BLS WORKING PAPERS

U.S. DEPARTMENT OF LABOR
Bureau of Labor Statistics

OFFICE OF PRICES AND LIVING
CONDITIONS

A State Space Model-Based Approach to Intervention
Analysis in the Seasonal Adjustment of BLS Series
Some Empirical Results

Raj K. Jain

Working Paper 228
June 1992

A STATE SPACE MODEL-BASED APPROACH TO INTERVENTION ANALYSIS IN THE SEASONAL ADJUSTMENT OF BLS SERIES: SOME EMPIRICAL RESULTS

Raj K. Jain
U.S. Bureau of Labor Statistics, 600 E Street, N.W., Room 4013, Washington, D.C. 20212

Keywords: Explanatory variables, Kalman Filter, EM Algorithm

The Intervention and Explanatory variables are incorporated in the statistical models of seasonal adjustment of BLS series. The State-Space/Kalman Filter methodology is used to estimate these models. EM algorithm and a quasi-Newton algorithm are employed to estimate the hyper-parameters of those models. Two BLS series are seasonally adjusted using these models and the empirical results relating to the effects of intervention analysis and explanatory variables on seasonal adjustment presented.

1. INTRODUCTION

In the last decade, two events occurred which significantly affected the prices of crude and refined petroleum products. In January 1981, the U.S. Price Allocation and Entitlement regulation were lifted which effectively decontrolled the oil prices. Immediately, the price indexes of such products registered significant upward jumps. Again, in January 1986, Saudi Arabia made a political decision to almost double the production of crude oil and the prices of crude oil fell sharply. Consequantly, the BLS price indexes of these products registered a very sharp one month decline. Such events called "Interventions" in statistical terminology, are external to the market forces of demand and supply which determine the prices of these products. Now the quality of seasonal adjustment of any series which is affected by such interventions will be adversely affected if the effect of interventions is not accounted for by the method used to seasonally adjust that series. The X-11 ARIMA which is the method currently being used by BLS to seasonally adjust all price index series, does not have any mechanism in it to eliminate the effects of interventions; hence their effect is first eliminated from a series by ad-hoc methods and the series is then seasonally adjusted by the X-11 ARIMA method.

The state space structural model-based method of seasonal adjustment , on the other hand, can simultaneously adjust for the interventions as well as for the seasonal effects. This is done by introducing in the structural models, intervention variables (to be defined later) as explanatory variables in addition to the seasonal and trend components and other explanatory variables.

In section 2, the structural model for seasonal adjustment with intervention and other explanatory variables is presented. The state space form of this model and its estimation is also discussed. In section 3, empirical results from the estimation of this model using three different gasoline series are discussed. Final section includes the summary and conclusions.

2. STRUCTURAL MODEL OF SEASONAL ADJUSTMENT

A complete seasonal adjustment model consists of a decomposition equation for the time series into its unobserved components, intervention variables and other explanatory variables, a model for each of its unobserved components, definitions of intervention and explanatory variables and the dynamic specification of the parameters of the intervention and other explanatory variables. The general form of this structural model is as follows:

$$y_t = \sum_{i=1}^{k_1} \beta_{ti} x_{ti} + \sum_{j=1}^{k_2} + \delta_{tj} I_{tj} + \mu_t + \gamma_t + \varepsilon_t \tag{1}$$

$$\beta_{ti} = \beta_{t-1,i} + \theta_{ti} \qquad i = 1, 2, \ldots, k_1 \tag{2}$$

$$\delta_{tj} = \delta_{t-1,j} + \phi_{tj} \qquad i = 1, 2, \ldots, k_2 \tag{3}$$

$$\mu_t = 2\mu_{t-1} - \mu_{t-2} + \eta_t \tag{4}$$

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + \omega_{t1} \tag{5}$$

$$t = 1, 2, \ldots, n$$

where y_t is the observed series, x_{ti} and I_{tj} are explanatory and intervention variables respectively, which are different for each series fitted to the model; β_{ti} and δ_{tj} and the corresponding parameters which are stochastic variables following a random walk without drift in equations (2) and (3). Some other models of unobserved components μ_t and γ_t is equations (4) and (5) were tried but the model above provided the best fit and forecasting performance for all series. The errors of all the five equations are assumed to be mutually and serially uncorrelated random variables having zero mean and constant variance. This structural model of seasonal adjustment is cast into a state space form and Kalman filtering and smoothing technique is used to estimate all the unobserved components including μ_t, γ_t and parameters β_{ti} and δ_{tj}. The State space form of this model is:

$$y_t = Z_t \alpha_t + \varepsilon_t \tag{6}$$

$$\alpha_t = T\,\alpha_{t-1} + R\xi_t \tag{7}$$

where

$$Z_t = [x_{t1},\dots, x_{tk1}, I_{t1},\dots, I_{tk2}, 1\ 0\ 1\ 0\ 0 \dots 0]_{1x\,(k_1 + k_2 + s + 1)}$$

$$\alpha_t = [\beta_{t1},\dots, \beta_{tk1}, \delta_{t1},\dots,\delta_{tk2}, \mu_t, \mu_{t-1}, \gamma_t, \gamma_{t-1}, \dots, \gamma_{t-s+2}]'_{(k_1 + k_2 + s + 1)\,x\,1}$$

$$T = \begin{bmatrix} I_{k1+k2} & 0 \\ 0 & T_{22} \end{bmatrix}$$

$$T_{22} = \begin{bmatrix} 2 & -1 & 0 & 0 & - & - & 0 & 0 \\ 1 & 0 & 0 & 0 & - & - & 0 & 0 \\ 0 & 0 & -1 & -1 & - & - & -1 & -1 \\ 0 & 0 & 1 & 0 & - & - & 0 & 0 \\ 0 & 0 & 0 & 1 & - & - & 0 & 0 \\ - & - & - & - & - & - & - & - \\ - & - & - & - & - & - & - & - \\ 0 & 0 & 0 & 0 & - & - & 1 & 0 \end{bmatrix}_{(s+1)x(s+1)}$$

I_{k1+k2} is the $k_1 + k_2$ identify matrix.

$R = \text{zero}\,(k_1 + k_2 + s + 1, k_1 + k_2 + 3)$ except

$R(i,i) = 1,\ \ i = 1, 2, \dots, k_1 + k_2 + 1, k_1 + k_2 + 3;$

$\xi_t = [\theta_{t1},\dots, \theta_{tk1}, \phi_{t1},\dots, \phi_{tk2}, \eta_t, 0, \omega_{t1}]'$;

$E\varepsilon_t = 0; E\varepsilon_t^2 = \sigma_\varepsilon^2$

$$E\xi_t = 0 \quad E\xi_t\xi_t' = \sigma_\varepsilon^2 \; Q = \sigma_\varepsilon^2 \, \mathrm{diag}\left(\frac{\sigma_{\theta 1}^2}{\sigma_\varepsilon^2},\ldots,\frac{\sigma_{\theta k1}^2}{\sigma_\varepsilon^2},\frac{\sigma_{\phi 1}^2}{\sigma_\varepsilon^2},\ldots,\frac{\sigma_{\phi k2}}{\sigma_\varepsilon^2},\frac{\sigma_\eta^2}{\sigma_\varepsilon^2},0,\frac{\sigma_{\omega 1}^2}{\sigma_\varepsilon^2}\right)$$

Equations (6) and (7) are respectively the measurement and state transition equations. The variance σ_ε^2 and the diagonal covariance matrix Q whose diagonal elements are relative variances (relative to σ_ε^2)(also called hyper-parameters) of the random errors of the model components and explanatory and intervention variable equations are unknown parameters to be estimated. To do this, it is assumed that ε_t and ξ_t $t = 1, 2, \ldots$ n, are uncorrelated normally disitrbuted random variable and vector respectively. Z_t, R and T are given matrices which are obtained from the structure of the model.

The estimation of the seasonal adjustment model in the state space form (6) and (7) is done in two parts:

(i) estimation of the state vector α_t and it's covariance matrix P_t, given the initial value of the state vector (α_0), and it's covariance matrix (P_0) and initial values of the matrix of hyper-parameters (Q_0), by Kalman filtering and smoothing. This technique is discussed in the literature. See, for example, Kalman (1960), Anderson and Moore (1979), Jaswinski (1970), and Harvey (1981) and (1991).

(ii) estimation of Q and σ_ε^2 by EM algorithm and BFGS numerical optimization techniqes. For a discussion of EM algorithm, see Dempster et. al. (1977) and Shumway and Stoffer (1982). The likelihood function which is optimised to estimate Q is obtained via prediction error decomposition using Kalman Filter.

The Kalman Filter is initialized with α_0 and $P_0 = kI$ where k is chosen to be a large but finite number and I is the identity matrix. To start the filter, Q_0 is set to equal the identity matrix.

3. EMPIRICAL RESULTS

The seasonal adjustment model presented above is estimated from the following three price index series (i) CPI of gasoline, (ii) PPI of gasoline, and (iii) PPI of domestic crude petroleum. The sample period is from January 1979 to December 1986 for each series. These data series were chosen because interventions of January 1981 and January 1986 had significantly affected these series. Several structural models with different specifications for unobserved component models for trend and seasonal were fitted to these series but the model in section (2) gave the best fit and best forecasting performance.

1. CPI of Gasoline: This is an nonstationary time series and is quite volatile as can be seen from figure 1. The wholesale price of gasoline has a very direct effect on the retail price of gasoline. Any significant change in the wholesale price of gasoline is almost immediately translated into a change in the retail price of gasoline. Hence, PPI of gasoline is included as an explanatory variable in the model. The CPI of gasoline has also been affected by external interventions mentioned in the introduction. Two intervention variables are included in the model which are defined as step functions as follows:

$$
\begin{aligned}
I_{1t} &= 0 \quad \text{for} \quad t \leq \quad \text{Feb. 1981} \\
&= 1 \quad \text{for} \quad t > \quad \text{Feb. 1981} \\
I_{2t} &= 0 \quad \text{for} \quad t \leq \quad \text{Feb. 1986} \\
&= 1 \quad \text{for} \quad t > \quad \text{Feb. 1986}
\end{aligned}
$$

The coefficients of the two intervention variables as well as the coefficient of the explanatory variable x_1 are assumed to follow a random walk process without drift. The empirical estimates of the hyper-parameters in Table 1 indicate that the seasonal component and the coefficients of I_1 and x_1 are deterministics because the hyper-parameter estimates are close to zero. The coefficients of I_2, on the other hand, is stochastic in character because the corresponding hyper-parameter estimate is quite large. The performance statistics of this model are given in Table 2. The Ljung-Box $Q^*(k)$, $k = 12,24$ which has a χ^2 distribution with 7 and 19 degrees of freedom has values in the region of acceptance of the null hypothesis of no autocorrelation in errors. This, together with the graph of CUSUM lying within 5 percent limits in figure 1 show that the errors do not have any systematic component in them. Hence, the model has adequate explanatory power. The goodness of fit statistics AIC, BIC, PEV, and R^2 seem quite reasonable which means the fit is quite good. The figure 1 shows that the observed series lies within 95 percent prediction intervals constructed from 12 month one step ahead forecasts and 12 months multistep ahead forecasts. The low values of post sample prediction error sum of squares (both one step ahead and multistep ahead) and the very low value of predictive F tests which indicate that the post sample PEV and within sample PEV are not significantly different, indicate that the model forecasts the series quite well. The low values of the estimates of revision (both AMR and RMSR) indicate that the revision of the seasonal adjusment would be minimal. Although the seasonal pattern estimated for this series is deterministic because the corresponding hyper-parameter estimate is zero, the seasonality in this series is not very pronounced.

2. PPI of Gasoline. This series is quite similar to CPI of gasoline series as can be seen in figure 2. The price of crude petroleum is the main component of the wholesale price of gasoline; hence the PPI of crude petroleum (domestic) is used as an explanatory variable. The same intervention variables as defined above for CPI of gasoline are used in estimating the model using PPI of gasoline. The estimtes of hyper-parameters corresponding to the two intervention variables indicate that the coefficients of the intervention variables are stochastic in character. The seasonal pattern, however, is found to be deterministic. The trend and the coefficients of explanatory variables are not deterministic but the low values of the estimates of their hyper-parameters indicate that these are not very random either. The performance of this model in seasonally adjusting the PPI of gasoline in the presence of interventions can be gauged from the various statisitcs in Table 2. The value of Q^* (k) for both k=12 and 24 are below the critical value of χ^2 for 7 and 19 degrees of freedom respectively. Also the graph of CUSUM lies within 5 percent limits. This indicates that the residuals are random and hence the model explaines the PPI of gasoline adequately. The goodness of fit statistics AIC, BIC, PEV, and R^2 indicate that the model fits the series well. The forecasting performance of this model for this series is quite good. The observed series lies within 95 percent intervals both one-step ahead and multi-step ahead. The post sample predictive F test and the post sample prediction error sums of squares indicate reasonably good forecast performance of this model. Low estimates of ARM and RMSR in Table 2 together with low standard error of revision in figure 2 suggests that this model is acceptable for seasonal adjustment of PPI of gasoline.

3. PPI of Crude Petroleum (Domestic): In the case of this series, the one period and two period lagged values of this series itself were used as explanatory variables. Since ninety-six observations are used to estimate the seasonal adjustment model, the small sample bias is the esitmation of the coefficient of these variables (if coefficients are assumed to be constant) may not be serious. The same two intervention variables, used in the previous two cases, are used to account for the effects of external interventions of January 1981 and January 1986. The estimates of the hyer-parameters indicate that the seasonal component, the coefficients of the two lagged dependent variables and the coefficiet of the first intervention variable are non-stochastic; however, the coefficient of the second intervention variable is stochastic because of the high value of the relative variance of

the corresponding error. The relative variance of the error corresponding to the trend component is small which may indicate some randomness in the trend.

Since the economic time series are highly autocorrelated, the presence of lagged dependent variables as explanatory variables is expected to produce very high coefficient of determination. But this is not true in this case because the interventions may have distorted the auto-correlation structure of this series. The value of Ljung-Box Q^* (k) in table 2 and the CUSUM graph in figure 3 indicate that the model is adequate to explain this series. In addition, the goodness of fit statistics AIC, BIC, PEV, and σ_ε^2 seem quite reasonable. However the post sample forecasting performance of this model is poor. This can be seen from the movements of the observed series in the 95 percent prediction intervals both one-step ahead and multi-step ahead. In the latter case, the observed series is outside the prediction interval for most of the post sample period. The post sample prediction error sum of squares are very large which adds uncertainty to the performance of this model. The low value of the estimates of revision indicate that the model is adequate to seasonally adjust this series in the presence of interventions.

4. SUMMARY AND CONCLUSIONS

This paper analyses the empirical results of estimating a structural model for seasonal adjustment in the presence of external interventions. Over all the performance of this model is quite good as a tool of seasonal adjustment in the presence of interventions. The effects of interventions and seasonality are simultaneously estimated and accounted for. The forecasting performance of this model especially in the case of CPI and PPI of gasoline series is quite remarkable. Probably experiments with alternative specifications of trend and seasonality might improve this model even more.

REFERENCES

Anderson, B.D.O., and Moore, J.B. (1979), Optimal Filtering, Englewood Cliffs, N.J.: Prentice Hall.

Box, G.E.P., and Tiao G.C. (1975), "Intervention Analysis with Applications to economic and Environmental Problems," Journal of the American Statistical Association, 70, 70-79.

Dempster, A.P., N.M. Laird and D.B. Rubin (1977), "Maximum Likelihood from Incomplete Data Via the EM Algorithm," Journal of Royal Statistical Society, Ser B., 39, 1-38.

Fletcher, R. (1980), Practical Methods of Optimization, Vol. 1, New York, John Wiley & Sons.

Harvey, A.C. (1981), Times Series Models, Oxford, U.K. Philip Allan Publishers Limited.

(1984), "A Unified View of Statistical Forecasting Procedures," Journal of Forecasting, 3, 245-275.

(1990), Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge, Cambridge University Press.

Harvey, A.C., and Todd, P.H.J. (1983), "Forecasting Economic Time Series With Structural and Box-Jenkins Models: A Case Study," Journal of Business & Economic Statistics, 1, 299-306.

Jazwinski, A.H. (1970), "Stochastic Processes and Filtering Theory", New York, Academic Press.

Kalman, R.E. (1960), "A New Approach to Linear Filtering and Prediction Problems," Trans. ASME, Journal of Basic Engineering, Ser. D, 80, 35-45.

Kitagawa, G., and W. Gersch (1984), "A Smoothness Priors-State Space Modelling of Time Series with Trend and Seasonality", Journal of American Statistical Association, 79, 378-389.

Shumway, R.H. and D.S. Stoffer (1982), "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm," Journal of Time Series Analysis, 3, 253-264.

Shumway, R.H. (1988), "Applied Statistical Time Series Analysis", Englewood Cliffs, N.J.: Prentice Hall.

Schweppe, F. (1985), "Evaluation of Likelihood Functions for Gaussian Signals", IEEE Transactions on Information Theory, 11, 61-70.

Table 1. Estimates of Hyper-parameters of the Seasonal Adjustment Model

Series	Model Error	Trend	Seasonal	Intervention Variables		Explanatory Variables	
				I_1	I_2	X_1	X_2
CPI-Gasoline	0.7726	0.0026	0.0000	0.0000	2.2464	0.0003	-
PPI-Gasoline	0.8172	0.0048	0.0000	1.2466	7.3914	0.0006	-
PPI-Crude Petroleum	0.8457	0.0355	0.0000	0.0001	9.1639	0.0002	0.0000

Table 2. Adequacy, Goodness of Fit and Forecasting Performance of the Seasonal Adjustment Model

Series	Ljung-Box $Q^*(k)$		AIC	BIC	PEV	R^2		Post-Sample	Post-Sample Predictive F-Test df=(12,96)		Post-Sample Prediction Error Sum of Squares		Estimates of Revision	
	12	24				Diff.	Seas.	PEV	One- Step Ahead	Multi-Step Ahead	One-Step Ahead	Multi-Step Ahead	AMR	RMSR
CPI-Gasoline	1.88 (7)	4.08 (9)	254.22	267.04	3.98	0.47	0.50	39.45	0.49	0.34	25.30	54.14	0.13	0.15
PPI-Gasoline	2.93 (7)	11.36 (19)	337.50	350.32	10.32	0.19	0.23	124.70	0.46	0.19	61.87	149.14	0.35	0.45
PPI-Crude Petroluem	3.43 (6)	6.24 (18)	339.62	355.01	10.61	0.31	0.33	181.88	1.34	2.80	174.74	1981.72	0.74	0.86

Note: Numbers in parentheses under the values of Q*(k) refer to the degrees of freedom. k is the number of lags used in computing Q*.

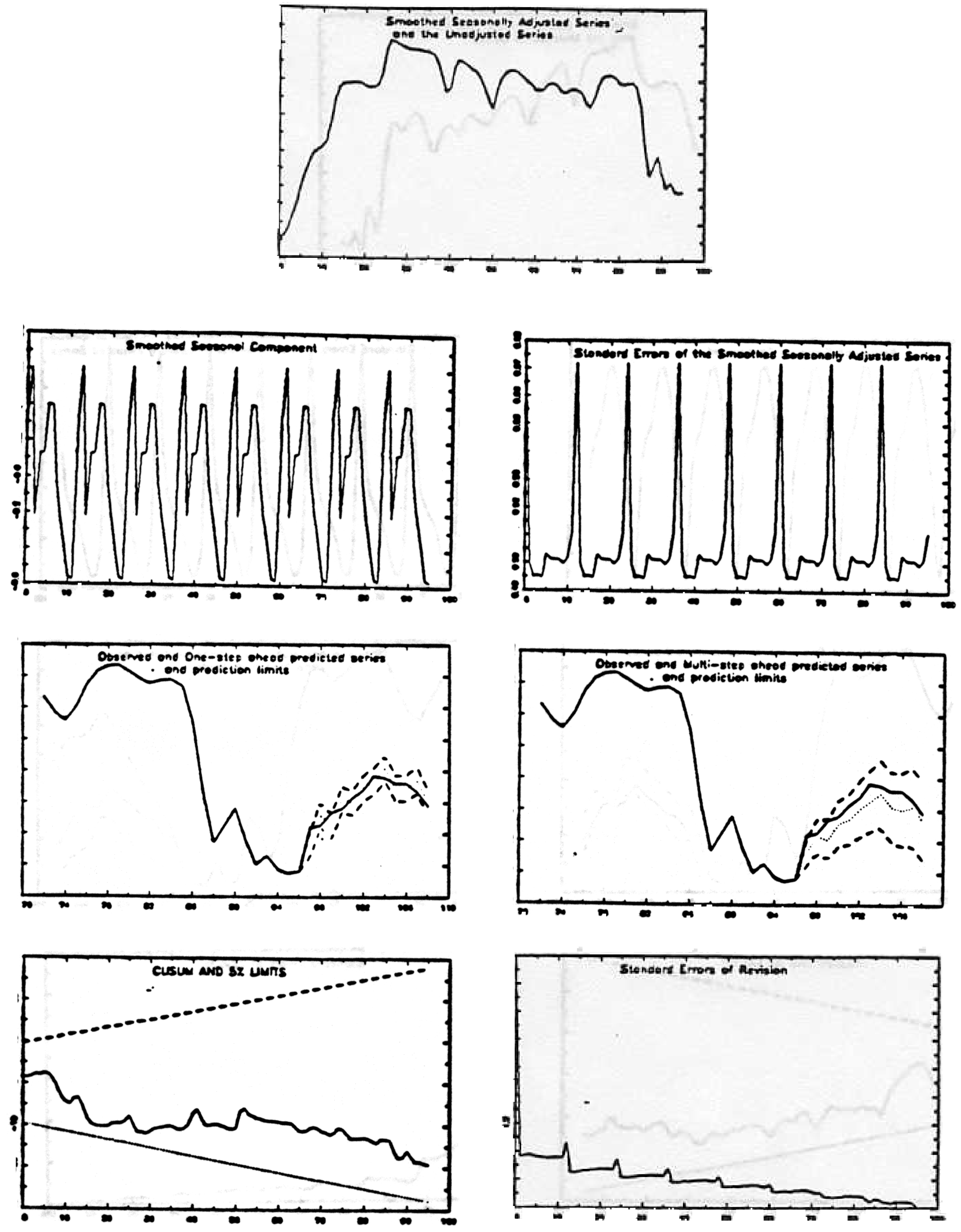

Figure The Consu Price Index of Gasoline

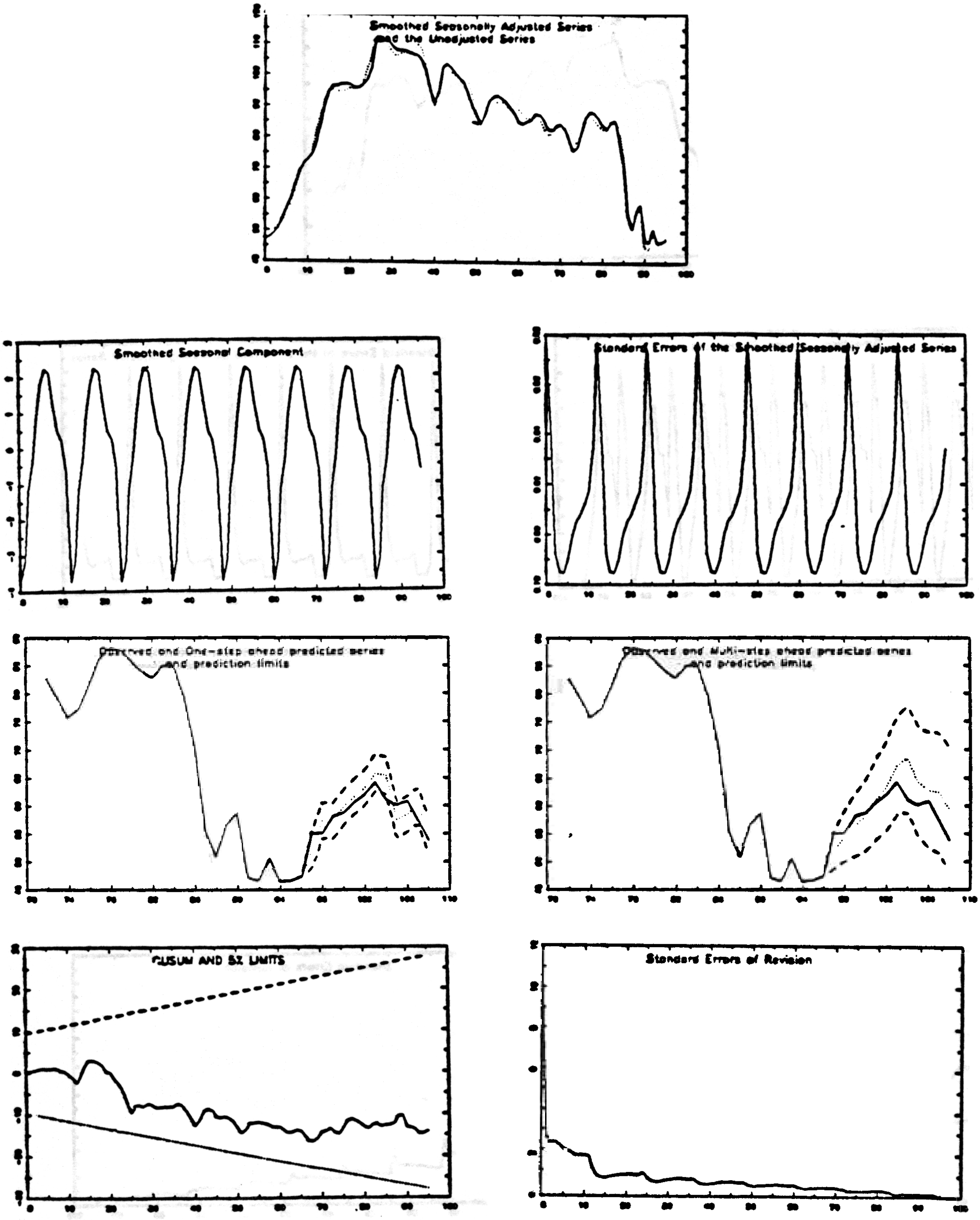

Figure 2. The Producer Price Index of Gasoline

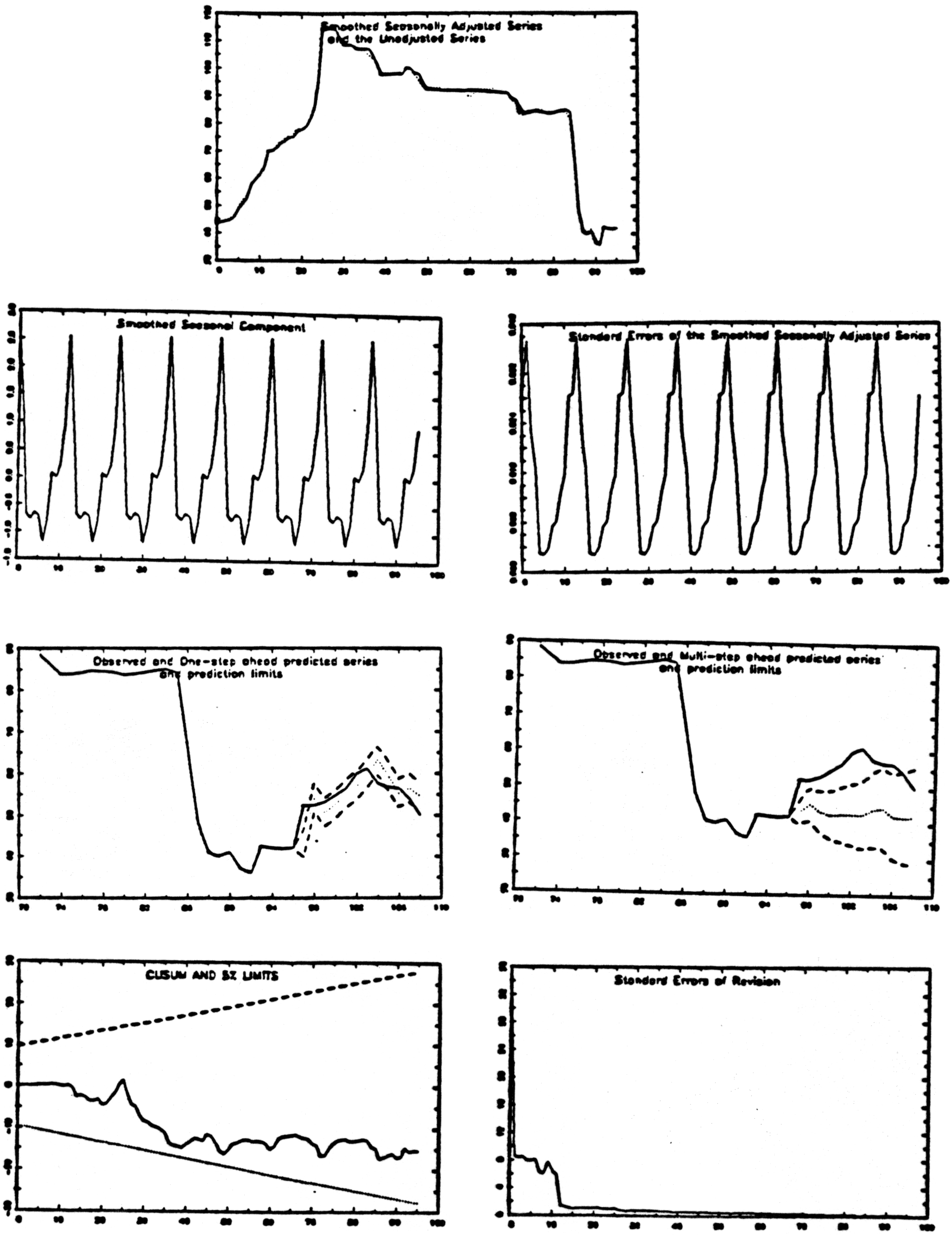

Figure 3. The Producer Price Index of Crude Petroleum