COMBINING DATA TO PRODUCE TIMELY ESTIMATES

Nancy C. Hassett and Pedro J. Saavedra, Macro International, Paula Weir, U.S. Department of Energy
Nancy C. Hassett, Macro International, 11785 Beltsville Dr., Calverton, MD 20705

KEY WORDS: Sample, Petroleum, Time Series

The Energy Information Administration (EIA) publishes prices for crude oil and petroleum products in the Petroleum Marketing Monthly. Crude oil prices are obtained from three EIA surveys. Petroleum product prices are based on the EIA-782 survey, a monthly survey of refiners, resellers, and retailers of various petroleum products. Due to the nature of the surveys and the required data processing time, results are not available until over two months after the reference period. Alternate price series and methodologies are examined as a means of producing more timely price estimates at the national and regional levels. One such method is the use of time series models using available prices from a variety of sources.

Estimates are desired at both the wholesale and retail level for several petroleum products. This paper will focus primarily on wholesale regular unleaded gasoline prices. Forecasts are required at both the national level and the Petroleum Administration for Defense District (PADD) level[1].

Background

The EIA-782 survey, which is conducted by mail, collects data at the company level on petroleum product sales to end users and resellers. The survey has two components: the EIA-782A, a census of refiners and gas plant operators, which accounts for a large percentage of petroleum products sales volume, and the EIA-782B. The EIA-782B survey is mailed to a sample of resellers/retailers who have been selected according to a complex sampling design involving multiple stratifications. The sample design is based on eight target products and estimates are produced by state/product combination. The sample is redrawn each year and rotated with 50 percent overlap each year.

Data Sources/Availability

The development of the forecasting models is subject to constraints on data availability, including the limitations of historical time series (e.g., length and accuracy), and processing system and publication system deadlines. Spot market prices are considered to be potential predictors of wholesale prices. However, spot price series may not be available at all, such as for midgrade gasoline, or the data series may contain only two to three years of data, which is not sufficient for the estimation of time series models. However, there are advantages to using spot prices as a predictor. Whereas EIA-782 prices are not available until at least two months after the reference month, daily spot prices are available on a real-time basis. The average monthly spot price is calculated as the average of the daily spot prices for that month.

ARIMA Models

A Box-Jenkins time series approach was selected for modeling. A univariate Box-Jenkins model may be expressed as:

$$Y_t = \mu + \frac{\theta(B)}{\phi(B)} a_t$$

where:

μ = mean
B = backshift operator: $B(X_t) = X_{t-1}$
$\Theta(B)$ = moving-average operator
$\Phi(B)$ = autoregressive operator
a_t = random error

Transfer function models, ARIMA models using input data series, were utilized for the prediction of wholesale regular unleaded gasoline prices. Transfer function models are appropriate when forecasts of a data series (EIA-782) are dependent on both past values of that series as well as current and historical values of an independent data series (e.g., spot prices). The fitting of a transfer function model requires the determination of the functional form relating the two data series (the transfer function) and the fitting of a noise model to the remaining error series.

A transfer function model may be expressed as:

$$Y_t = \mu + \sum \frac{\omega_i(B)}{\delta_i(B)} B^k X_{i,t} + \frac{\theta(B)}{\phi(B)} a_t$$

where:

$X_{i,t}$ = ith predictor time series

w_i = transfer function model numerator polynomial (analogous to moving-average operator)

δ_i = transfer function model denominator polynomial (analogous to autoregressive operator)

Wholesale Regular Unleaded Gasoline

Daily spot price series from New York Harbor, the Gulf Coast, and Los Angeles were aggregated to the monthly level for direct comparison to the reported monthly EIA wholesale prices. These price series at the national level are shown in Exhibit 1. Both the New York Harbor and Gulf Coast spot price series were available back to 1986. However, the Los Angeles spot price was available back only to late 1990.

Transfer function models were developed using the SAS ARIMA procedure. The price series were differenced to achieve stationarity of the error series. Models were developed at the national level and for each of the five PADD's, using the spot price series geographically aligned with that region. The models are shown in Exhibit 2. A plot of the forecasts with 95 percent confidence intervals is shown in Exhibit 3 for the national level. Similar results were obtained at the PADD level.

Criteria for evaluating and comparing models

Models were checked for consistency over time. After fitting a model to the full data (June 1986 through January 1995), the model was refit using only data through the end of 1993. The parameter estimates were compared between the full data model and the reduced data model. If the estimates were not stable, the model was rejected and the model identification and estimation process was repeated until a consistent model was obtained. In addition, both one-month ahead and two-month ahead forecasts for 1994 were compared to actual reported values as a means of measuring the adequacy of the models. The differences between the predicted and actual prices for 1994 are shown in Exhibit 4.

Results

The national and PADD level models were extremely sensitive to volatility in the spot price. When spot prices were relatively stable, prices were forecast within a couple of cents. When the spot price changed by about 15 percent, the prices were underestimated by as much as 9 cents. Additional predictor variables were considered in order to produce more accurate forecasts. Monthly ending gasoline stocks were obtained back to 1986. ARIMA models were developed using both spot prices and ending stocks as predictors. The models produced slightly better forecasts but were still extremely sensitive to large changes in the spot price.

Rack prices, associated with wholesale truckload sales at a terminal, have been suggested as an alternative predictor series. When this data becomes available, new models will be estimated using these prices.

Once improved PADD-level models are obtained, it will be possible to take a weighted average of the individual PADD price forecasts as an estimate of the national price. This estimate can then be compared to the estimate produced by a national-level forecasting model.

---

[1]PADD 1:

Subdistrict 1A: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont.

Subdistrict 1B: Delaware, District of Columbia, Maryland, New Jersey, New York, Pennsylvania.

Subdistrict 1C: Florida, Georgia, North Carolina, South Carolina, Virginia, West Virginia.

PADD 2: Illinois, Indiana, Iowa, Kansas, Kentucky, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, Oklahoma, South Dakota, Tennessee, Wisconsin.

PADD 3: Alabama, Arkansas, Louisiana, Mississippi, New Mexico, Texas.

PADD 4: Colorado, Idaho, Montana, Utah, Wyoming.

PADD 5: Alaska, Arizona, California, Hawaii, Nevada, Oregon, Washington.

EXHIBIT 1

Wholesale Regular Unleaded Gasoline Price Series

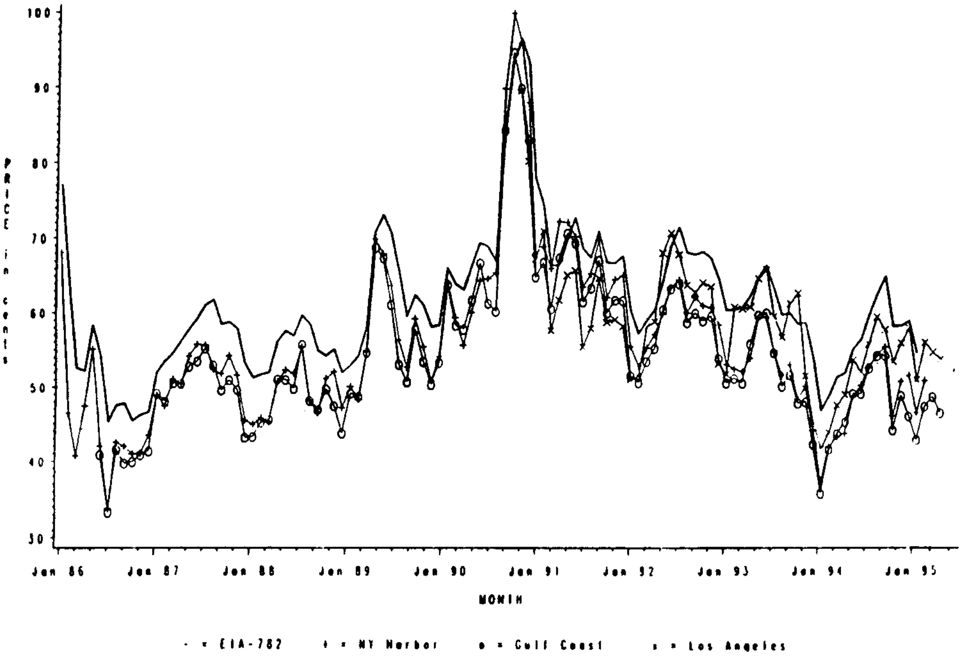

EXHIBIT 2

NATIONAL AND PADD-LEVEL ARIMA TRANSFER FUNCTION MODELS FOR WHOLESALE UNLEADED REGULAR GASOLINE

U.S.:	Y'_t	$= (0.10-0.16B)(1+0.64B^2)X'_t + (0.56+0.34B)(1-0.21B^2)W'_t +$ $(0.02+0.09B)(1+0.57B^2)Z'_t + e_t$
PADD 1:	Y'_t	$= (0.64+0.40B-0.09B^3)(1-0.29B+0.35B^2)X'_t + (1+0.21B^2)^{-1}e_t$
PADD 2:	Y'_t	$= (0.81-0.12B)/(1-0.41B)W'_t + (1+0.34B+0.32B^2)^{-1}e_t$
PADD 3:	Y'_t	$= (0.67+0.28B)/(1-0.11B^2)W'_t + (1+0.80B)(1+1.20B+0.45B^2)^{-1}e_t$
PADD 4:	Y'_t	$= (0.41+0.30B)W'_t + (0.09+0.30B)/(1-0.42B^2)Z'_t + e_t$
PADD 5:	Y'_t	$= (0.41+0.28B)/(1-0.35B)Z'_t + (1-0.61B+0.43B^2)^{-1}e_t$

where:

Y'_t	= difference in EIA-782 price $= (Y_t - Y_{t-1})$
X'_t	= difference in average monthly New York Harbor spot price for wholesale unleaded gasoline $= (X_t - X_{t-1})$
W'_t	= difference in average monthly Gulf Coast spot price for wholesale unleaded gasoline $= (W_t - W_{t-1})$
Z'_t	= difference in average monthly Los Angeles spot price for wholesale unleaded gasoline $= (Z_t - Z_{t-1})$
B	= Backshift operator such that $B(X_t) = X_{t-1}$

EXHIBIT 3

Forecasts for U.S. Wholesale Regular Unleaded Gasoline Prices with 95% Confidence Intervals

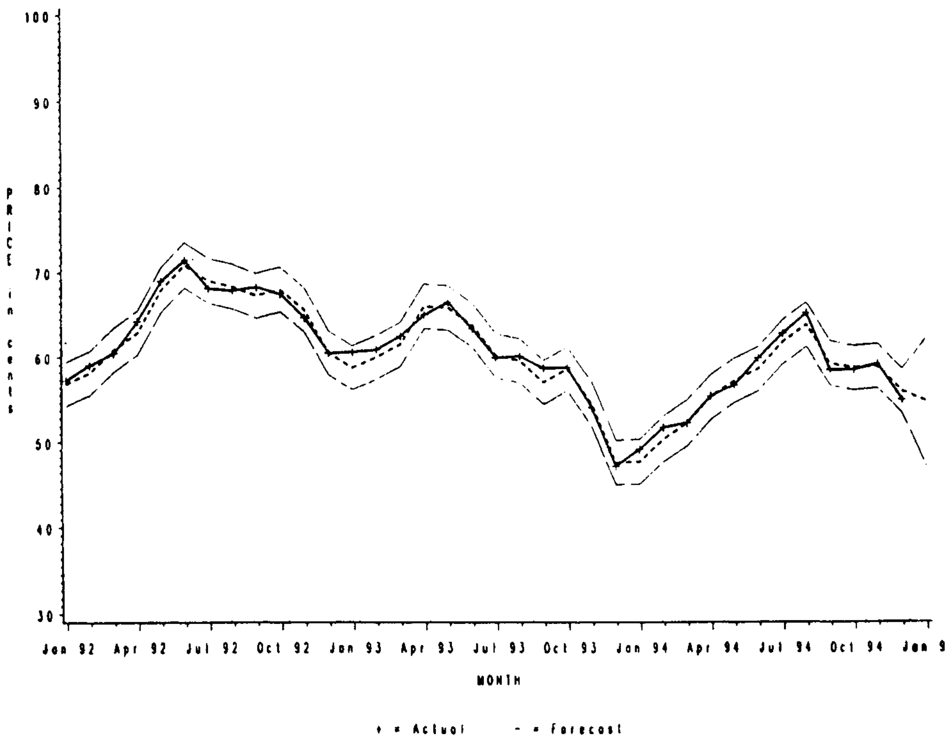

EXHIBIT 4

COMPARISON OF WHOLESALE REGULAR UNLEADED GASOLINE PRICE FORECASTS

MONTH	DIFFERENCE (ACTUAL - FORECAST) in cents											
	U.S.		PADD 1		PADD 2		PADD 3		PADD 4		PADD 5	
	1-MONTH AHEAD	2-MONTH AHEAD	1-MONTH AHEAD	2-MONTH AHEAD	1-MONTH AHEAD	2-MONTH AHEAD	1-MONTH AHEAD	2-MONTH AHEAD	1-MONTH AHEAD	2-MONTH AHEAD	1-MONTH AHEAD	2-MONTH AHEAD
Dec 93	2.3	3.1	3.9	7.0	5.1	8.4	3.4	7.6	1.6	4.7	2.3	5.4
Jan 94	0.3	-0.8	2.2	2.4	2.3	2.8	2.4	4.2	1.3	1.8	1.0	1.5
Feb 94	-0.6	1.0	-0.1	2.5	0.3	3.7	1.6	4.1	-1.3	-0.1	-0.8	1.5
Mar 94	2.0	2.0	2.5	3.1	2.8	3.8	2.2	2.8	0.4	4.2	2.7	3.2
Apr 94	-1.0	1.3	-0.5	1.7	0.2	3.8	-0.6	2.1	1.2	6.3	-2.1	-1.2
May 94	2.6	4.5	1.9	3.4	3.3	7.1	0.5	4.9	5.7	9.1	1.7	6.0
Jun 94	1.2	3.3	1.0	2.8	3.1	5.9	1.5	3.9	1.3	5.4	1.0	0.7
Jul 94	2.1	-4.6	1.5	-5.3	3.1	-6.9	2.1	-5.9	2.4	-0.5	-1.5	-2.7
Aug 94	-7.3	-6.9	-7.1	-5.3	-9.0	-10.2	-7.3	-7.7	-2.2	-10.3	0.1	1.3
Sep 94	3.7	3.6	3.7	7.1	0.4	0.4	2.0	2.3	-4.6	-4.8	1.9	2.7
Oct 94	-2.1	-6.4	2.4	-1.3	-2.2	-7.4	-1.4	4.9	-2.5	-8.0	-0.8	-4.4
Nov 94	-3.2	0.0	-3.9	-1.3	-4.6	-1.0	-2.5	0.6	-5.5	-5.6	-3.5	-2.7
Avg. Diff.	0.0	0.0	0.6	1.4	0.4	0.9	0.3	2.0	-0.2	0.2	0.2	0.9
Avg. Absolute Diff.	2.4	3.1	2.6	3.6	3.0	5.1	2.3	4.2	2.5	5.1	1.6	2.8