

NBER WORKING PAPER SERIES

SEASONAL ADJUSTMENT WITH MEASUREMENT ERROR PRESENT

Jerry A. Hausman

Mark Watson

Working Paper No. 1133

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

May 1983

Both authors thank the NSF for research support. Hausman has received support from the NBER. This paper was written while he was on leave at Harvard, 1982-1983. The research reported here is part of the NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

SEASONAL ADJUSTMENT WITH MEASUREMENT ERROR PRESENT

Abstract

Seasonal adjustment procedures attempt to estimate the sample realizations of an unobservable economic time series in the presence of both seasonal factors and irregular factors. In this paper we consider a factor which has not been considered explicitly in previous treatments of seasonal adjustment: measurement error. Because of the sample design used in the CPS, measurement error will not be a white noise process, but instead it will be characterized by serial correlation of a known form. We first consider what effect the serially correlated measurement error has on estimation of the non-seasonal component in seasonal adjustment models. We also consider the effect of measurement error on the widely used seasonal adjustment process X11. X11 which is the seasonal adjust procedure used by the BLS will implicitly reduce the effect of measurement error because of the averaging process used. However, this treatment will not be optimal in general. We therefore specify a seasonal adjustment model which takes explicit account of the measurement error. For examples on the unemployment rate, we find that X11 does almost as well as the optimal filter on some series but its efficiency is less than 10% for the teenage unemployment series. We also find that optimal treatment of the measurement error which accounts for the serial correlation can reduce the overall mean square error of the seasonally adjusted series below the variance of the measurement error which is often used as the benchmark for the sampling procedure.

Jerry A. Hausman
Massachusetts Institute of Technology
Department of Economics - E52-271A
Cambridge, MA 02139
617-253-3644

Mark Watson
Harvard University
(617) 495-5148

SEASONAL ADJUSTMENT WITH MEASUREMENT ERROR PRESENT

Jerry Hausman - MIT and NBER

Mark Watson - Harvard

Seasonal adjustment procedures attempt to estimate the sample realizations of an unobservable economic time series in the presence of both seasonal factors and irregular factors. Only the sum, or the product, of the three components is observed. Within an additive framework the procedures may be represented as

$$(1) \quad x_t = n_t + s_t + e_t$$

where n_t is the non-seasonal component, s_t is the seasonal component, and e_t is a stochastic disturbance. Seasonal adjustment procedures differ in their treatment of each of the components in equation (1). Many procedures treat s_t as deterministic up to some unknown parameters and set e_t to zero. The unknown parameters of s_t are estimated and the n_t series is then determined by $n_t = x_t - s_t$. Other seasonal adjustment procedures treat s_t as stochastic and after estimation by signal extraction methods, again determine n_t by subtraction. In either case, little attention is given to the properties of the stochastic disturbance, e_t . Now, e_t can arise from two sources as in most econometric models. The first source of e_t can be model specification error which we will test for subsequently. The other source of e_t can be measurement error. The concept of measurement error is somewhat non-standard here since we do not have a

'real' series which we are attempting to measure. That is, n_t does not exist apart from the model specification of the components of equation (1). But it still makes sense to consider an underlying series, say x_t^* , where we can only observe x_t which contains measurement error. In this reformulated model we would have $x_t = x_t^* + e_t$ so that n_t and s_t determine x_t^* . In the presence of measurement error and thus estimation will need to account for the properties of e_t .

The usual model of measurement error in statistics and econometrics is one of white noise. However, because of government data collection procedures which use overlapping sample designs, the measurement error can be serially correlated. For instance, the sample design used by the Bureau of Labor Statistics (BLS) to measure unemployment consists of 8 sub-samples, seven of which have been included in previous samples. This sample design is chosen presumably to induce positive correlation across months which leads to a reduction of measurement error in estimated changes in the unemployment rates. So far as we know, no seasonal adjustment procedure takes explicit account of this measurement error which becomes confounded with the seasonal component s_t because of the 50% overlap from year to year induced by the BLS sample design.

The natural question is what effect this serially correlated measurement error has on estimation of the non-seasonal component n_t from equation (1). This question cannot be answered apart from an examination of the goals of seasonal adjustment in the specification of the components of equation (1) since the true n_t is not observable and does not exist apart from a specific model formulation. Nerlove et al. (1979) and Pierce (1976) discuss the question of the goal of seasonal adjustment, with Pierce setting forth a set of assumptions to guide his procedure. We choose the minimum mean square error estimator of n_t given

hypothesized ARMA models for both n_t and s_t . Thus we allow for both deterministic and stochastic components in the model for both components. When measurement error is ignored the estimates of both n_t and s_t will be determined by the similarity of their spectra and that of the measurement error. Our empirical examples demonstrate that the increase in mean square error can be substantial in some cases while in other cases the increase can be quite small.

Besides comparing optimal seasonal adjustment procedures which ignore measurement error to an optimal procedure which takes account of measurement error, we also consider the effect of measurement error on X11. X11 is the seasonal adjustment procedure used by the BLS and many other government agencies. Since the X11 procedure uses a moving average of measured unemployment it will implicitly reduce the effect of measurement error because of the averaging procedure used. However, this treatment will not be optimal because the explicit form of the measurement error has not been incorporated. Our empirical examples demonstrate again that the effect of measurement error on the efficiency of the X11 procedure can vary widely across series. The efficiency of X11 with respect to the optimal filter which takes account of measurement error is found to be as low as .09 in our examples so that consideration of measurement error is an important topic. Also, of interest is that we calculate the root mean square error of X11 so that one can assign confidence intervals to estimated levels or changes of the official unemployment rate.

The plan of the paper is as follows. In Section 2 we consider the signal extraction problem for n_t when n_t , s_t , and e_t have been assigned ARMA specifications. We derive expressions for the effect of omission of general

measurement error on otherwise optimal filters. In Section 3 we discuss the sampling procedure used by the BLS and derive the specification of measurement error for this particular sample design. We then estimate the unknown parameters of the model by maximum likelihood estimation via the Kalman filter. We estimate models for both the overall unemployment rate as well as the teenage unemployment rate. Since the latter series has considerably more variance, it offers a good comparison to the overall series. We test for the specification of our model both by using diagnostic statistics and by seeing how well it can predict the estimates from a model which ignores measurement error. Then in Section 4 we compare the optimal procedure with X11 and with an optimal filter which ignores measurement error. For the overall unemployment rate the three procedures given quite similar results with the efficiencies varying from .71 to .98. For teenage unemployment the three procedure give very dissimilar results with the efficiency of the optimal filter which ignores measurement error as low as .2 and X11 as low as .09. In the conclusion we discuss other potential applications of the model as well as implications for sample design.

Section 2

In the last section we defined seasonal adjustment, in the context of an additive components model, as a procedure for estimating the unobserved non-seasonal component using data on the observed composite series. Measurement error simply adds another component to the model. The observed series is now "contaminated" not only with seasonality but with measurement error as well. The presence of this extra component reduces the precision with which the non-seasonal component can be estimated. If part of the non-seasonal component is deterministic, but a function of unknown parameters (e.g., a time trend), measurement error will decrease the precision of the estimated parameters. In general, estimates of these parameters will remain consistent so that this problem diminishes as the sample size grows. On the other hand, measurement error will have an effect on the estimate of the stochastic part of the non-seasonal which no amount of data can eliminate. This fact is easily demonstrated.

We will ignore any deterministic components and assume that the non-seasonal component, n_t , and the seasonal component, s_t are generated by independent stationary and invertible moving average processes of the form

$$(2) \quad n_t = \theta_n(B)\varepsilon_t$$

$$(3) \quad s_t = \theta_s(B)\eta_t$$

where ε_t and η_t are white noise and $\theta_n(B)$ and $\theta_s(B)$ are polynomials in the

backshift operator B . (The stationarity assumption is not crucial for what follows. An investment in additional notation would allow us to assume that $\Delta_n(B)n_t$ and $\Delta_s(B)s_t$ were stationary, where $\Delta_n(B)$ and $\Delta_s(B)$ are differencing operators with no common factors. This generalization is discussed in Cleveland and Tiao (1976) and Pierce (1979).) We will assume that the measurement error, e_t , is generated independently of n_t and s_t by another stationary and invertible moving average process

$$(4) \quad e_t = \theta_e(B)\xi_t$$

where ξ_t is white noise. This allows e_t to be serially correlated, a characteristic of the measurement error processes for the two series discussed in the next section. Finally, let x_t be the series composed of only n_t and s_t , and let x_t^* be composed of n_t , s_t , and e_t .

$$(5) \quad x_t = n_t + s_t$$

$$(6) \quad x_t^* = n_t + s_t + e_t.$$

The problem of estimating n_t given x_t or x_t^* (i.e., seasonally adjusting x_t or x_t^*) can be solved using well known signal extraction techniques. In both cases n_t is estimated from "noisy" measurements. When x_t is observed the noise is s_t ; when x_t^* is observed the noise is the sum of s_t and e_t .

In general, the signal extraction procedure for n_t depends on the time subscript t . When t is not near the beginning or the end of the sample the procedure for forming the minimum linear mean square estimates of n_t and n_{t+1} are very similar. Indeed, they can be made arbitrarily close by extending the sample

at both ends. To avoid the notational burden of extra time subscripts we will assume that a complete realization of the observed series is available. If the observed series is uncontaminated with measurement error the linear minimum mean square estimate of n_t is given by

$$(7) \quad n_t = V(B)x_t$$

where

$$(8) \quad V(B) = \sum_{i=0}^{\infty} V_i(B^i + B^{-i})$$

and the coefficients V_i can be formed from

$$(5) \quad V(Z) = \frac{\sigma_{\varepsilon}^2 \theta_n(Z) \theta_n(Z^{-1})}{\sigma_{\varepsilon}^2 \theta_n(Z) \theta_n(Z^{-1}) + \sigma_{\eta}^2 \theta_s(Z) \theta_s(Z^{-1})}$$

(See Whittle [1963] or Grether and Nerlove [1970] for the derivation of the filter $V(B)$.) The seasonal adjustment error associated with the filter $V(B)$ is given by

$$(10) \quad a_t = n_t - \hat{n}_t = (1 - V(B))n_t - V(B)s_t,$$

and the variance of a_t of this seasonal adjustment error is

$$(11) \quad \sigma_a^2 = \int_{-\pi}^{\pi} \frac{f_n(w) f_s(w)}{f_x(w)} dw$$

where $f_n(w)$, $f_s(w)$, and $f_x(w)$ are the spectra of n_t , s_t , and x_t respectively.

When the observed series is contaminated by measurement error as well as seasonal noise the optimal seasonal adjustment filter changes. In this case the

linear minimum mean square estimate of n_t is given by

$$(12) \quad n_t^* = v^*(B)x_t^*$$

where

$$(13) \quad v^*(Z) = \frac{\sigma_\varepsilon^2 \theta_n(Z) \theta_n(Z^{-1})}{\sigma_\varepsilon^2 \theta_n(Z) \theta_n(Z^{-1}) + \theta_\eta^2 \theta_s(Z) \theta_s(Z^{-1}) + \sigma_\xi^2 \theta_e(Z) \theta_e(Z^{-1})}$$

The seasonal adjustment error in this case is

$$(14) \quad a_t^* = n_t - n_t^*$$

and

$$(15) \quad \sigma_{a^*}^2 = \int_{-\pi}^{\pi} \frac{f_n(w)(f_s(w) + f_e(w))}{f_{x^*}(w)} dw$$

where $f_e(w)$ and $f_{x^*}(w)$ are the spectra of e_t and x_t^* . The increase in the variance of the seasonal adjustment caused by the measurement error is easily shown to be

$$(16) \quad \sigma_{a^*}^2 - \sigma_a^2 = \int_{-\pi}^{\pi} f_e(w) \left(\frac{f_n(w)}{f_x(w)} \right) \left(\frac{f_n(w)}{f_{x^*}(w)} \right) dw.$$

Since

$$(17) \quad f_x(w) = f_n(w) + f_s(w)$$

and

$$(18) \quad f_{x^*}(w) = f_n(w) + f_s(w) + f_e(w)$$

the terms in parenthesis are both less than one so that the increase in mean

square error is less than the variance of e .

In the discussion of optimal seasonal adjustment procedures found in the literature measurement error is usually (always) ignored. The literature has focused on the construction and evaluation of seasonal filters when the models for n_t and s_t are known (e.g., Grether and Nerlove [1970], Cleveland Tiao [1976]) or the estimation of models for n_t and s_t using the composite series (e.g., Pierce [1976], Engle [1976], Burman [1980], Hillmer and Tiao [1982].) When measurement error is ignored seasonal adjustment filters formed using the actual or estimated models for n_t and s_t will be sub-optimal. In the first case the optimal filter $V(B)$, for seasonal adjustment of x_t is used to seasonally adjust x_t^* . This produces a seasonally adjusted series

$$(19) \quad \tilde{n}_t = V(B)x_t^*$$

with seasonal adjustment error

$$(20) \quad \tilde{a}_t = n_t - \tilde{n}_t.$$

The increase in mean square error which arises from the use of this sub-optimal filter is given by

$$(21) \quad \sigma_{\tilde{a}}^2 - \sigma_{a^*}^2 = \int_{-\pi}^{\pi} f_e(w) \left(\frac{f_e(w)}{f_{x^*}(w)} \right) \left(\frac{f_n(w)}{f_x(w)} \right)^2 dw.$$

When the processes generating the measurement error and n_t are very dissimilar, so that $f_e(w)$ is large when $f_n(w)$ is small and visa-versa, the increase in mean square error will be small. This is unlikely to be true for most series where we would expect both n_t and e_t to be positively serially correlated. In this case

the spectra of the two series have a similar shape, and the increase in mean square error can be as large as

$$\sigma_e^2 \left(1 + \frac{\sigma_n^2}{\sigma_e^2}\right)^{-1}.$$

When measurement error is ignored and the contaminated data, x_t^* , are used to estimate models for n_t and s_t the comparison of filters is less straightforward. The comparison depends in an obvious way on whether the measurement error is attributed to the seasonal or to the non-seasonal component. At one extreme the measurement error is purely seasonal and is totally attributed to the seasonal component. The seasonal adjustment filter in this case will filter the "seasonal" series $s_t + e_t$ from x_t^* leaving an estimate of n_t . This is precisely what the optimal filter does. At the other extreme the measurement error is not seasonal (i.e., its spectrum has no peaks near the seasonal frequencies) and is incorrectly attributed to the non-seasonal component. The seasonal adjustment will now be an estimate of the filter which extracts $(n_t + e_t)$ from x_t^* which yields a seasonally adjusted series

$$(22) \quad \bar{n}_t = \bar{V}(B) x_t^*$$

with error

$$(23) \quad \bar{a}_t = n_t - \bar{n}_t.$$

The increase in mean square error caused by the misspecification of the model (i.e., ignoring e_t) is

$$(24) \quad \frac{\sigma^2}{a} - \sigma_{a^*}^2 = \int_{-\pi}^{\pi} f_e(w) \left(\frac{f_e(w)}{f_{x^*}(w)} \right) dw.$$

Again the exact magnitude of this increase depends on the processes generating n_t , s_t , and e_t .

In the next section we estimate models for two economic time series which are measured with error. We estimate models incorporating and ignoring the measurement error. These models are used to construct seasonally adjusted series. The mean square error of these series are then calculated and compared.

Section 3

One of the most closely watched indicators of macroeconomic performance is the civilian unemployment rate. Each month this rate is estimated by the Bureau of Labor Statistics using a rotating sample composed of approximately 56,000 households. The data exhibit clear seasonal behavior. The unadjusted rate in January and June of each year is roughly one percentage point higher than the rate in May. This empirical regularity suggests the presence of a seasonal component. Since the data published by the BLS are estimates constructed from a sample, they also contain a sampling error or measurement error component.

In this section we present a model which decomposes the series into three components - a non-seasonal component, a seasonal component, and a measurement error component. Using the observed data we estimate the parameters of the model. The estimated model is then used to construct the optimal seasonal adjustment filter. We compare this filter to one which we estimate ignoring the measurement error. The exercise is repeated for another series, the teenage unemployment rate. This series exhibits more dramatic seasonal behavior, and is subject to more severe measurement error.

We begin by describing the sample design used by the BLS. Each month the BLS surveys eight sub-samples, each composed of approximately 7,000 households. Seven of these sub-samples have been included in some past survey and one sub-sample is new. The new sub-sample is included in the survey for four months, left out of the survey for the next eight months, and then included for four final months. This rotation procedure produces a 75% overlap in the sample from month to month and a 50% overlap from year to year.

This sample design produces measurement error which may be serially correlated. The degree of serial correlation will depend on the "memory" in the measurement error for any sub-sample. Given random sampling we can assume that errors across sub-samples are uncorrelated, but we want to allow some fraction of the error in any sub-sample to persist from period to period. To capture this persistence we assume that the error for the i 'th sub-sample at time t is

$$(25) \quad e_{it} = \gamma_i + w_{it}$$

where γ_i is a time invariant random effect with mean zero and variance σ_γ^2 , and w_{it} is a white noise error, uncorrelated with γ_i , and has variance σ_w^2 . If we choose our indices so that sub-sample "i" is surveyed for the first time at time $t = i$, then the total measurement error at time t is given by

$$(26) \quad e_t = (1 + B^{12})(1 + B + B^2 + B^3)\gamma_t + \xi_t$$

where ξ_t is the sum of the eight white noise terms, e_{it} .

This model produces a very parsimonious representation of the process generating the measurement error. The rotation incorporated in the sample design implies that error follows an MA (15) process. The size of the sample determines the standard deviation of the sampling error. (It is approximately .12% for the civilian unemployment rate and .60% for the teenage unemployment rate.) Our model represents this MA(15) process with known variance in terms of only one unknown parameter, $\sigma_\xi^2/\sigma_\gamma^2$.

While much is known about the process generating the measurement error from the sample design we have little a priori knowledge concerning the processes generating the non-seasonal and seasonal components. There are some "common

sense" constraints that seem reasonable to impose on these processes. First, the deterministic component of the seasonal should sum to zero over any twelve month period. Second, the stochastic component of the seasonal should have non-zero autocorrelations only at the seasonal lags, so that it represents seasonal noise. Third, the deterministic component of the non-seasonal should not be seasonally periodic. Analogously, its stochastic component should not have a spectrum with extra power at the seasonal frequencies.

Unfortunately these four constraints are not sufficient to identify the seasonal and non-seasonal processes, even if the process generating their sum is known. There is little problem in identifying deterministic components and the autoregressive portion of the stochastic processes. The problem arises in the identification of the moving average portions of the model. This lack of identification has led to various normalizations and canonical forms (e.g., see Pierce [1976], Burman [1980], Hillmer and Tiao [1982].) In this application we found that low order autoregressive models adequately described the data, so that identification was not a problem. In other applications this may not be the case so that a "minimal extraction principle", as in Pierce [1976], or some other normalization would have to be employed.

Our model for the deterministic part of the seasonal was very simple. We chose a flexible form consisting of twelve monthly constants which summed to zero. We did not include a non-seasonal deterministic component; rather we assumed the non-seasonal to be generated by an integrated autoregressive process:¹

¹ More elaborate models are certainly possible. Within our framework we can allow any of the components to depend on observed variables. The non-seasonal component, for example, could be related to changes in GNP or an index of real wages as well as to its own lagged values. We have chosen to use a univariate model for n_t so that our results are directly comparable with other widely used "auto-adjustment" procedures.

$$(27) \quad \phi(B)(1 - B)n_t = \varepsilon_t$$

where the roots of $\phi(Z)$ are outside the unit circle. We also allowed the stochastic seasonal component to be generated by an autoregressive process

$$(28) \quad \psi(B^{12})s_t = \eta_t$$

In all models that are presented below we found

$$(29) \quad \phi(B) = (1 - \phi_1 B - \phi_2 B^2)$$

and

$$(30) \quad \psi(B^{12}) = (1 - \psi B^{12})$$

adequately described the data. Combining the non-seasonal, seasonal, and the measurement error we have a model for the observed unemployment rate

$$(31) \quad x_t^* = d_t' \beta + n_t + s_t + e_t$$

where d_t is (12 x 1) vector of monthly dummy variables and β is a (12 x 1) vector of unknown constants which sum to zero. The unknown parameters which must be estimated are β , ϕ_1 , ϕ , σ_ε^2 , σ_η^2 , $\sigma_\xi^2/\sigma_\gamma^2$, and the initial value for non-seasonal component n_0 .

To estimate the unknown parameters of the model we assumed that the disturbances were normally distributed and calculated maximum likelihood estimates. A very useful computational device for forming the likelihood function and for carrying out seasonal adjustment is the Kalman filter.¹

¹ Engle [1976] uses the Kalman filter in a similar application. See also Pagan [1975] and Harvey [1981].

The Kalman filter recursively calculates various estimates including the current estimates of the components as well as

$$(32) \quad x_t^*|_{t-1} = E[x_t^*|x_{t-1}^*, x_{t-2}^*, \dots, x_g^*]$$

and

$$(33) \quad h_t = \text{var}[x_t^*|x_{t-1}^*, x_{t-2}^*, \dots, x_1^*].$$

Since the disturbances are normally distributed the innovations in x^* , $x_t^* - x_t^*|_{t-1}$, are independent and normally distributed with mean zero and variance h_t . The likelihood function can then be formed in terms of these innovations. A discussion of techniques for maximizing this likelihood function can be found in Watson and Engle [1983].

The parameters of the models for the civilian unemployment rate and the teenage (age 16-19) unemployment rate were estimated using data from 67:1 to 83:1. The results are shown in Table 1 in the column labeled Model 1.

TABLE 1

	Civilian Unemployment Rate		Teenage Unemployment Rate	
	Model 1	Model 2	Model 1	Model 2
ϕ_1	.264 (.158)	.239	.278 (1.302)	-.153
ϕ_2	.366 (.117)	.317	.408 (1.038)	.108
σ_ε^2	.028 (.009)	.036	.057 (.1032)	.411
ψ	.525 (.383)	.566	.758 (.114)	.678
σ_η^2	.004 (.005)	.008	.106 (.060)	.167
σ_ξ^2	.0021(.005)		.190 (.061)	
σ_γ^2	.0016(.004)		.021 (.007)	
Q(d.s.)	25.6(18)	22.7(20)	26.5(18)	27.2(20)

Note: Asymptotic standard errors in parentheses. Coefficients on twelve seasonally dummy variables were also estimated.

The last entry in this column is the Box-Pierce Q-statistic followed by the degrees of freedom.¹ The Q-statistics suggest that we have committed no grievous dynamic misspecification; neither statistic is significant at the 5% level. The point estimates for the parameters of the non-seasonal processes are similar for both series. They are much more precisely estimated for the civilian unemployment rate. This series is subject to a much smaller seasonal effect (the variance of s_t is .005 for the civilian unemployment rate and .250 for the teenage unemployment rate) and much less measurement error. The persistence of the sampling error is also much different across the two series. The persistent random effect, γ , accounts for 85% of the variance in the civilian unemployment rate measurement error but only 47% of the measurement error for the teenage unemployment rate.

Charts 1 and 2 show the estimated decomposition of the observed series. These estimates were formed using the optimal signal extraction filter constructed from the parameter values in Table 1. The seasonal and measurement error variation is much more dramatic for the teenage unemployment rate. The estimated seasonal factor for the teenage unemployment rate varies from 4.8% to -2.7% and from .75% to .59% for the civilian unemployment rate. Teenage unemployment rate sampling error varies from .95% to -.96%, and civilian unemployment rate sampling error varies from .10% to -.11%.

¹Our component models implies an ARIMA model for the observed series. The parameters of the ARIMA model are functions of $(\phi_1, \phi_2, \psi, \sigma_\gamma^2, \sigma_\xi^2, \text{ and } \frac{\sigma_\epsilon^2}{\sigma_\eta^2})$.

As a first step in constructing seasonal adjustment filters which ignore measurement error we re-estimated the models leaving out the sampling error component. The results are presented in Table 1 in the columns labeled Model 2. The measurement error in the civilian unemployment rate is small and neglecting it has only a small effect on the model. The estimated seasonal process is essentially unchanged, and the non-seasonal process has changed slightly to capture the measurement error dynamics. The estimated process for the seasonal component of the teenage unemployment rate is slightly changed. The variance of the seasonal component has increased to capture some of the effect of the measurement error. The process for the non-seasonal component has changed markedly. The variance of its driving noise, ε_t , is over seven times larger than in Model 1, and the point estimates of ϕ_1 and ϕ_2 are different.

While Model 1 and Model 2 may be quite different in one sense they should be very similar in another sense. Both models are descriptions of the same observed series, so that they should imply the same aggregate dynamic model. Comparing the implied aggregate model from both models can therefore serve as a check on our specification. The stochastic portion of Model 1 is $u_t = n_t + s_t + e_t$, and let us write the stochastic portion of Model 2 as $u_t = n_t^* + s_t^*$.

Since $(1 - B)u_t$ is stationary it has a Wold representation

$$(34) \quad (1 - B)u_t = \lambda(B)a_t$$

where a_t is white noise with unit variance so that

$$(35) \quad (1 - B)(n_t + s_t + e_t) = (1 - B)(n_t^* + s_t^*) = \lambda(B)a_t.$$

Using the estimated parameters of Model 1 and Model 2 we have calculated the implied moving average parameters $\lambda_0, \lambda_1, \dots$. These are plotted in Chart 3 for the civilian unemployment rate and in Chart 4 for the teenage unemployment rate. The implied models are very similar. That is, when we 'predict' the parameters of Model 2, which ignores measurement error, in terms of the parameters of Model 1 the results are quite close. This finding should increase the confidence placed in Model 1 since we are able to predict well the properties of the 'misspecified model', Model 2, as a function of the estimated parameters of Model 1 which is hypothesized to be correct.

A different "optimal" seasonal adjustment is associated with both Model 1 and Model 2. The Model 1 filter constructs an optimal estimate of the true non-seasonal component, n_t , while Model 2 filter constructs an optimal estimate of the misspecified non-seasonal component, n_t^* . Given a complete realization of the observed series both of these filters will be of the form

$$(36) \quad \pi(B) = \pi_0 + \sum_{i=1}^{\infty} \pi_i (B^i + B^{-i}).$$

In Chart 5 we plot the filter weights implied by Model 1 and Model 2 for the civilian unemployment rate. The seasonal adjustment filters are quite similar so that the filters will produce similar seasonally adjusted values for the civilian rate. Indeed in the next section we show that Model 2 filter produces an

estimate of n_t which is only slightly less precise than the Model 1 filter. In Chart 6 we compare the filter for the teenage unemployment rate. The filters are quite different for this series. The misspecified filter puts far too much weight on the current observation, and in part compensates for this by a large negative weight at the twelfth lead-lag.

We can analyze other seasonal adjustment in a similar manner. The most widely used seasonal adjustment procedure is produced by the Census X-11 program. While the filter constructed by this program contains non-linearities (e.g., adjustments for outliers) it can be well approximated for many series by the symmetric 84 term linear filter which is given in Wallis [1974]. Charts 7 and 8 compare this filter with the Model 1 filter for our series. For both series X-11 puts too much weight on the first observation and compensates for this with large negative weight at the seasonal lead-lags. This characteristic is much more dramatic for the teenage unemployment rate. A mere glance at these charts suggests that the increase in precision of the optimal filter from the X-11 filter will not be too large for the civilian unemployment rate, but may be quite large for the teenage rate. In the next section we calculate the mean square error associated with the X-11 procedure and find this to be the case.

Section 4

The unobserved components model of seasonality allows us to view a seasonal adjustment procedure as an estimation method, and a seasonally adjusted series as a sequence of estimates of the underlying non-seasonal component. Alternative seasonal adjustment procedures can be evaluated on the basis of how precisely they estimate the non-seasonal component. In this section we compare various seasonal adjustment procedures for the two series analyzed in the previous section using mean square error as the measure of precision. In particular, we calculate the m.s.e. of the estimates calculated by Census X-11 program. This comparison allows us to discuss the accuracy of the official seasonally adjusted series and the increase in accuracy that could be achieved using an optimal filter. We also calculate the increase in m.s.e. that arises from the measurement error component, both when it is ignored and when it is accounted for in an optimal manner.

It is useful to identify five different sources of seasonal adjustment uncertainty. First, there is a certain irreducible signal extraction uncertainty, that arises when both n_t and s_t are stochastic. This uncertainty makes it impossible to deduce the value of either when only their sum is observed. Next, there is additional uncertainty which arises from measurement error. Its presence increases the "noise" in the observed series. A further increase in uncertainty may arise from the use of a suboptimal seasonal adjustment procedure. There are also two sources of uncertainty which arise from the quantity of data available. Since the components are serially correlated, future values of the observed series will contain information on the value of

today's non-seasonal component. Data at time $t+1$, etc. can therefore be used to reduce the uncertainty surrounding n_t . Finally, since the parameters used in constructing seasonal adjustment filters may be unknown, there is uncertainty that arises from the use of estimated parameters.

In usual calculations of mean square error it is possible at some point in time to observe the realization of the variable being estimated. In calculating one-step-ahead forecast mean square error, for example, the variable being forecast is observed with a period lag. This set-up makes it possible to calculate a sample mean square error which incorporates all sources of error. In seasonal adjustment the case is somewhat different. The variable being estimated is never observed. Sample mean square error cannot be calculated. Mean square error can only be calculated or estimated if the processes generating the components are known or are estimated. It makes little sense to discuss the precision of an estimate of something that is not observed nor defined apart from a particular model specification¹. Our estimates of mean square error will be based on the models that we presented in the last section. We will assume that the specification used in Model 1 is correct.

If we assume that the data were generated by Model 1 with parameter values shown in Table 1 then the root mean square error of various seasonally adjusted series are straightforward to calculate. All of the seasonal adjustment procedures that we will consider first subtract a constant, c_t , from the observed series. The result is then passed through a linear filter, $\pi(B)$, to produce seasonally adjusted value. Consider for example a crude form of seasonal adjustment in which only the deterministic seasonality is removed. In this case c_t represents the deterministic seasonality which we denoted $d_t^1\beta$ in the last

¹Summers [1981] attempts such a discussion.

section and $\pi(B) = 1$. This procedure produces a seasonal adjustment error

$$(37) \quad b_t = n_t - \pi(B)(x_t - c_t) = -s_t$$

if the observed series is not measured with error and

$$(38) \quad b_t^* = n_t - \pi(B)(x_t^* - c_t) = -s_t - e_t$$

if the observed series is measured with error. An arbitrary c_t and $\pi(B)$ will yield seasonal adjustment errors

$$(39) \quad b_t = (1 - \pi(B))n_t - \pi(B)s_t - \pi(B)(d_t'\beta - c_t)$$

and

$$(40) \quad b_t^* = b_t - \pi(B)e_t.$$

Recall that n_t followed an integrated process so that a necessary condition for the m.s.e. of b_t to be finite is that $(1 - \pi(B))$ contain the factor $(1 - B)$ or equivalently that $\pi(1) = 1$. All seasonal adjustment filters that we consider will have this property. Given these expressions for b_t and b_t^* their m.s.e. can be easily calculated from the constants $(d_t'\beta - c_t)$ and autocovariances of $(1 - B)n_t$ and s_t .

The first row of Table 2 shows the root mean square error (r.m.s.e.) associated with the optimal filter applied to a complete realization of the series without measurement error. This model corresponds to the series n_t and error a_t given in Section 2. Notice that the irreducible r.m.s.e. is much larger for the teenage unemployment rate. The larger r.m.s.e. is caused by the large variance of the stochastic seasonal component for this series as compared

TABLE 2
Seasonal Adjustment Root Mean Square Error

	Civilian Unemployment Rate		Teenage Unemployment Rate	
	<u>Levels</u>	<u>Changes</u>	<u>Levels</u>	<u>Changes</u>
1. M1, no measurement error	.060	.075	.222	.173
2. M1, with measurement error	.114	.103	.387	.202
3. Deterministic adjustment, no meas. error	.074	.105	.499	.706
4. Deterministic adjustment, with meas. error	.141	.127	.780	.981
5. M2, with measurement error	.124	.104	.511	.451
6. X-11, no measurement error	.090	.091	.286	.337
7. X-11, with measurement error	.126	.122	.552	.680
8. Current adj. using M1 with meas. error	.131	.104	.529	.241
9. M1 with parameter uncertainty*	.124	.105	.391	.265
10. Current adj. with parameter uncertainty*	.139	.110	.540	.270

Notes:

*Varies over the sample period
Average value is presented.

to the component for the civilian unemployment rate.

For both series notice that the r.m.s.e. is considerably smaller than the r.m.s.e. of the measurement error component. The r.m.s.e. of the level of the civilian rate is roughly 1/2 the size the r.m.s.e. of its measurement error, and for the teenage rate the r.m.s.e. is only 1/3 its measurement error. This emphasizes the point that the seasonal adjustment procedures significantly reduces both sources of noise - seasonal and measurement error. It also

illustrates the point made by equations (15) and (16): while the size of the measurement error is important, its dynamic properties and those of n_t and s_t are equally important.

In row 2 of the table we show the r.m.s.e. associated with the optimal filter applied to a complete realization of the series measured with error. This model corresponds to the series n_t^* and error a_t^* given in Section 2. Again the figures for the teenage unemployment rate are larger than those for the civilian unemployment rate.

Comparing row 2 and row 1 of Table 2 shows the increase in r.m.s.e. due to measurement error is larger for the teenage rates. Recall, however, that the measurement error variance was .36 for the teenage rate as compared to .0144 for the civilian rate. The increase in m.s.e. for the level of the teenage rate is only 28% of its measurement error variance, while the increase in m.s.e. for the level of the civilian rate is 65%. The reason for this differential increase is shown in equation (16). The seasonal measurement error components are relatively unimportant in the civilian rate so that a large fraction of the spectrum of x and x^* is accounted for by the n_t . The terms $(f_n(w)/f_x(w))$ and $(f_n(w)/f_{x^*}(w))$ in equation (16) are, on average, larger for the civilian rate than the teenage rate.

The next two rows of Table 2 present the r.m.s.e. of the seasonally adjusted series when only the deterministic seasonal component is extracted. The estimator was calculated for the series with and without measurement error. Table 3 presents the efficiency of the optimal estimator relative to this estimator. The relative efficiency is much smaller for the teenage rate, again reflecting the importance of stochastic seasonality and measurement error for this series. The next row of Table 1 presents the results using the filter

TABLE 3

Relative Efficiency of Seasonal Adjustment Procedures

	<u>Levels</u>	<u>Changes</u>	<u>Levels</u>	<u>Changes</u>
Deterministic adjustment, no meas. error	.66	.51	.20	.06
Deterministic adjustment, meas. error	.65	.66	.25	.04
M2, meas. error	.85	.98	.57	.20
X-11, no meas. error	.44	.68	.60	.26
X-11, with meas. error	.82	.71	.49	.09

constructed from the Model 2 parameter estimates and applied to a complete realization of the series. This filter would be estimated if measurement error were present but was ignored. For the civilian unemployment rate the variance of the (change in the) measurement error is small relative to the variance of the (change in the) observed series, so that the increase in m.s.e. from the Model 1 filter is not large. (See equation (24).) The relative efficiency of this filter is quite high. In the case of teenage rate, measurement error is much more important, and the Model 2 filter produces a much less precise estimate than the Model 1 filter. The relative efficiency for estimating the change in n_t is only .20. Our finding emphasizes the potential importance that measurement error can have on seasonal adjustment procedures.

The plots presented at the end of the last section also suggested a more dramatic reduction in efficiency from the use of census X-11 for the teenage unemployment. We should point out the X-11 procedure does not subtract a deterministic component from the row series before filtering. The seasonally

adjusted series will therefore contain the deterministic component $\pi(B)d_t'\beta$, which is a weighted average of monthly constants. The X-11 filter is constructed so that the weights across months are very nearly equal. Therefore, $\pi(B)d_t'\beta$ is very nearly the sum of the seasonal constants, which is zero. Rows 6 and 7 of Table 2 present the r.m.s.e. for the linear approximations to the X-11 filter. The performance of the X-11 filter is close to the optimal filter for the level of civilian unemployment rate (a relative efficiency of .82),¹ but it performs very poorly for the change in the teenage rate (a relative efficiency of .09). Indeed for the change in the teenage rate one is better off applying the Model 1 filter to the series measured with error than to apply the X-11 filter to the series without measurement error.

The results thus far have assumed that a complete realization of the series was available so that two sided filters could be applied. When adjusting current values, or values from the recent past, symmetric two-sided filters cannot be used. It is possible to construct optimal one-sided filters and indeed the Kalman filter does just this. In row 8 of Table 2 we present the r.m.s.e. of currently adjusted values using the optimal filter. Comparing these figures with those in row 2 shows the value of future data for current seasonal adjustment. For both series future data decreases m.s.e. for the level of the series rather substantially (25% for the civilian rate, nearly 50% for the teenage rate), but has a smaller effect on the m.s.e. of the change in n_t .

The filter that we have constructed from our estimated Model 1 is only an

¹Our estimate of the r.m.s.e. of the X-11 process is approximately one-half of the estimate of Summers (1981). However, he failed to take account of the time series properties of the unemployment rate or of the X-11 filter.

estimate of the optimal filter. Imprecise estimates of the parameters will produce an imprecise estimate of the filter which in turn will produce an imprecise seasonally adjusted series. A bit more notation will make the discussion of the effects of parameter uncertainty easier. Let

$$(41) \quad \theta = (\phi_1, \phi_2, \sigma_\varepsilon^2, \psi, \sigma_\eta^2, \sigma_\gamma^2/\sigma_\xi^2)'$$

be the vector of unknown parameters used to construct the filter for extracting the stochastic components. The other unknowns are the deterministic seasonal components, β . Let θ_0 and β_0 denote the true values of the parameters, and θ and β denote their maximum likelihood estimates. Finally we write the optimal seasonally adjusted series, n_t , as a function of (θ, β) .

$$(42) \quad n_t(\theta, \beta) = \sum_{i=-\infty}^{\infty} \pi_i(\theta)(x_{t-1} - d'_{t-i}\beta).$$

Expanding our estimate, $n_t(\theta, \beta)$, about the true optimal estimate, $n_t(\theta_0, \beta_0)$, and subtracting from n_t we have

$$(43) \quad n_t - n_t(\theta, \beta) = n_t - n_t(\theta_0, \beta_0) - (\theta - \theta_0)' \sum_{i=-\beta}^{\infty} \frac{\partial \pi_i(\theta_0)}{\partial \theta} (x_{t-i} - d'_{t-i}\beta_0) \\ + (\beta - \beta_0)' \sum_{i=-\beta}^{\infty} \pi_i(\theta_0) d_{t-1}$$

as our seasonal adjustment error. The first term on the left-hand side is the error that arises from the optimal filter, the final two terms reflect parameter estimation error. As the sample grows these last two terms converge in probability to zero leaving only the first term. A large sample approximation to

the m.s.e. of $n_t - n_t(\theta, \beta)$ follows from the asymptotic disturbances of $\sqrt{T}(\theta - \theta)$ and $\sqrt{T}(\beta - \beta)$. We will approximate the m.s.e. of $n_t - n_t(\theta, \beta)$ using the covariance matrix of θ and β from the asymptotic normal distribution.

The information matrix between is easily shown to be block diagonal between θ and β so that

$$(44) \quad \text{m.s.e.}(n_t(\theta, \beta)) \approx \text{m.s.e.}(n_t(\theta_0, \beta_0)) \\ + \sum_{i=-\infty}^{\infty} (x_{t-i} - d'_{t-i}\beta_0)^2 \left(\frac{\partial \pi_i(\theta_0)}{\partial \theta} \right)' M_{\theta\theta} \left(\frac{\partial \pi_i(\theta_0)}{\partial \theta} \right) \\ + \sum_{i=-\infty}^{\infty} (\pi_i(\theta_0))^2 d'_{t-i} M_{\beta\beta} d_{t-i}$$

where $M_{\theta\theta}$ and $M_{\beta\beta}$ are estimates of the covariance matrices of θ and β derived from the information matrix.¹

In the final two rows of Table 2 we present these estimates for both of our series. Row 9 shows the results for the estimated two-sided filters, so that this result corresponds to historical seasonal adjustment. The final row shows the results using the estimated optimal one-sided filter. These results correspond to the adjustment of current data. Comparing rows 2 and 9, and rows 8 and 10, suggests that parameter uncertainty increased m.s.e. only slightly. This finding is true even though our point estimates of the parameters were not very precise, as the large standard errors in Table 1 indicated.

¹ Formally, of course, $M_{\theta\theta}$ and $M_{\beta\beta}$ should be indexed by T and the range of the summations depend on sample size and t .

5. Conclusions

Given our empirical findings a more full-fledged study of the effect of the use of X11 on the seasonally adjusted unemployment rate seems called for. BLS actually uses X11 on each of the four component series separately and then adds to find the overall unemployment rate. Our evidence on the teenage unemployment rate indicates that X11 could be improved on significantly by a filter that takes account of measurement error. The further question of whether it is a good idea to filter each of the four series by X11 separately should also be investigated. Since the overall unemployment rate is arguably the most watched government series, such an investigation seems worthwhile. If unemployment is not the most watched series, then the inflation rate must be the recipient of the most watched-over award. Our techniques seem applicable to this series also since measurement error may be an important factor. Further investigation seems called for.

A related question which arises is the optimal sample design of the BLS Survey, i.e., the CPS. The National Commission on Employment and Unemployment Statistics Report (1980), vol II considered the question of measurement error, but it did not do so in the context of seasonal adjustment. If an accurate estimate of the non-seasonal component of unemployment is an important goal for the data collection procedure, which it almost surely is, then the sample design could be substantially increased in efficiency by taking account of the simultaneous presence of both measurement error and seasonal components. Our estimated model of Section 3 would be useful in choosing the optimal design for

samples used to measure unemployment statistics.

A last potential unadjusted application arises when revisions take place between the first announced number and the final estimate of an economic time series, e.g., the money supply series. If the measurement error model were applicable to these series in the sense that the measurement error is independent of the final number, then our procedures could be used to predict optimally the final number given the initial estimate. It seems quite likely that the measurement error has a seasonal component which would contaminate the estimates of the seasonal component and non-seasonal components using current seasonal adjustment procedures. Procedures which model the measurement error in an explicit manner should lead to an improvement over current methods.

CHART 1

Components of the Civilian Unemployment Rate

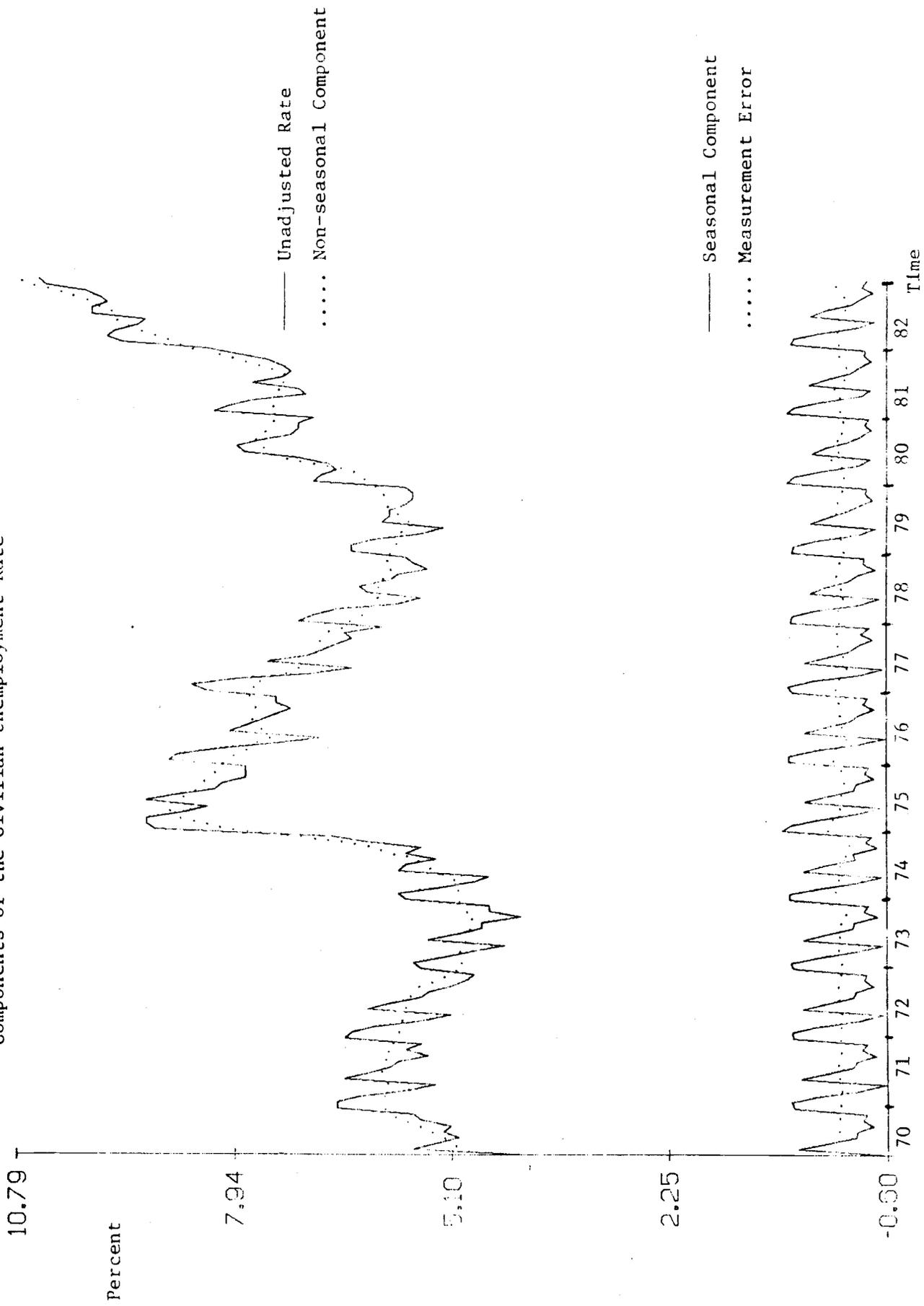


CHART 2

Components of the Teenage Unemployment Rate

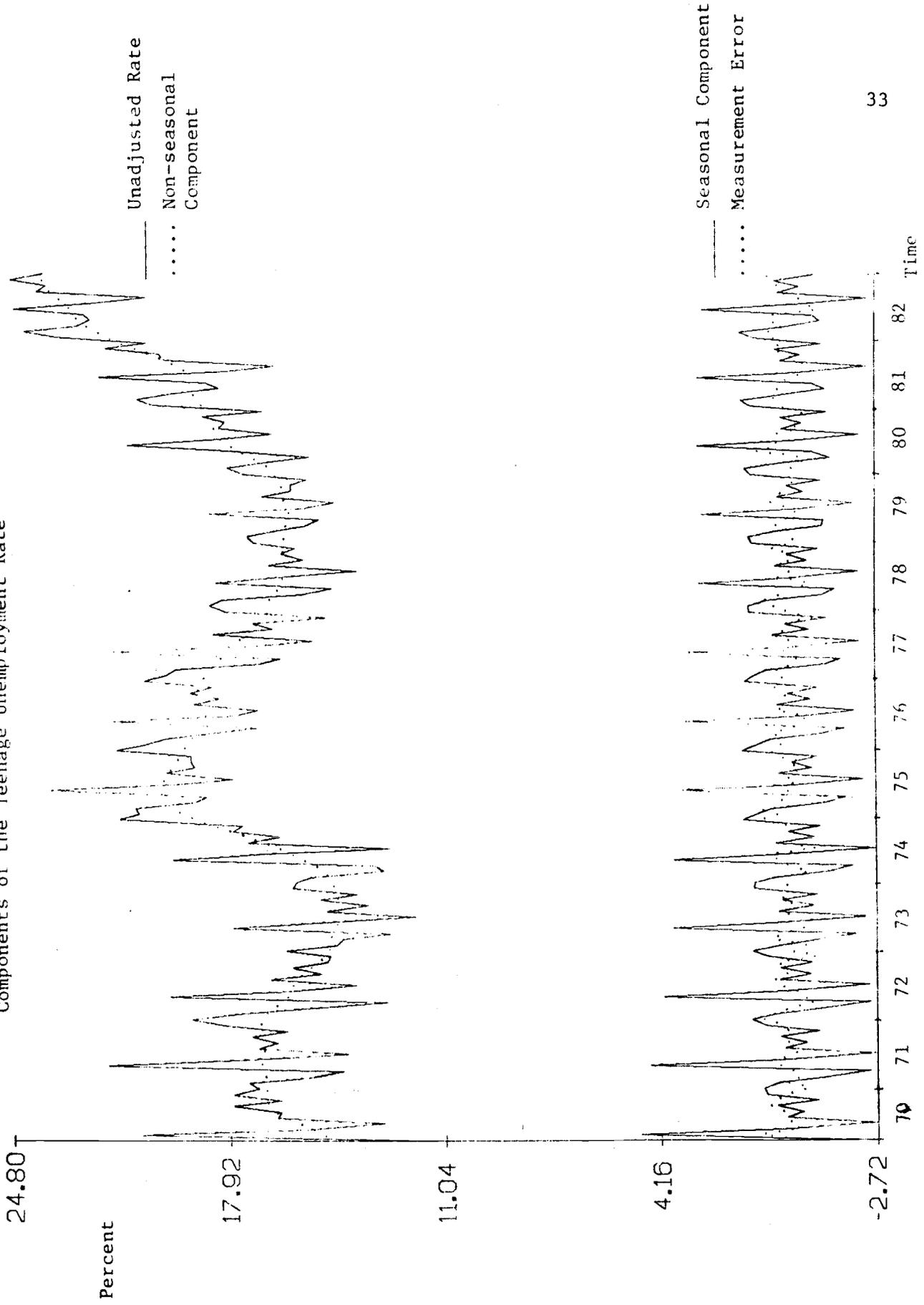


CHART 3

Moving Average Representations of the Civilian Unemployment Rate

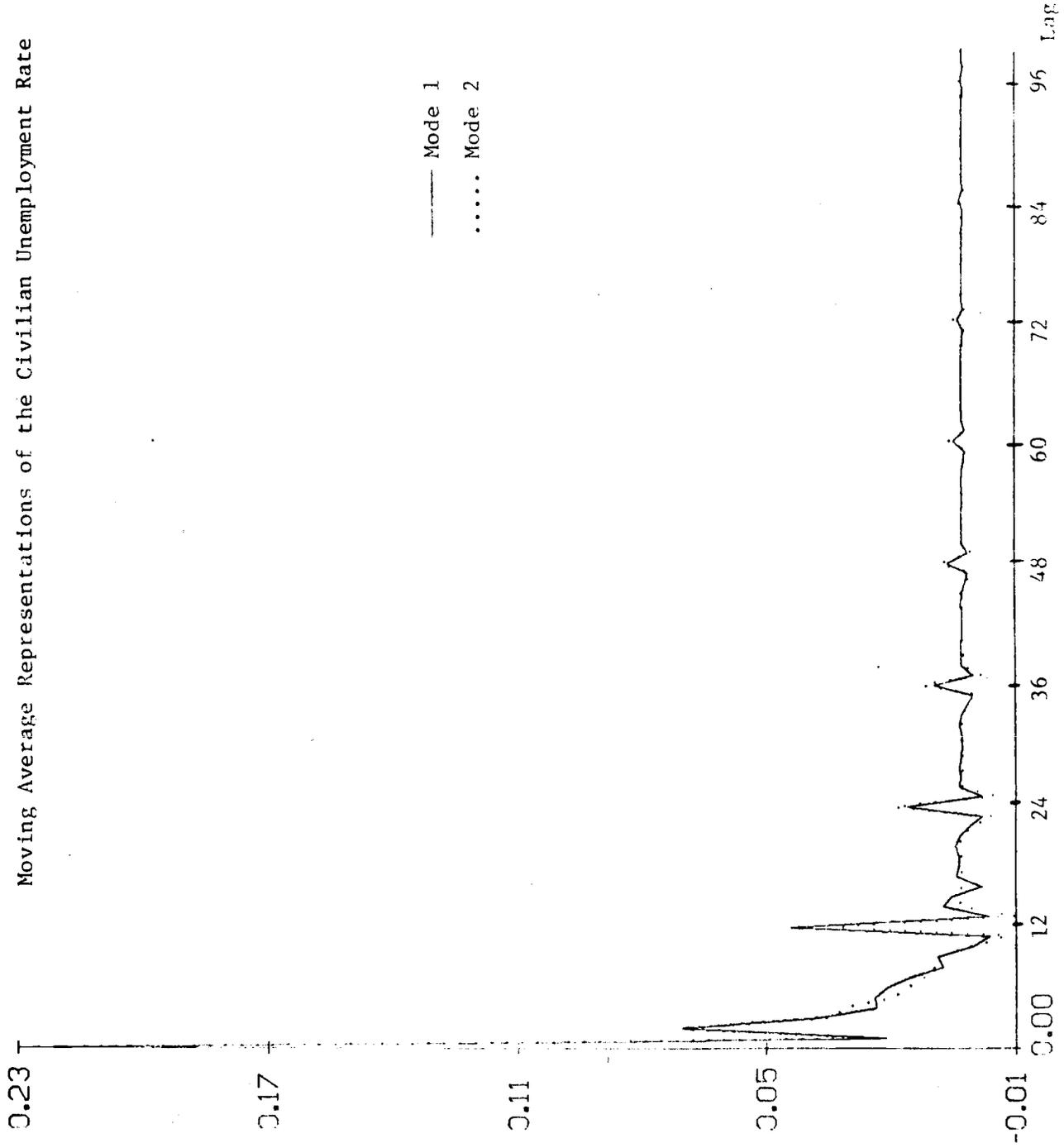


CHART 4

Moving Average Representations of the Teenage Unemployment Rate

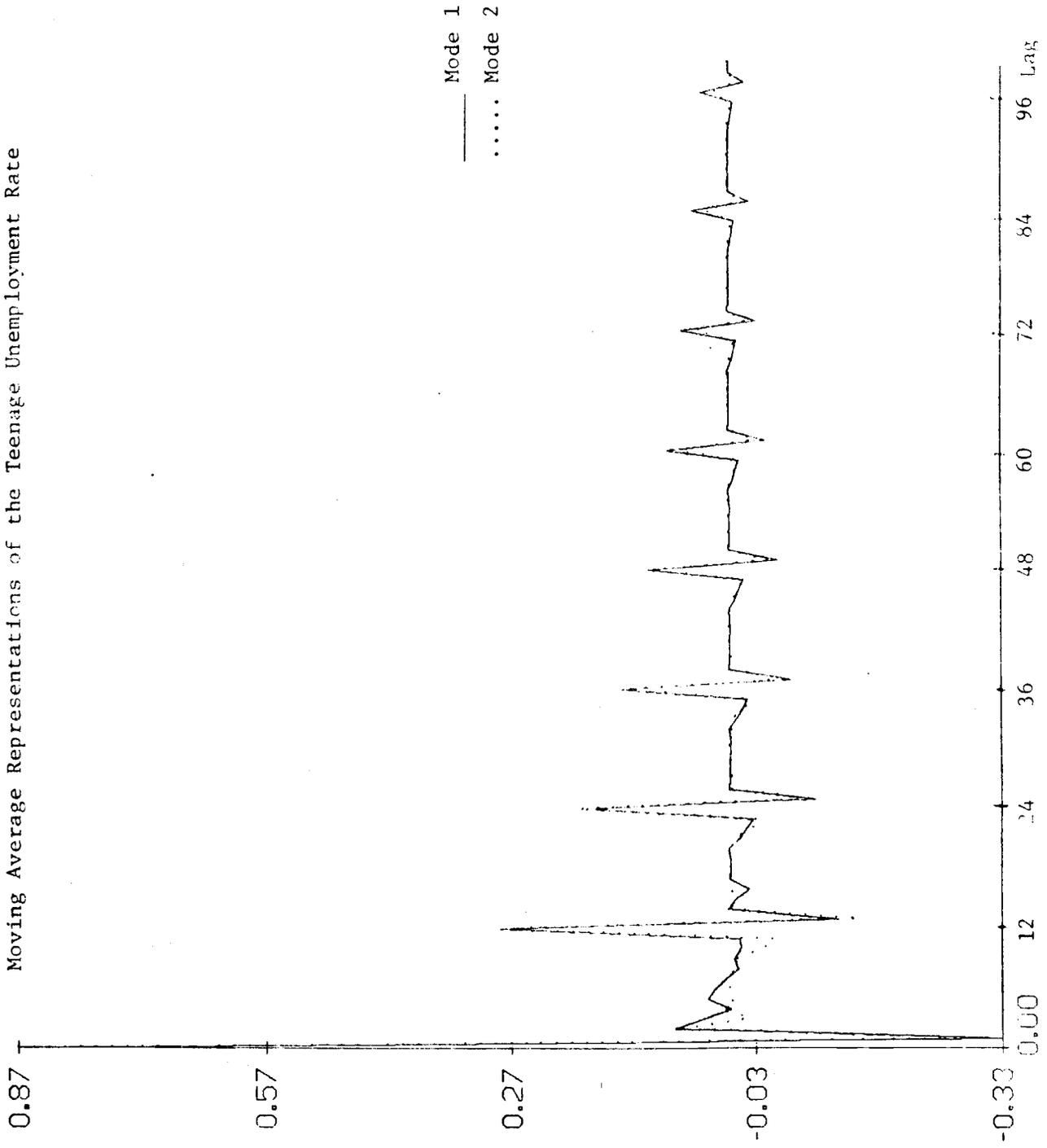


CHART 5

Optimal and Model-2 Filters for the Civilian Unemployment Rate

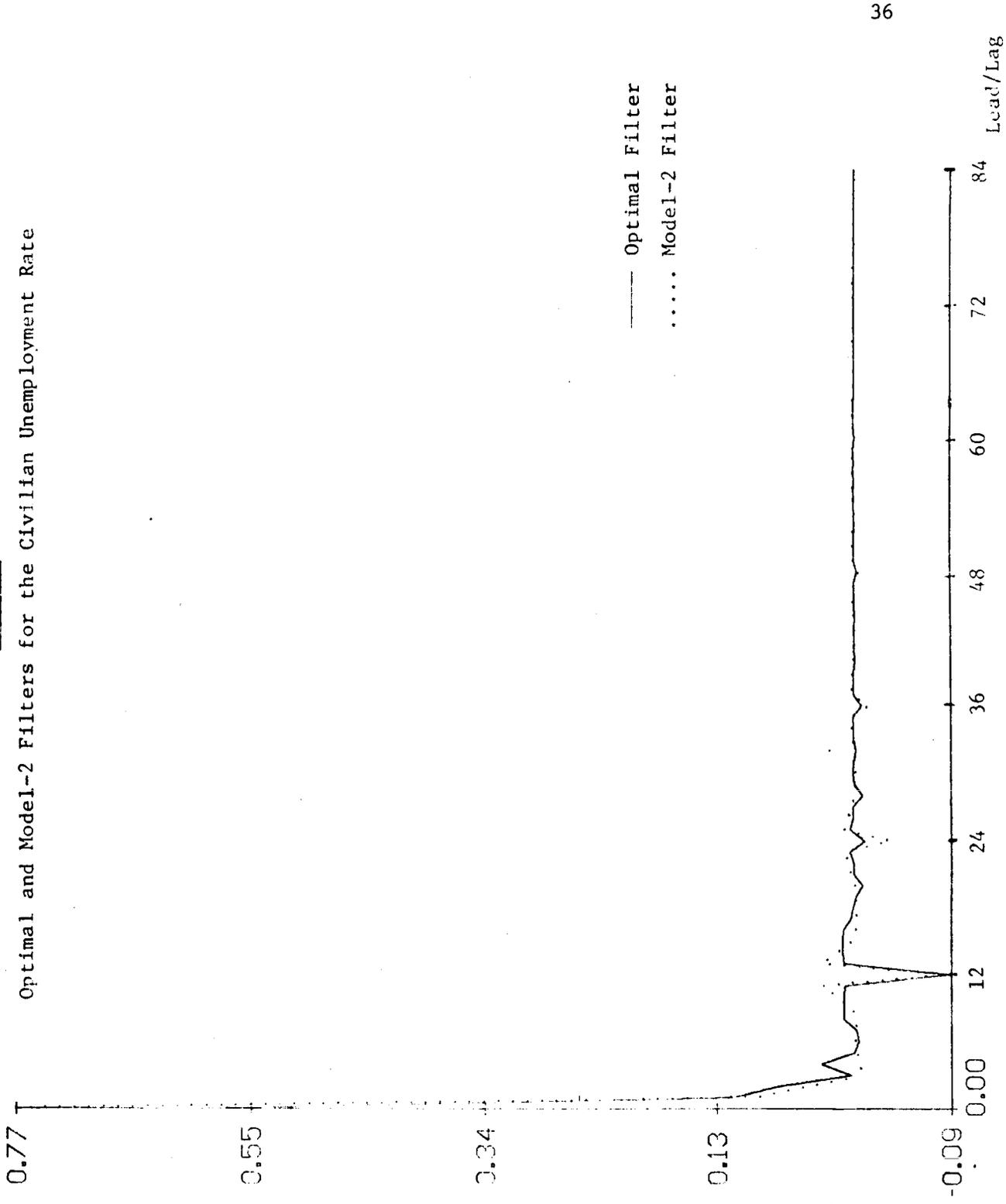


CHART 6

Optimal and Model-2 Filters for the Teenage Unemployment Rate

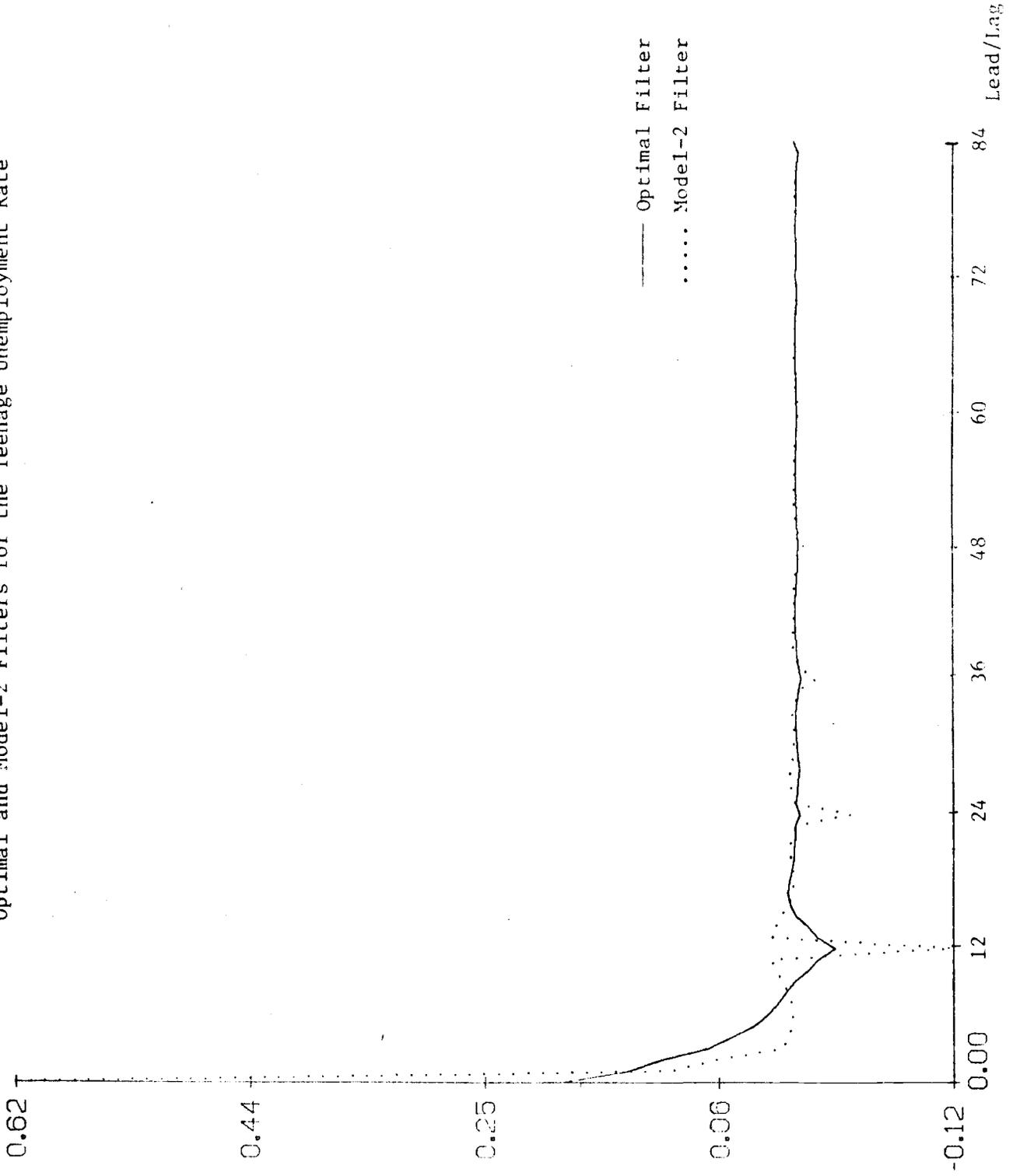


CHART 7

Optimal and X-11 Filters for the Civilian Unemployment Rate

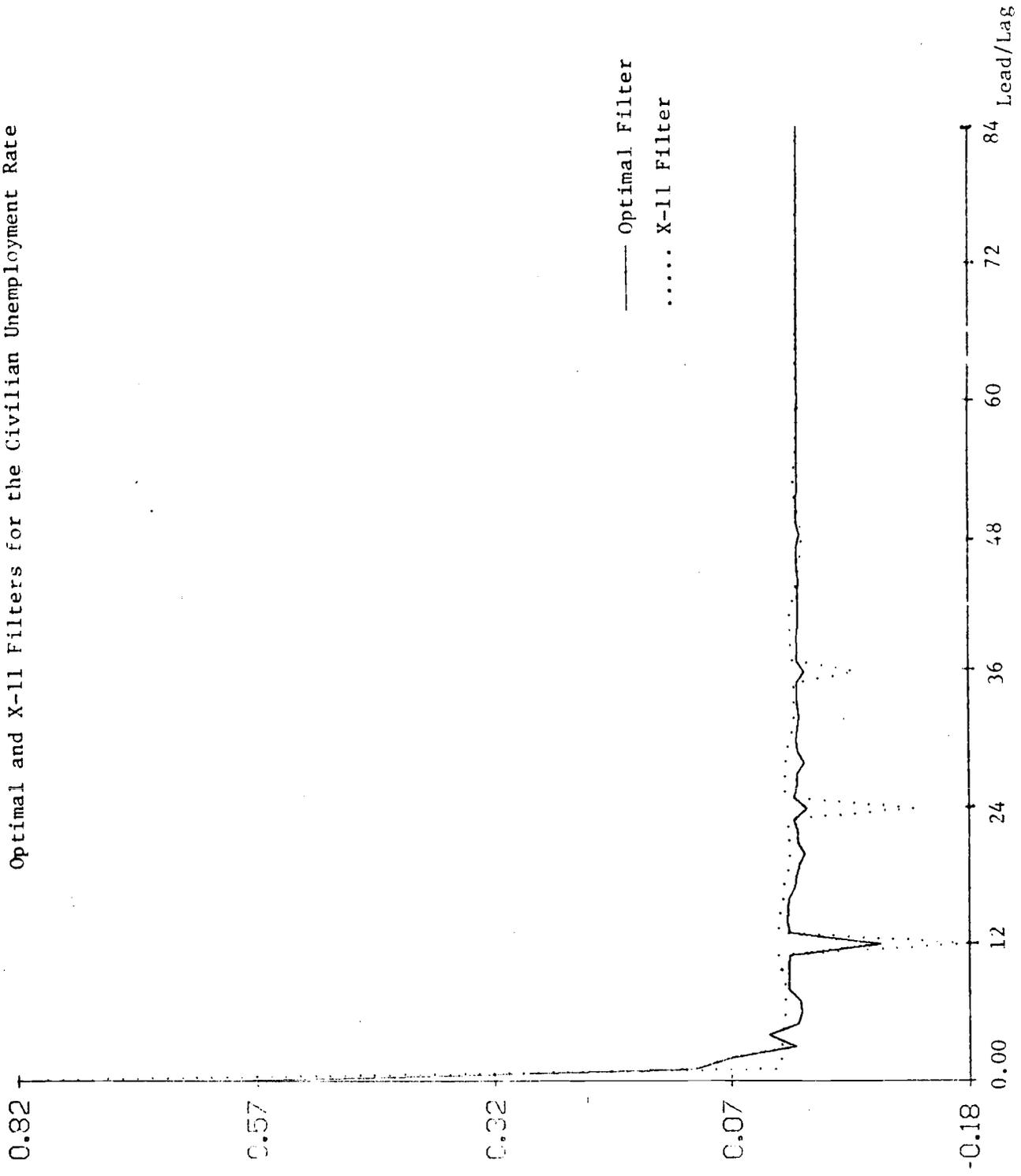
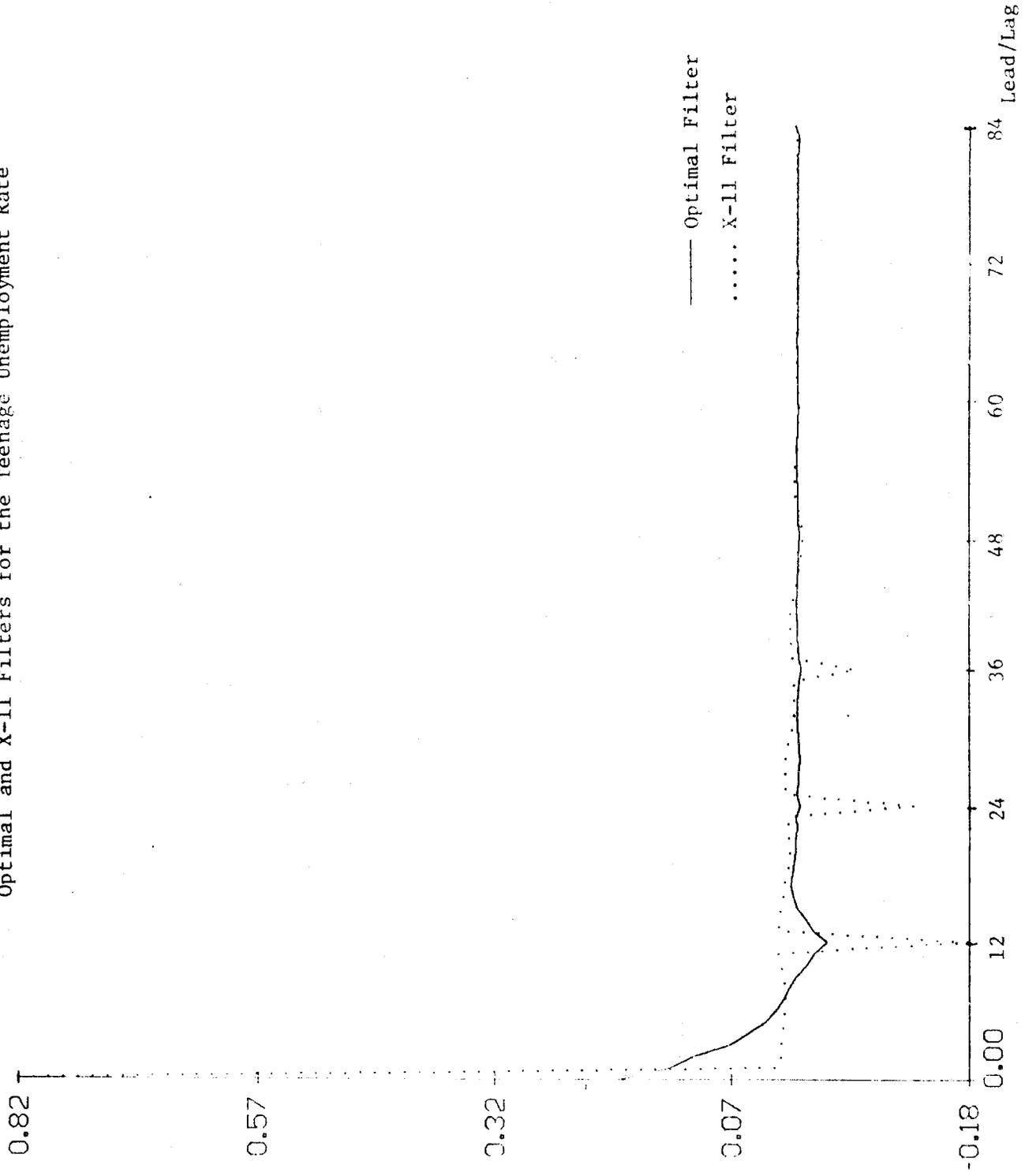


CHART 8

Optimal and X-11 Filters for the Teenage Unemployment Rate



REFERENCES

- Burman, J. P. (1980), "Seasonal Adjustment by Signal Extraction", JRSS, Series A, 143, pp. 321-337.
- Cleveland, W. P. and G. C. Tiao (1976), "Decomposition of Seasonal Time Series: A Model for the Census X-11 Program", JASA, 71, pp. 581-587.
- Engle, R. F. (1976), "Estimating Structural Models of Seasonality", in Seasonal Analysis of Economic Time Series, edited by A. Zellner, U.S. Dept. of Commerce, Bureau of the Census.
- Grether, D. M. and M. Nerlove (1970), "Some Properties of 'Optimal' Seasonal Adjustment", Econometrica, 38, pp. 682-703.
- Harvey, A. C. (1981), Time Series Models, Halstead Press, New York.
- Hillmer, S. C. and G. C. Tiao (1982), "An ARIMA-Model-Based Approach to Seasonal Adjustment", JASA, 77, pp. 63-70.
- Nerlove, M., D. M. Grether, and J. L. Carvalho (1979), Analysis of Economic Time Series, A Synthesis, Academic Press, New York.
- Pagan, A. (1973), "A Note on the Extraction of Components from Time Series", Econometrica, 43, pp. 165-168.
- Pierce, D. A. (1976), "Seasonal Adjustment When Both Deterministic of Stochastic Seasonality is Present," in Seasonal Analysis of Economic Time Series, edited by A. Zellner, U.S. Dept. of Commerce, Bureau of the Census.
- (1979), "Signal Extraction Error in Nonstationary Time Series", Annals of Statistics, 7, pp. 1303-1326.
- Summers, L. H. (1981), "Measuring Unemployment", Brookings Papers on Economic Activity, pp. 609-620.
- Wallis, K. F. (1974), "Seasonal Adjustment and Relations Between Variables", JASA, 69, pp. 18-31.
- Watson, M. W. and R. F. Engle (1983), "Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC, and Varying Coefficient Regression Models", forthcoming Journal of Econometrics.