesel fuel price survey started about four years later in response to the introduction of EPA regulations requiring low sulphur diesel as a standard for all on-highway diesel fuel. The principal objective of these surveys was to collect, process and release the data to Congress and a variety of users in a rapid quick-turnaround mode. The surveys were originally designed to provide estimates of only one price for each fuel type at the national level. Over time they have expanded to include estimates for additional types and grades of fuel at several geographic levels.

## Description of Surveys

- Retail prices at the pump for three grades of gasoline and diesel fuel
- Fixed set of respondents ~ 800 gas stations and 325 truck stops
- Conducted each Monday – national, regional, state and city level estimates released by 5:00 p.m.

4

Today, the gasoline survey collects retail prices at the pump for regular, midgrade, and premium conventional and reformulated gasoline. The diesel survey collects prices for low sulfur diesel fuel and in January of this year, we began collecting prices for ultra low sulfur diesel fuel which has recently been introduced in response to new EPA requirements for cleaner emission standards. The sample of survey respondents is fixed from week-to-week and consists of approximately 800 gasoline stations for the gas survey and 325 truckstops for the diesel survey. Respondents may report for only one or multiple locations, depending on the size and geographic coverage of the company, and their reporting preference. The samples for the surveys are drawn independently, however, there are some respondents who report prices for both surveys. Data collection begins each Monday at 8:00 a.m. and is closed by 3:00 p.m. National, regional, state and city level estimates are released by EIA by 5:00 p.m. on the same day.

## Uses for Survey Data

- Government: Congress, Federal, State and Local
- Media: Television, radio, newspapers, syndicated wire services
- Industry: Commercial firms, Consulting and Financial firms, Research and Academic community
- Public

5

The data is used at all levels of government and in Congress with numerous Senate and House staff inquiries during the year. The data appears in the media, on national and local television news and radio stations and newspapers and is distributed via syndicated wire services such as Reuters, Bloomberg News, Dow Jones and the Associated Press. In Industry the data is used by commercial, consulting and financial firms as well as by researchers and academia. Recently, the Chicago Board of Trade has begun using the data to set futures prices for Trade futures contracts and for Hedgefund derivations. The data is accessed frequently by the general public.

## Data Collection Methods

- Primary data collection method is CATI
- E-mail, Fax, Internet, 800 call-in line
- Alternate methods used by respondents reporting for several stations
- Non-standard format for e-mail and fax submissions

6

The primary data collection method for both surveys is Computer Assisted Telephone interviewing. Responses are also collected via e-mail, fax, the internet and an 800 line is available for respondents electing to call in their prices. The alternate methods are primarily used by larger companies that report for several locations. The primary goal in providing alternative modes of data submission for the larger respondents is to promote ease and convenience for them, which in turn reduces their burden and increases the likelihood that they will respond consistently each week. As a result, e-mail and fax responses are received in many different formats.

## Data Collection Modes

	<b>Contacts</b>	<b>Locations</b>
<b>Gasoline Survey</b>		
Telephone	523	529
Fax	31	142
E-mail	16	124
Web	1	2
<b>Diesel Survey</b>		
Telephone	139	145
Fax	7	65
E-mail	6	26
Web	4	94

7

This slide shows the distribution by data collection mode for each of the surveys. For the gasoline survey, fax and e-mail and the web account for about 20 percent of reported prices, while for the diesel survey, they account for almost 60% of reported prices. Because a small number of respondents account for a significant proportion of collected prices, it is critical that responses from these respondents are received each week.

## Integration of Data

- All data integrated via CATI system
- Concurrent data collection via telephone and alternative methods
- Respondents are classified by mode of data collection
- Prices collected via fax, e-mail and internet are entered via CATI screens

8

The data received through all collection modes are integrated via the CATI system. Throughout the day, the data is collected concurrently via the telephone and the alternative methods. Respondents are classified by the mode of response and assigned to separate work stations based on this classification. All prices collected via fax, e-mail and the internet are entered via CATI screens and are processed through the same editing criteria as if the prices were collected over the phone.

## Processing of Data

- Front-end – current sample and prior week's prices loaded
- Mid-day – provides check of response rates and data quality
- Back-end – SAS programs for quality control, imputation and aggregation
- Transmission of data to EIA

9

The sequence of data processing starts with the front-end loading of the current sample and prior week's prices on Monday morning. The data is processed at mid-day and reports are generated which provide response rates by collection mode and several data quality measures. At the back-end, a series of SAS programs are run for quality control checks, imputation, aggregation, and calculation of standard errors and coefficients of variation. The data is then transmitted to EIA by 2:45 for the diesel survey and by 3:45 for the gasoline survey.

## Processing of Data – EIA

- Data imported into software systems
- Additional quality control programs to validate data
  - Identify outliers and assess impact
- Generate WEB, EPUB reports
- Distribute data and reports
  - Collection and Dissemination Division (CDD)
  - List-serve subscribers
  - Internally for analysis and comparison

10

At EIA, the data is imported into a set of software systems that provide additional quality control and validation checks. Outliers are identified and reviewed and their impact on the aggregate prices is assessed. A series of reports is generated in formats for distribution via the internet and other channels. The data is then forwarded to EIA's collection and dissemination division, to list-serve subscribers and internally for analysis and comparison with outside data sources.

## Data Dissemination

- WEB page reports [www.eia.doe.gov](http://www.eia.doe.gov)
  - Gasoline and Diesel Fuel Update
  - Petroleum Navigator
  - This Week in Petroleum (TWIP) electronic newsletter
  - Weekly Petroleum Supply Report (WPSR)
    - Via Web and limited hard copies
- Hot-line
- List-serve e-mails
- Wireless list-serve (brief- diesel only)

11

The data is disseminated in several ways. Over the internet, the data is accessible via the EIA website in a series of reports of i

## Management Tools - Overview

- Core requirement is the dissemination of accurate and timely prices each week
- Tools used during data collection to track survey progress and data quality
- Performance metrics developed to measure two dimensions over time:
  - data quality indicators: accuracy, response rates, and imputation for non-response
  - survey performance indicators: time, efficiency, cost, and data security

12

The rest of the presentation covers the tools that are used to manage the surveys. As was mentioned earlier, the core requirement for the surveys is the dissemination of accurate and timely prices each week. We have a set of management tools that are used during the data collection process and another set of performance indicators that are used to provide cumulative measures of data quality and survey performance over time.

## Tracking Survey Progress

- Early morning report – results of gasoline and diesel fuel price models
- Response rates continually monitored via call center
- Respondents reporting for four or more stations and respondents reporting via fax or e-mail closely monitored
- Mid-day check:
  - responses failing edit re-check criteria (outliers) are moved to re-check study
  - non-response is evaluated (back-up contacts)

13

# Data Re-checks

Price re-checks during CATI interview	
Small Difference	Prices that are more than <b>3 percent</b> different from the price provided during the previous week are verified. If the price given is incorrect, the interviewer asks for the correct price. If the price given is correct, the interviewer continues.
Large Difference	Prices that are more than <b>5 percent</b> different from the price provided during the previous week are verified. If the price given is incorrect, the interviewer asks for the correct price. If the price given is correct, the interviewer asks why there was a large difference, records the answer, and then continues.
Price Maximum	Prices higher than <b>5 dollars</b> are not accepted by the CATI system.
Price Minimum	Prices lower than <b>50 cents</b> are not accepted by the CATI system.
Prices flagged for re-check study	
Percent Difference	A price is marked as bogus when the current week's price is different by <b>12 percent or more</b> than the previous week's price. Differences are marked regardless of whether the change is in the negative or positive direction.
Flagged Prices	Prices are flagged and treated as bogus when the price for a lower grade of gasoline is equal to or greater than the price for a higher grade of gasoline.
Imputation Rule	A price different by <b>24 percent or more</b> from the previous week is treated as erroneous and replaced with an imputed price.

## Performance Metrics – Data Quality

Contacts Attempted	The number of individuals that provide prices. An individual contact may provide prices for more than one station.
Stations Attempted	The number of stations for which prices were requested.
Stations Completed	The number of stations that provided prices. (Does not include stations for which prices were imputed.)
<b>Station Response Rate</b>	<b>The percentage of stations that provided prices.</b>
Prices Asked	The total number of prices requested from all stations (3 prices per station)
Prices Received	The total number of prices provided from all stations (3 prices per station).
Prices Imputed	The number of prices that were imputed; the contact may have been unable to provide prices or refused to provide prices.
<b>Imputation Rate</b>	<b>The percentage of prices that were imputed.</b>
Prices Rechecked	The number of prices that were more than the pre-set parameter for cent difference from the price provided during the previous week.
<b>Recheck Rate</b>	<b>The percentage of prices that were more than the pre-set parameter for percentage difference from the price provided during the previous week.</b>
Recheck Corrections	The number of prices that were corrected after the interviewer mentions the price was more than the pre-set parameter for percentage difference from the price provided during the previous week.
<b>Recheck Correction Rate</b>	<b>The percentage of prices that were corrected after the interviewer mentions the price was more than the pre-set parameter for percentage difference from the price provided during the previous week.</b>

15

## Performance Metrics – Program Performance

<b>EIA Weekly Price Survey Program Performance Statistics</b>		
January 1, 2006 through December 31, 2006	Target	Actual
Percent of Weeks Data File Delivered on Time	>= 98%	98%
Percent of Weeks EIA Data Release Delayed	<= 2%	1%
Number Incorrect Prices Found After Delivery	0	0
Actual Cost/Projected Cost	<= 1.00	1.00
Percent Staff CIPSEA Trained	100%	100%
Number of Security Breaches	0	0

16

The program performance metrics are broader in scope and evaluate the survey performance in terms of timeliness, cost, efficiency, and data security.

CIPSEA – Confidential Information Protection and Statistical Efficiency Act

## **Future Directions**

- Evaluation of CATI re-check rate has lead to improved methodology for price rechecks based on expected regional differences
- Development of a standardized format for fax and e-mail submissions