

## Using Models to Detect Outliers in Refinery Data\*

Phillip Tseng, SMG, [phillip.tseng@eia.doe.gov](mailto:phillip.tseng@eia.doe.gov), 202-287-1600,  
Lawrence Stroud, SMG, [lawrence.stroud@eia.doe.gov](mailto:lawrence.stroud@eia.doe.gov), 202-287-1722  
October 5, 2006

### **Abstract**

Refinery operations in the United States provide useful information on the supply of petroleum products. The EIA Form-810 provides a very comprehensive set of data on inputs, outputs, and changes in inventory of all refineries in the U.S. In recent years, Form-810 data receives special attention by stakeholders as it provides useful information on the trends in demand for major products and can have profound implications on prices of crude oil and major petroleum products.

The goal of this paper is to review the refinery input and output data system that reports the demand for petroleum products, to identify possible data inconsistencies, and to make adjustments if necessary to the time series demand data. Several simple regression models will be used to examine the relationship between refinery gain and production of major products. The statistical relationships developed in these models can be used to identify potential outliers and to help data analysts verify data accuracy with respondents. The end product will be a methodology and a computer program that EIA may use to detect outliers and improve data quality.

The Energy Information Administration (EIA) defines demand as product supplied, which measures the disappearance of products from primary sources, i.e., refineries, natural gas processing plants, blending plants, pipelines, and bulk terminals. In general, product supplied of each product in any given period is computed as: field production, plus refinery production, plus imports, plus unaccounted for crude oil, minus stock change, minus crude oil losses, minus refinery inputs, minus exports. Refinery production is the largest component of product supplied and gasoline dominates U.S. refinery production, therefore, this paper will focus only on refinery operations of gasoline and its effects on refinery gain. Specifically, the focus of this paper includes inputs to refineries, production of gasoline, and refinery processing gain. Historically, refinery volumetric gain ranges from 4 to 7 percent and is closely related to gasoline production. In 2005, refinery production averages about 17.7 million barrels per day and a one percent increase in refinery gain amounts to more than 170 thousand barrels per day additional production or about 62 million barrels per year.

\*This is a working document prepared by the Energy Information Administration (EIA) in order to solicit advice and comment on statistical matters from the American Statistical Association Committee on Energy Statistics. This topic will be discussed at EIA's fall 2006, meeting with the Committee to be held October 5 and 6, 2006.

