WP. 44
ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Neuchâtel, Switzerland, 5-7 October 2009)

Topic (viii): Selective and macro editing

PROCESS ENHANCEMENT AND IMPROVEMENT OF A MACRO EDITING APPROACH

Supporting Paper

Prepared by the U.S. Energy Information Administration, United States[1]

I. INTRODUCTION

1. The U.S. Energy Information Agency (EIA) estimates monthly product supplied (demand) of petroleum products from the components of supply as reported in a family of petroleum surveys. However, respondent level data may vary significantly overtime and across regions, making it difficult to detect misreporting. While it is expected that some of this variation will be smoothed when the data across surveys and respondents are combined, this is not always the case.

2. As a result, EIA is developing a methodology for checking aggregate data for outliers to supplement its respondent-level or micro-editing process. As reported last year (Meyer et. al., 2008) in an initial study, time-series econometric models were developed to identify reporting issues or frame deficiencies in the aggregate survey data and to guide subsequent editing and imputation activities. The initial results were promising. However, in 2008 the approach did not perform as well in response to shifts in consumer usage between high and low sulfur distillate fuel oil and to unprecedented increases and decreases in petroleum product prices. The effects of which were illustrated by the reduced out-of-sample forecast performance at the end of the time period studied. This indicated that additional work was necessary for process enhancement and improvement to adequately account for possible future structural shifts. The other question addressed in this paper is how best to determine when a change in the model structure is needed to detect possible reporting anomalies and errors.

3. This follow-up study reports on the continuing examination of monthly observations from January 2008 through April 2009. The analysis and process were changed from the static models used in the previous study to using recursive estimation for making projections of

[1] Prepared by P. Mason, J. Shore, and J. Zyren, U.S. Energy Information Administration; F. Joutz, Dept. of Economics, GWU; and J. Meyer, SAIC. This report is released to inform parties of ongoing research and to encourage discussion. The views expressed are those of the authors and not necessarily of the U.S. Energy Information Administration.

petroleum product supplied. In addition alternative models have been developed for comparison purposes. The approaches selected are ones that allow the model(s) to adjust as new information becomes available and provide a means for testing the model and coefficients for stability. In the time period since the previous report, it appeared that structural change(s) occurred in the data series. This merited the change in approach.

4. Recursive statistical techniques were explored and utilized, because they provided statistical measures for evaluating the performance vis-à-vis the data generating process for the series and checking for possible reporting errors. Unfortunately, these are not two immediately distinct empirical issues because they are interdependent. But the techniques do provide insights into what might be useful "triggers" for re-evaluating the model structure. For example, a one-time large error may only signal a data reporting problem or an unusual event. A pattern of increasing within-sample errors which exceeds a pre-determined error trigger criterion would be useful to re-evaluate the model structure. In addition, the out-of-sample forecasts would provide information on possible forecast failure or reporting errors from the survey data and micro-editing process. Forecast failure in this case would be characterized as a significant deterioration in forecast performance relative to the anticipated outcome. See Hendry and Ericsson (2001) and Clements and Hendry (2001) for further discussion of forecast failure.

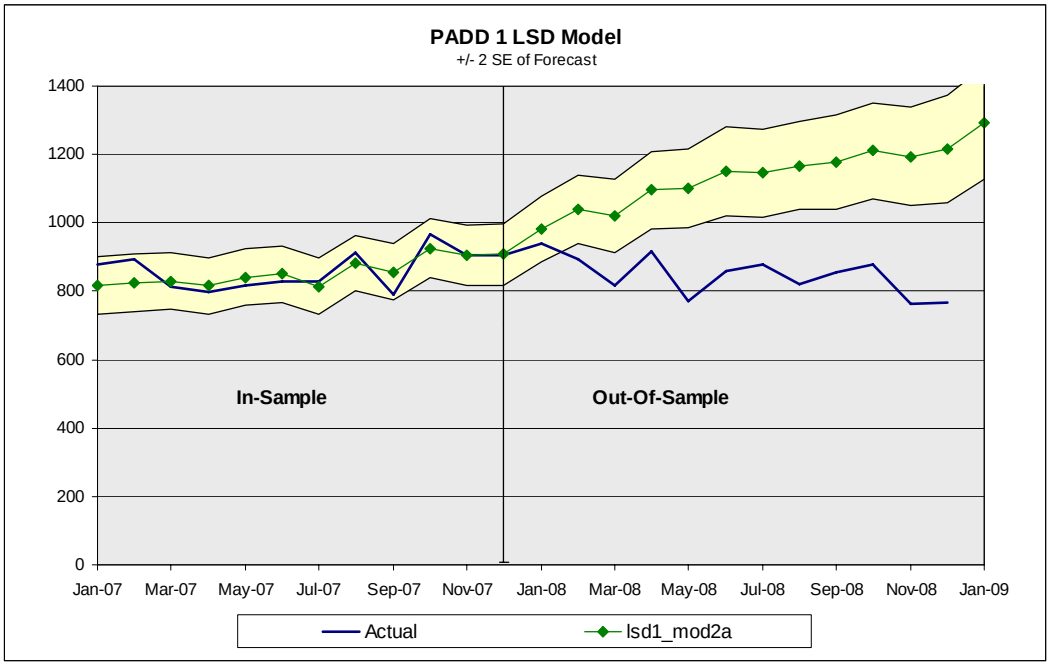
5. There are five additional sections to the paper plus the conclusion. The next section illustrates the performance problem of the macro-editing tool if not corrected or re-estimated. Section III demonstrates the potential benefits of recursive estimation for preliminary data and macro-editing. This is followed by discussion of triggers and indicators for use in the process. The next section briefly describes the alternative models developed for analysis. The last section presents the results from one month-ahead predictions using the original models and the alternative models.

II. UNCORRECTED MODEL PERFORMANCE

6. With every forecast month beyond the initial estimation period, there is an increasing possibility of forecast error, exposing possible misspecification of the model. This paper focuses on two representative Low-Sulfur Distillate (LSD) models, PADD 1 and PADD 2, corresponding to two geographic regions. The sample period for initial estimation was January 1996 through December 2007. This sample period was chosen because it pre-dates the extremes of the models' out-of-sample forecasting performance during 2008. In addition, the two regions appeared to capture different phenomenon in the reporting data and macro editing process.

7. At one extreme was the out-of-sample forecast of the PADD 1 LSD model which, if left uncorrected, deviated strongly from the actual survey results. This deviation occurred almost from the beginning of 2008, as shown below. The forecasting editing model projected a continuing upward trend from the historical data when the actual data were declining slowly. The decrease in product supplied was such that it consistently fell outside the forecast confidence intervals (Figure 1).

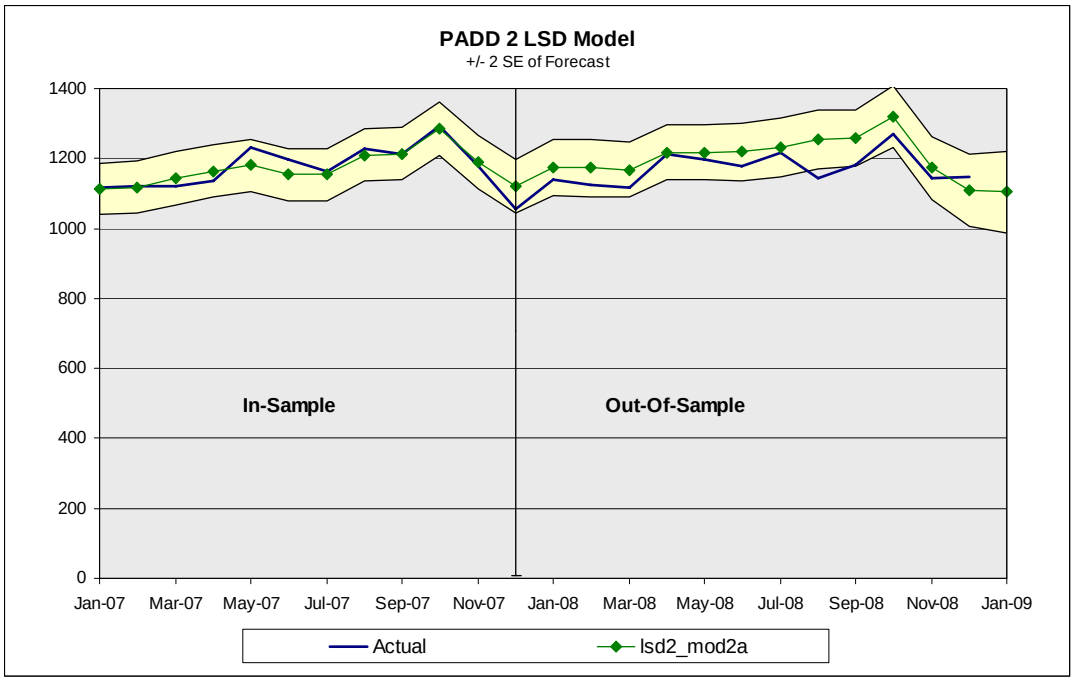
Figure 1. Out-of-Sample Performance PADD 1



8. At the other end of the spectrum was the PADD 2 LSD forecast. Here the reported or observed data fell within the 95% confidence interval in all but one of the months of the 2008 forecast period (Figure 2). Notwithstanding the superior performance of this model (particularly compared to the one above), it was not without fault. In the out-of-sample period, the PADD 2 LSD model consistently overestimated distillate demand. Both cases led to the exploration of corrective measures to improve the monthly forecast, and to consideration of what circumstances would dictate re-examination of each model's structure. When there appears to be a consistent pattern in the deviation of observed reports to forecast values, this may warrant action. It could come in the form of testing the model, considering alternative significance levels and establishing measures for notifying the survey production staff.

9. The data generating process for the LSD series may be one cause of the modeling and forecasting problem. For example, if the original data are stationary or not integrated, I(0), but in fact are non-stationary and integrated of order one, I(1), this creates specification problems in model estimation and further complications in out-of-sample forecasting. The former is known as the spurious regression problem and the later is due to the implication that the data generating process is subject to "permanent" shocks each period. Appendix A reports the preliminary test results for stationarity at the monthly and the annual frequency. They suggested that the evidence was mixed for PADD 1 LSD, while the PADD 2 LSD appears stationary, and neither exhibits integration at the annual frequency. The integration issue may be explored further in subsequent model reporting reviews.

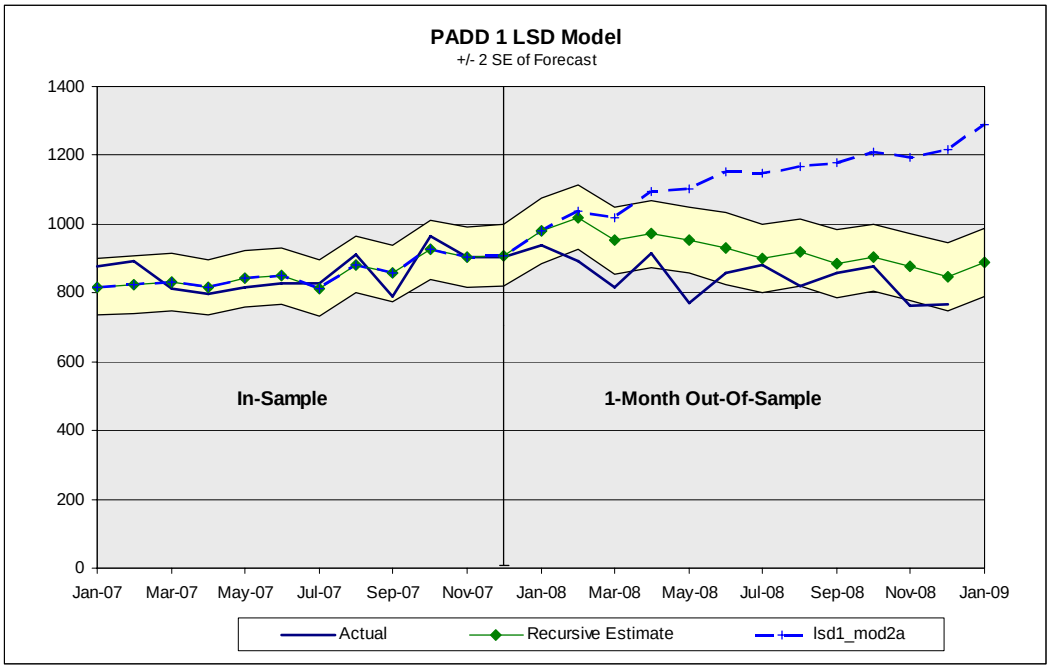
Figure 2. Out-of-sample LSD Forecast Performance PADD 2



III. RECURSIVE FORECASTS

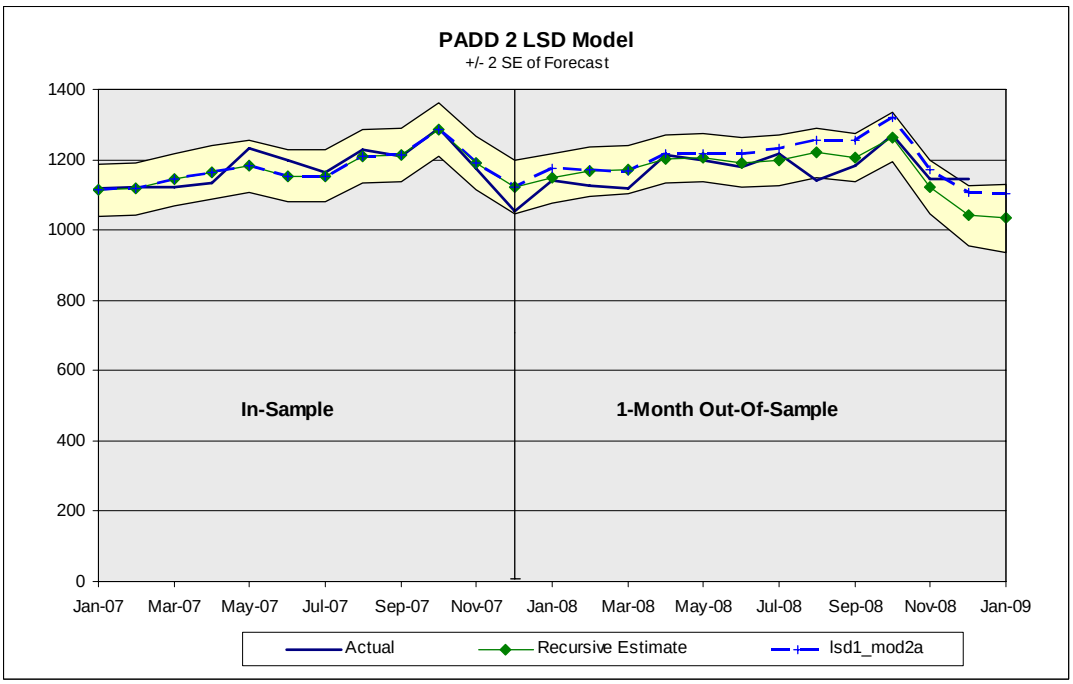
10. The corrective measure employed retained the original model specification, but re-estimated the coefficients on a monthly basis, thereby producing a true sequence of one month-ahead or period-ahead forecasts. Figure 3 depicts the result of this recursive estimation approach with the PADD 1 model; the unmodified forecast (lsd1_mod2a) is also shown for purpose of comparison. As can be seen in the figure, this approach provided a significant improvement over the original forecast. However, the problem of chronic over-prediction remained.

Figure 3. Recursive Estimation PADD 1



11. Using the same technique for the PADD 2 models also yielded some improvement in the quality of the sequence of one month-ahead forecasts, but the improvement is fairly modest (Figure 4). This improvement can be seen in a retrospective look at the RMSE of the forecasts during 2008, as shown in Table 1. There is a measurable improvement in the quality of the

Figure 4. Recursive Estimation PADD 2



monthly forecasts resulting simply from the re-estimation of the coefficients.

12. An alternative to the recursive estimation approach was to use rolling forecasts, trimming one month from the beginning of each forecast period as a new month was added to the end. This approach had the benefit of maintaining the same number of observations as the original model's estimation period, increasing the validity of directly comparing the corresponding RMSE's. The graphic results were virtually indistinguishable from those of the recursive estimates, above, so a tabular comparison is provided in Table 1 below.

Table 1.
One Month-Ahead RMSE Comparisons 2008

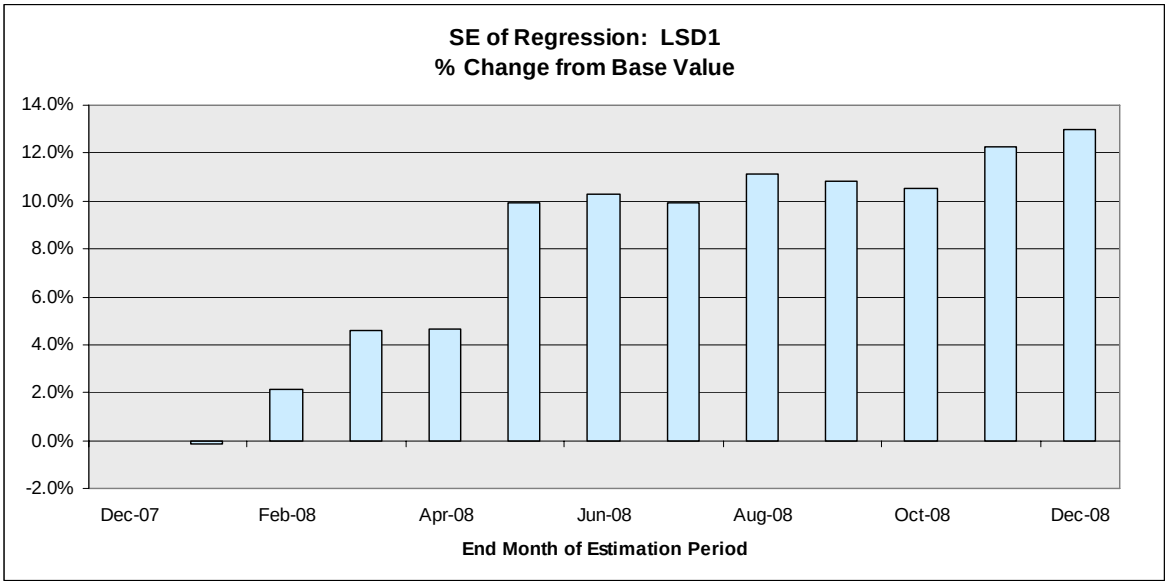
	PADD 1	PADD 2
Original Model	300.1	47.9
Recursive Estimates **(% Change)**	94.6 (-68%)	42.1 (-12%)
Rolling Estimates **(% Change)**	92.3 (-69%)	41.7 (-13%)

IV. TRIGGERS/INDICATORS

13. While the recursive/rolling estimation approach represented a significant improvement over the original forecast, the models still uniformly overestimated demand throughout the year. It was therefore important to develop a set of metrics for the future that would prompt a re-examination of the model's structure if there were an indication of persistent bias or other problems with the forecast. The idea was to come up with "thresholds" for sounding the alarm to discern when new reported estimates of petroleum products supplied suggest that there is a potential measurement problem from the situation where there is a potential problem with the model. In the first case, analysts would decide what corrective action to take regarding the reported data. In the second, the modelers would need to decide on the adequacy of the model.

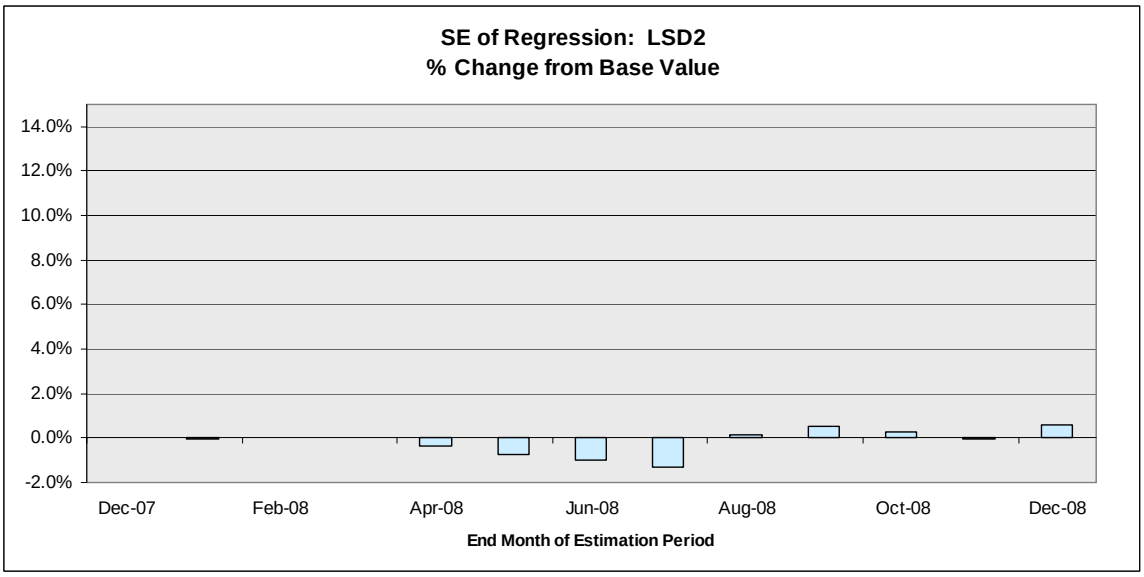
14. The recursive modeling approach or a periodic updating of the model yields new parameter estimates. One possible trigger or measure considered was to look at the trend or change in the standard errors (SE) of the regression which are used in the confidence intervals for a given model. In a production environment, the difference in the original estimated SE could be compared to the new estimate when additional information was used in the model forecast. Figure 5 depicts the percentage change from the initial value for PADD 1 (for the regression ending in December 2007). The percent change from the base value is for the one month-ahead forecast standard error values. A sharp jump in the SE might be indicative of the need to re-examine the model. For example, between April and May 2008 the percent change from the base value in the standard error increased from 4% to 10%. A jump of this numerical size had not historically occurred. Work toward investigating a means to test changes in the standard errors through techniques similar to interval estimation of the estimated variance term is still in progress.

Figure 5. Percent Change in Regression SE from Base Value, PADD 1



15. By comparison, a similar measure for PADD 2 (using the same scale) did not provide a strong indication of model failure (Figure 6).

Figure 6. Percent Change in Regression SE from Base Value, PADD 2



V. ALTERNATIVE MODELS

16. Three alternative models were separately developed for both PADD regions to address empirical and forecasting issues encountered in the Base model and the recursive versions of the Base model. The first two alternative models were based on a general-to-specific (GETS)

modeling methodology and differed only in their level of specification: one was in the original level and the second used the natural log transform. This method uses maximum likelihood to minimize the number of variables in order to simplify an initially general model which characterizes the empirical evidence within a theoretical statistical framework. The third model was estimated in state space form using STAMP 8 (Koopmans, Harvey, Doornik, and Shepard, 2007). The dependent variable was transformed to natural logarithms in each case, because the data generating process appeared to approximate a log-normal distribution.

17. These alternative models were considered for several reasons. First, the breakdown in the original Base model suggested that alternative specifications may yield better results. The market or structural shift from High Sulfur Distillate (HSD) to LSD in 2007 was the direct result of changes in environmental regulations. This appears to have been compounded by changes in demand as a result of the recession for much of the remaining out-of-sample periods. Second, the updating procedure and recursive modeling of the Base model were laborious and did not have a strong enough theoretical foundation for the macro-editing team. Third, the recent developments in the theory of forecasting and forecast failure cited earlier suggested simpler model specifications. Fourth, the state space modeling approach with the Kalman filter provides a useful and intuitive way of examining preliminary data from statistical surveys. The Kalman filter captures the tradeoff from model uncertainty with preliminary data uncertainty in generating the forecasts.

VI. COMPARISON OF ONE MONTH-AHEAD FORECASTS

18. The one month-ahead forecasts for PADD 1 and PADD 2 low sulfur distillate were compared using five different models. Two tables are reported for each PADD. The first provides the level series forecasts for each model and summary forecast error statistics. Percent forecast errors are provided in the second table for each PADD. The forecast evaluation covers a period of 16 months from January 2008 to April 2009. The models were re-estimated up to the previous month and then a one month-ahead forecast was made.

19. Table 2a reports the results for PADD 1 in level form, Table 2b contains the one month-ahead percent forecast errors. There are eight columns in each table. The forecast month is the first column; the next two give the published series and the unedited series, respectively. Columns five and six present the forecasts from the Base model and the Recursive model. The next two columns contain the GETS model forecasts from the level and natural log specifications of the regression models incorporating new developments in forecasting theory. The eighth and final column presents the forecast results from the State Space modeling (KF) approach. PADD 2 results are given in Tables 3a and 3b using the identical format.

20. The last four rows of Tables 2a and 3a report the summary forecast errors statistics for the five alternative models. The first is the average or mean error. Measures of dispersion for the forecast errors are given in the next two rows as the root mean square error and the mean absolute error. The mean percent error is reported in the last row; the individual percentage forecast errors are provided in Tables 2b and 3b.

21. The Base model in PADD 1 clearly missed the structural break in the trend of LSD as discussed earlier (Figure 1); the average error (average percent error) was nearly 320 thousand barrels per day (kbd) (almost 40%) for petroleum supplied which declined from over 900 kbd to 740 kbd by the end of the sample. The mean error and mean percent error for the Recursive model, the two GETS models and the State Space model were 69 (8.5%), -9.3 (-0.8%), -4.6 (-0.2%), and 34 (4.5%) respectively. None of the mean error terms exceeded their respective

RMSE's for the four forecasts; the bias measure was not statistically significant. The lowest RMSE's were for the GETS models followed by the State Space model. Mean absolute errors (MAE) did not suggest that the four error terms were positively or negatively skewed. However, the Base model was skewed toward over-prediction for all 16 forecasts. In addition, the Base model did not under-predicted the unedited series in any of the months. The two GETS models and the State Space model under-predicted in 9, 8, and 5 months, respectively. Except for the Base model, all forecasts had large numerical positive errors relative to the unedited series in November 2008 (7%-14%) and large numerical negative smaller (10%-24%) than in May 2008.

22. The Base model in PADD 2 over-predicted LSD in 15 of the 16 months; mean error was almost three times that of the other models. The Recursive model mean error was about 28 kbd nearly twice that of the alternative models, but still only had an average percent error of 2.6%. The mean error and mean percent error for the two GETS models and the State Space model were 15.6 kbd (1.5%), 16.4 kbd (1.6%), and -15.6 (1.3%) respectively. None of the mean error terms exceeded their respective RMSE's for the four forecasts; the bias measure was not statistically significant. The lowest RMSE's were for the State Space and the GETS models. There were negative forecast errors in 4, 7, 8, and 10 months, respectively, for the Recursive model, the GETS level model, the GETS log level model, and the State Space model. The month with the largest over-predictions was August 2008 and the largest under-predictions were in December 2008. The State Space model appeared to have much smaller relative percent forecast errors than the GETS or Recursive model.

Table 2a
Comparison of Actual Series with
Alternative One Month-Ahead Forecasting Models
PADD1 Product Supplied - Low Sulfur Distillate (kbd)

Month	Published Series	Unedited Series	Base Model	Recursive Model	GETS Model Level	GETS Model Nat. Log	State Space (KF)
Jan-08	971	938	985	985	856	865	902
Feb-08	884	893	1,041	1,020	902	904	934
Mar-08	817	818	1,022	953	880	878	855
Apr-08	921	917	1,096	969	879	887	872
May-08	820	772	1,103	954	869	872	853
Jun-08	844	859	1,149	927	834	827	903
Jul-08	857	880	1,143	896	840	839	848
Aug-08	802	822	1,163	914	854	845	917
Sep-08	857	857	1,172	883	826	824	875
Oct-08	884	878	1,205	901	839	844	900
Nov-08	828	763	1,184	870	821	821	872
Dec-08	708	766	1,201	840	803	810	839
Jan-09	911	911	1,290	894	783	800	851
Feb-09	764	764	1,245	848	760	773	888
Mar-09	818	818	1,247	802	744	752	809
Apr-09	737	737	1,276	841	755	778	815
Mean Error			320.5	69.0	-9.3	-4.6	33.7
RMSE			344.4	87.4	58.2	55.4	65.1
MAE			320.5	73.1	48.3	47.8	56.3
Mean % Error			39.4%	8.5%	-0.8%	-0.2%	4.5%

Table 2b
Actual Series with One Month-Ahead Percent Forecast Errors from Alternative Forecasting Models
PADD1 Product Supplied - Low Sulfur Distillate (kbd)

Month	Published Series	Unedited Series	Base Model	Recursive Model	GETS Model Level	GETS Model Nat. Log	State Space (KF)
Jan-08	971	938	5.1%	5.1%	-8.8%	-7.8%	-3.8%
Feb-08	884	893	16.6%	14.3%	1.0%	1.2%	4.6%
Mar-08	817	818	25.0%	16.5%	7.5%	7.4%	4.5%
Apr-08	921	917	19.5%	5.7%	-4.2%	-3.2%	-4.9%
May-08	820	772	42.9%	23.6%	12.6%	12.9%	10.5%
Jun-08	844	859	33.7%	7.9%	-2.9%	-3.8%	5.1%
Jul-08	857	880	29.9%	1.9%	-4.5%	-4.6%	-3.6%
Aug-08	802	822	41.4%	11.2%	3.9%	2.8%	11.5%
Sep-08	857	857	36.7%	3.0%	-3.6%	-3.8%	2.1%
Oct-08	884	878	37.3%	2.7%	-4.4%	-3.8%	2.5%
Nov-08	828	763	55.1%	14.0%	7.6%	7.6%	14.3%
Dec-08	708	766	56.8%	9.6%	4.8%	5.7%	9.5%
Jan-09	911	911	41.5%	-1.8%	-14.1%	-12.2%	-6.6%
Feb-09	764	764	63.0%	11.0%	-0.5%	1.2%	16.2%
Mar-09	818	818	52.4%	-2.0%	-9.1%	-8.0%	-1.1%
Apr-09	737	737	73.1%	14.2%	2.4%	5.5%	10.5%
Mean Error			39.4%	8.5%	-0.8%	-0.2%	4.5%

Table 3a
Comparison of Actual Series with
Alternative One Month-Ahead Forecasting Models
PADD2 Product Supplied - Low Sulfur Distillate (kbd)

Month	Published Series	Unedited Series	Base Model	Recursive Model	GETS Model Level	GETS Model Nat. Log	State Space (KF)
Jan-08	1083	1140	1,176	1,176	1,119	1,089	1101
Feb-08	1120	1125	1,174	1,168	1,112	1,100	1122
Mar-08	1126	1118	1,162	1,155	1,147	1,162	1162
Apr-08	1209	1212	1,216	1,207	1,186	1,207	1185
May-08	1229	1197	1,213	1,202	1,174	1,187	1193
Jun-08	1203	1180	1,220	1,206	1,187	1,189	1193
Jul-08	1176	1218	1,224	1,207	1,172	1,160	1138
Aug-08	1139	1142	1,236	1,223	1,210	1,213	1185
Sep-08	1166	1182	1,255	1,236	1,223	1,235	1202
Oct-08	1270	1270	1,309	1,285	1,280	1,312	1256
Nov-08	1158	1145	1,160	1,127	1,138	1,132	1122
Dec-08	1134	1146	1,086	1,060	1,100	1,079	1024
Jan-09	1064	1064	1,083	1,101	1,114	1,096	990
Feb-09	1025	1025	1,100	1,096	1,109	1,109	1012
Mar-09	1031	1031	1,134	1,117	1,128	1,134	1041
Apr-09	1039	1039	1,145	1,115	1,084	1,093	1060
Mean Error			41.4	28.0	15.6	16.4	-15.6
RMSE			58.6	51.5	46.4	52.7	46.8
MAE			48.8	42.9	38.3	45.0	34.5
Mean % Error			3.8%	2.6%	1.5%	1.6%	-1.3%

Table 3b
Actual Series with One Month-Ahead Percent Forecast Errors from Alternative Forecasting Models
PADD2 Product Supplied - Low Sulfur Distillate (kbd)

Month	Published Series	Unedited Series	Base Model	Recursive Model	GETS Model Level	GETS Model Nat. Log	State Space (KF)
Jan-08	1083	1140	3.2%	3.2%	-1.8%	-4.4%	-3.4%
Feb-08	1120	1125	4.4%	3.8%	-1.1%	-2.2%	-0.3%
Mar-08	1126	1118	3.9%	3.3%	2.6%	4.0%	3.9%
Apr-08	1209	1212	0.4%	-0.4%	-2.2%	-0.4%	-2.3%
May-08	1229	1197	1.3%	0.4%	-1.9%	-0.9%	-0.3%
Jun-08	1203	1180	3.4%	2.2%	0.6%	0.8%	1.1%
Jul-08	1176	1218	0.5%	-0.9%	-3.8%	-4.8%	-6.6%
Aug-08	1139	1142	8.2%	7.1%	5.9%	6.2%	3.8%
Sep-08	1166	1182	6.2%	4.6%	3.5%	4.4%	1.7%
Oct-08	1270	1270	3.1%	1.2%	0.8%	3.3%	-1.1%
Nov-08	1158	1145	1.3%	-1.6%	-0.6%	-1.1%	-2.0%
Dec-08	1134	1146	-5.2%	-7.5%	-4.0%	-5.8%	-10.7%
Jan-09	1064	1064	1.8%	3.5%	4.7%	3.0%	-7.0%
Feb-09	1025	1025	7.3%	6.9%	8.2%	8.2%	-1.3%
Mar-09	1031	1031	10.0%	8.4%	9.5%	9.9%	1.0%
Apr-09	1039	1039	10.2%	7.3%	4.3%	5.2%	2.1%
Mean Error			3.8%	2.6%	1.5%	1.6%	-1.3%

VII. CONCLUSIONS

23. Initial results for developing and using a model-based macro editing approach were promising but the models' performance deteriorated as a result of shifts in consumer usage between high and low sulfur distillate and large declines in petroleum product prices. This follow-up study focused on examination of possible improvements that could be gained from recursive estimation of the original model each month or the consideration of alternative models. With the longer out-of-sample history available, it was demonstrated that the Base models were no longer useful.

24. The recursive models which retained the Base model specification but re-estimated the coefficients each month provided significant improvements but consistently over-predicted. As a result, three alternative models were developed in an attempt to overcome the breakdown in the Base model by accounting for the market and structural shifts and to provide a strong theoretical foundation capturing recent developments in theory with simpler model specifications.

25. The first two alternative models developed were OLS based and only differed in the level of specification in that one used raw data and the other the natural log transform. The two GETS models performed better than the recursive models, producing lower RMSE's. The third alternative made use of the state space modeling approach with the Kalman filter to capture the

tradeoff between model uncertainty and preliminary data uncertainty in generating the forecasts. The state space models appeared to have much smaller relative percent forecast errors than the GETS models or the recursive models.

26. While triggers were original viewed as necessary for re-evaluating the model structure within a production environment, the approach for such triggers was not evident. Therefore, priority was placed instead on improving the models, particularly through the use of modeling approaches that were more self correcting. These early results from the alternative models were encouraging. Further research is required to confirm their performance under new market and structural shifts as well as their long term usefulness in a survey production environment.

References

Clements, Michael, Hendry, David F., (2001) Forecasting Non-Stationary Time Series, The MIT Press, Cambridge, MA.

Harvey, Andrew C., (1989) Forecasting Structural times Series Models and the Kalman filter, Cambridge University Press.

Hendry, David F., Ericsson, Neil R., (2001) Understanding Economic Forecasts, The MIT Press, Cambridge, MA.

Joutz, F., Stekler, H., (2000) "An evaluation of the predictions of the Federal Reserve", International Journal of Forecasting 16, 17-38.

Koopman Siem-Jan, Harvey, Andrew C., Doornik, Jurgen A. and Shepard, Neil, (2007) Structural Time Series Analyser and Modeller and Predictor: STAMP 8.0, Timberlake Consultants Ltd.

Meyer, J., Shore, J., Weir, P. and Zyren, J. (2008) "The Development of a Macro Editing Approach", Conference of European Statisticians, Work Session on Statistical Data Editing, Invited Paper, Vienna 2008.

Appendix A: Stationarity or order of Integration of the Low Sulfur Distillate Series

Testing for Integration

Table A.1 contains the results from the Augmented Dickey-Fuller (ADF) tests. The top half of the table reports the test results for whether the series in levels are stationary I(0) and the bottom half of the table is for whether the first difference of the series is stationary I(1). All variables have been converted to natural logarithms. The dependent variable is the first difference of the variable of interest. It is regressed on the lag level of the series, a constant, trend, centered seasonal or monthly dummy variables, and up to six lags of the dependent variable. The specification is given below.

$$\Delta y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 * t + \sum_{i=1}^{6} \gamma_i \Delta y_{t-i} + \sum_{j=1}^{11} \delta_j Month_{jt} + \varepsilon_t$$

There are six columns in the table. The first column gives the variable. Columns two and three provide the t-statistic from the ADF test and the implied coefficient on the lagged level term. The last three columns help to explain the lag specification of the model; they are the t-statistic on the maximum significant lag of the differenced variable, the maximum lag length chosen for the test and the associated AIC value. The t-DY lag and AIC measures are used in evaluating the appropriate number of lags in the testing procedure which remove serial correlation in the residuals.

The actual null hypothesis in the each test is that there is a unit root or the series is I(1); this implies that the coefficient α_1 is insignificant from zero. Because of the properties of non-stationary series, the distribution of the statistics is non-standard. The critical values are found at the bottom of the table.

We cannot reject the null of a unit root in the level of the PADD 1 low sulfur variables and the unedited PADD 2 low sulfur variable. However, numerically the estimated coefficients are not particularly close to unity. The adjusted PADD 2 low sulfur series appears to be stationary; the associated t-statistic for the ADF is significant at 1%. The same is true for both the U.S. low sulfur variables. The second half of Table 2 reports the ADF test for the same variable with the hypothesis that they are non-stationary in first differences. We can reject the null hypothesis in all cases. The evidence is mixed for the two PADD variables; they could be treated as I(0) or I(1) series for modeling purposes.

Table A.2 considers a slightly different hypothesis. The null hypothesis is that the variables have unit roots at the annual frequency. The ADF regression equation above has been modified as follows:

$$\Delta^{12} y_t = \alpha_0 + \alpha_1 y_{t-12} + \alpha_2 * t + \sum_{i=1}^{6} \gamma_i \Delta^{12} y_{t-i} + \sum_{j=1}^{11} \delta_j Month_{jt} + \varepsilon_t$$

The dependent variable is given as the annual difference of the variable. The lagged level term is from twelve months ago not the previous month. This table only reports the level form of the test. The null hypothesis that the series were I(1) at the annual frequency was rejected in all cases.

Table A.1: Augmented Dickey-Fuller Tests for Unit Roots
Levels - sample 1995m1 - 2008m12

Variable	t-adf	alpha Y_lag	t-DY_lag	Maximum Lags	AIC
LLSD1_unedited	-2.26	0.77	-1.87	5	-5.42
LLSD1	-2.22	0.77	-1.44	5	-5.45
LLSD2_unedited	-2.26	0.77	-1.87	5	-5.42
LLSD2	-4.21**	0.64	-3.59	2	-6.43
LLSDUS_unedited	-3.57*	0.65	-4.44	2	-6.79
LLSDUS	-3.69*	0.64	-3.87	2	-6.80
Critical Values for: Constant+Trend+Seasonals; 5%=-3.44 1%=-4.02					
First Differences - sample 1995m1 - 2008m12					
LLSD1_unedited	-9.086**	-2.43	2.30	4	-5.40
LLSD1	-12.44**	-1.59	3.25	2	-5.41
LLSD2_unedited	-17.27**	-1.20	5.86	1	-6.28
LLSD2	-10.99**	-1.35	1.50	2	-6.33
LLSDUS_unedited	-17.56**	-1.21	6.18	1	-6.72
LLSDUS	-16.40**	-1.10	5.60	1	-6.73
Critical Values for: Constant+Trend+Seasonals; 5%=-3.44 1%=-4.02					

Table A.2: Augmented Dickey-Fuller Tests for Annual Unit Roots
Levels - sample 1995m1 - 2008m12

Variable	t-adf	alpha Y_lag	t-DY_lag	Maximum Lags	AIC
LLSD1_unedited	-6.94**	-0.42	3.97	3	
LLSD1	-6.66**	-0.39	3.60	3	
LLSD2_unedited	-6.94**	-0.42	3.97	3	
LLSD2	-6.66**	-0.39	3.60	3	
LLSDUS_unedited	-7.24**	-0.49	3.27	3	
LLSDUS	-7.21**	-0.51	2.70	3	
Constant+Trend+Seasonals; 5%=-3.44 1%=-4.02					