SEASONAL ADJUSTMENT OF PRELIMINARY DATA

Jerry A. Hausman, MIT and Mark W. Watson, Northwestern University

KEY WORDS: Kalman Filter, Signal Extraction, Measurement Error

1. Introduction

The classic representation for economic time series decomposes an observed series x_t into a seasonal and a nonseasonal component. The seasonal component captures the regular variation in the series over the course of the year and the nonseasonal component captures the residual variation in the series associated with the trend or the business cycle. For an additive model, the decomposition can be written as:

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

where n_t and s_t denote the nonseasonal and seasonal components, respectively. In this model, seasonal adjustment is a procedure for purging x_t of its seasonal component, or equivalently of forming an estimate of n_t, which from (1) can be written as $\hat{n}_t = x_t - \hat{s}_t$.

Since the work of Grether and Nerlove (1970), it has become standard practice to specify independent ARIMA processes for the components n_t and s_t so that (1) becomes an unobserved components ARIMA (UCARIMA) model. Univariate seasonal adjustment procedures—procedures that use information on $\{x_t\}$ only—can then be thought of as signal extraction procedures for estimating the signal, n_t, from noisy measurements, x_t. Grether and Nerlove (1970) and Engle (1978) used this observation to construct optimal seasonal adjustment procedures using signal extraction methods developed by Weiner (1950), Whittle (1963) and Kalman (1960). Indeed, Cleveland and Tiao (1976) show that it is possible to rationalize official seasonal adjustment procedures based on Census X-11 as an optimal signal extraction filter. They solve the inverse optimal filtering problem and find a UCARIMA of the form (1) in which Census X-11 is the optimal signal extraction filter. Since many economic time series are measured with error, a more useful decomposition is of the form

$$(2) \quad x_t = n_t + s_t + e_t$$

where e_t represents the measurement error. In the classic decomposition of time series, e_t is called the irregular component and is modeled as a white noise process. However, white noise measurement error is probably the exception and not the rule for most economic time series. For example, Hausman and Watson (1985) show that the sample design used in the Current Population Survey introduces measurement error that follows and ARMA(1,15) process into the U.S. unemployment data, and Wilcox (1989), Bell and Hillmer (1990) and Bell and Wilcox (1990) find that similar complicated ARMA models describe the measurement error in U.S. aggregate consumption data.

In this paper we extend previous work on seasonal adjustment in two directions. First, we consider the seasonal adjustment of "preliminary," "revised" and "final" data. Preliminary data and revised data are modeled using an equation like (2) with e_t representing the error in these data, while the final data are assumed to be measured without error, so that equation (1) is appropriate. For the data series that we study, New Housing Authorizations (Building Permits), preliminary estimates are revised one month after their initial publication; in turn, these revised estimates are modified after an annual census yields the final value. The second novelty in this paper is a new model for the seasonal component. Following standard practice we allow the seasonal component to follow an ARIMA process, but we allow the innovation in the process to be seasonally heteroskedastic. Thus, as an example, the innovation in the seasonal component may have a very low variance in February, but a high variance in August. This heteroskedasticity introduces a periodic nonstationarity in the univariate representation for x_t that is similar in some respects to the periodic models discussed by Tiao and Grupe (1980), Osborn and Smith (1989), and Hansen and Sargent (1990). This periodic nonstationarity means that standard time invariant seasonal adjustment filters are not

Reprinted from the 1990 Proceedings of the Section on Survey Research Methods of the American Statistical Association

optimal. Optimality requires that a different seasonal adjustment filter be used for each month of the year.

The plan of the paper is as follows. In the next section we discuss the building permit data and in Section 3 we develop a model of the revision error process. In Section 4 we specify and estimate two unobserved component models for the final data. The first model is the traditional homoskedastic UCARIMA model; the second allows the heteroskedastic innovation in the seasonal component discussed above. In the fifth section we compare optimal seasonal adjustment procedures for the two models and compare these procedures to the official Census X-11 procedure. Standard errors for the estimates of $\hat{n}_t$ are presented for both models and for X-11. Some concluding remarks are offered in Section 6.

2. The Data

The data series used in this paper is "Private Housing Units Authorized by Building Permits," published monthly by the U.S. Bureau of the Census.[1] Usually called "Building Permits," the series is an important leading indicator of future macroeconomic activity: it is one of the twelve components used by the Department of Commerce's Index of Leading Indicators and is one the seven components used to construct the National Bureau of Economic Research's Experimental Leading and Recession Indexes (see Stock and Watson (1988)). Economic forecasters rely heavily on the preliminary and one month revisions of the series. Final values of the series, which are available with a considerable lag, have little impact on economic forecasts.

Since 1985, the series has been constructed from the results of a survey of 8,300 of the 17,000 permit-issuing places currently in the permit-issuing universe. The same sample is used every month; the places not included in the monthly sample are surveyed once a year. Although the sample includes fewer than one half of the permit-issuing places, it covers the vast majority of permits issued because of the selection procedure used. In 1989, the sample covered over 92% of permit activity. All places in metropolitan areas and all places that authorized more than a specified number of units in 1978, 1981, and 1982 are surveyed monthly. This lower bound is sixty per year for all but ten sparsely populated states, where it is forty per year. The remaining places are stratified by state: one-tenth are included in the sample.

The preliminary permit data are compiled and published before all the surveys have been returned. During 1986 and 1987, data from an average of 66% of the permit-issuing places were included in the preliminary figure. One month after the preliminary data are published, they are revised to incorporate additional survey responses; an average of 76% of the places in the sample returned their surveys in time to be included in the revised figure. As the surveys come in, the actual data replace the imputed data and revisions are made. In addition, a final revision based on the annual reports from all 17,000 places is made during the middle of the following year. The final revision incorporates late survey responses, corrections, and the results of benchmarking from the annual surveys.

Data from nonresponding areas are imputed. The imputation is carried out in three steps. First, permit-issuing places are sorted by region (Northeast, North Central, South and West) and location (inside or outside an SMSA), yielding eight cells. Second, within each cell, a factor is formed from the ratio of the sum of reported data for the cell for the current month to the sum of the data for the reported places in the cell for the previous year. Third, the previous year's figure for the nonreported place is multiplied by the appropriate factor.

The final estimates incorporate information from the results of the annual census of all 17,000 places in the "universe." This annual census collects data on the annual total of permits issued. This additional information is used to rescale the monthly estimates so that the monthly figures sum to the annual figures. The rescaling is carried out by a link-relative-benchmark method which chooses the revisions in the monthly data to minimize the revisions in the monthly growth rates, subject to the constraint that, over the calendar year, the monthly figures sum to annual census values. Summary statistics for the data are presented in Table 1. The series are volatile and trendless. The standard deviation of monthly growth rates is over 16% for all series, and the twelve month growth rates have standard deviations exceeding 25%.

TABLE 1: Descriptive Statistics

Variable	Mean	Standard Deviation
Y_t	121.44	31.42
Y_t^p	119.83	30.88
Y_t^r	120.24	31.35
Δy_t	-.002	0.163
Δy_t^p	-.002	0.165
Δy_t^r	-.002	0.165
$\Delta^{12} y_t$	-.018	0.252
$\Delta^{12} y_t^p$	-.019	0.252
$\Delta^{12} y_t^r$	-.018	0.253

Notes: Y_t, Y_t^p and Y_t^r are the levels of the final, preliminary and revised data, respectively. Lower case letters denote logarithms, Δ is the first difference operator (1-L) and Δ^{12} is the seasonal difference operator ($1-L^{12}$). The sample period is 1978:1-1989:12.

3. A Model of the Revision Errors

The description of the data suggests that the major source of error is the survey nonresponses in the preliminary and revised data. Over any calendar year, it seems reasonable to model the monthly totals of final data as free from error. As an approximation, we will assume that the monthly values of the final data are measured without error. This assumption greatly simplifies the analysis, since it allows us to calculate historical values of the error in the preliminary and revised data. While this approximation undoubtedly abstracts from some error in the month-to-month changes in the final data, it does allow us to focus on the major source of error in the preliminary and revised data.

The imputation procedure used to construct the preliminary and revised data suggests that a multiplicative model of the error is appropriate. Thus, letting y_t denote the logarithm of the true data at time t, and letting y_t^p and y_t^r denote the logarithms of the preliminary and revised data, we write:

$$(3) \quad y_t^p = y_t + u_t^p$$

$$(4) \quad y_t^r = y_t + u_t^r$$

where u_t^p and u_t^r denote the error in the preliminary and the revised data. Since we assume that the final published data are measured without error, u_t^p and u_t^r can be constructed by subtracting the logarithm of the final data from y_t^p and y_t^r, respectively.

The first and second moment properties of the joint (y_t, u_t^p, u_t^r) process determine the properties of the seasonally adjusted series. In the remainder of this section we analyze the properties of $\{u_t^p, u_t^r\}$ conditional on $\{y_t\}$ and in the next section we analyze the properties of the marginal $\{y_t\}$ process.

Table 2 presents some summary statistics for u_t^p and u_t^r and the correlation between these errors and y_t. Because of significant change in the data collection process beginning in 1978, we limit the sample period to January 1978 through December 1989. The first column of the table shows that u_t^p and u_t^r have sample means of -1.3% and -1.1%, respectively. Serial correlation robust standard errors for the sample means (calculated using an estimated AR(12) model for the errors) yield t-statistics that exceed 3, suggesting a statistically significant bias in the preliminary and final estimates. Moreover, from column 2 of the table, biases are of same order of magnitude as the standard deviations of the errors. The third column presents the F-statistic testing for deterministic seasonality in the errors. This was calculated as the F-statistic on eleven seasonal dummies from a regression of the errors on a constant, the seasonal dummies and 12 lags of the dependent variable. There is no evidence of deterministic seasonality in the errors.

Columns 4-6 of Table 2 investigate the correlation between the errors $\{y_t\}$. Column 4 presents the OLS regression coefficient of u_t^p and u_t^r onto y_t. The coefficient is very close to zero. Interpreting the statistical significance of the OLS coefficient is difficult because, as we show in the next section, y_t is reasonably modeled as a seasonally integrated process. Thus unit root problems complicate the asymptotic distribution of the OLS regression coefficient. In column 5 of the table, we present the dynamic GLS estimate of the coefficient, obtained as the regression of u_t^p and u_t^r on y_t and six leads and lags of $(1-L^{12})y_t$. (The regression is estimated with a correction for an AR(1) error term.) Following the analysis in Stock and Watson (1989), the estimated coefficient on y_t (denoted γ_y in the table), divided by its standard error will have the usual asymptotic normal distribution.

TABLE 2: Preliminary and Revised Measurement Error

Variable	Statistic					
	$\overline{X}$	σ	F_s	β_y	γ_y	F_y
u_t^p	-.013	.017	0.694	-.007	-.010	1.370
	(.004)		(.742)		(.006)	(.178)
u_t^r	-.011	.011	1.019	.009	.010	1.235
	(.004)		(.434)		(.004)	(.259)

Notes: The values in parentheses under $\overline{X}$ and γ_y are standard errors; their construction is described in the text. The values in parentheses under F_s and F_y are the p-values for the F-statistics. The sample period was 1978:1-1989:12.

Inference can be carried out in the dynamic GLS regression without complicated unit root asymptotic distributions. The asymptotic t-statistics on the y_t coefficients are 1.6 for u_t^p and 2.2 for u_t^r. The results suggest a small, but statistically significant correlation between u_t^r and the level of y_t. F-statistics testing that coefficients on y_t and all leads and lags of $(1-L^{12})y_t$ are equal to zero are shown in column 6. Here there is no evidence of significant correlation between the errors and $\{y_t\}$. Taken together, the results in the table suggest that a model specifying u_t^p and u_t^r as uncorrelated with all leads and lags of y_t is broadly consistent with the data. We use this specification in our subsequent analysis, but note that an interesting extension of our analysis would incorporate a nonzero correlation from the regression of the errors onto the level of y_t.

The goal of this section is a complete specification of the process of $\{u_t^p,u_t^r\}$ given $\{y_t\}$. To complete this section we need a model characterizing the joint $\{u_t^p,u_t^r\}$ process. A simple model adequately captured the serial correlation in u_t^p and u_t^r.

$$(5)\quad u_t^p = \underset{(.004)}{-.013} + a_t^p + \underset{(.123)}{.766}\, a_t^r$$

$$(6)\quad a_t^p = \underset{(.083)}{.243}\, a_{t-1}^p + \epsilon_t^p, \qquad \sigma_{\epsilon p} = .015$$

$$(7)\quad u_t^r = \underset{(.004)}{-.011} + a_t^r$$

$$(8)\quad a_t^r = \underset{(.080)}{.290}\, a_{t-1}^r + \epsilon_t^r, \qquad \sigma_{\epsilon r} = .010$$

$F_1 = .176$, $F_2 = .106$, $F_3 = .353$,
$F_4 = .776$, $F_5 = .218$, $F_6 = .315$.

The numbers under the estimated coefficients are standard errors, and the F-statistics listed after the model are p-values for diagnostic tests that will be described below. Equations (5) and (6) represent the regression of u_t^p onto u_t^r; a_t^p is the AR(1) error in the regression. Equations (7) and (8) represent the AR(1) process for u_t^r. The specification assumes that ϵ_t^p and ϵ_t^r are uncorrelated at all leads and lags. The statistics F_1 through F_6 test various restrictions implicit in the specification. F_1 tests the null hypothesis that a_t^r enters equation (5) against the alternative that $a_t^r, a_{t-1}^r, \ldots, a_{t-6}^r$ belong in the equation. The resulting F-statistic had a value of 1.52 with the p-value of .176 shown above. F_2 tests the AR(1) null in equation (6) versus an AR(2) alternative. F_3 tests the AR(1) null in equation (6) versus an AR(12) alternative. F_4 tests the null that lagged values of u_t^p do not belong in (7) against the alternative that lags 1-6 do belong in the equation. F_5 tests the AR(1) null in equation (8) versus and AR(2) alternative, while F_6 tests the AR(1) null versus an AR(12) alternative. These diagnostics suggest that (5)-(8) provide a reasonable characterization of the revision error process.

This section has developed a model of the process characterizing the $\{u_t^p,u_t^r\}$ given $\{y_t\}$. In summary, the model assumes that $\{u_t^p,u_t^r\}$ are uncorrelated with $\{y_t\}$. The first and second moment properties of $\{u_t^p,u_t^r\}$ are given implicitly by (5)-(8). To complete the specification of the joint process for $\{y_t^p,y_t^r,y_t\}$ we require a model for the $\{y_t\}$ process. This is the subject of the next section.

4. Models of the Final Data

We assume that the y_t data are measured without error so that the decomposition given in equation (1) forms the basis of the models that we estimate. We estimate two related models. The first is:

$$(9)\qquad y_t = n_t + s_t$$

$$(10) \qquad n_t = \tau_t + \epsilon_t^i$$

$$(11) \qquad (1-L)\tau_t = \epsilon_t^\tau$$

$$(12) \qquad \left(1+L+L^2+ \ldots +L^{11}\right)s_t = \epsilon_t^s$$

where ϵ_t^i, ϵ_t^τ, and ϵ_t^s are uncorrelated gaussian white noise processes with variance σ_i^2, σ_c^2, and σ_s^2, respectively. Equations (10) and (11) represent the nonseasonal process as an ARIMA(0,1,1) parameterized as the sum of a random walk plus independent white noise; from (12) the seasonal component follows a seasonal autoregressive process. The initial level of the process is captured by the initial value of τ_0, while the initial seasonal pattern is captured by $s_0, s_{-1}, \ldots, s_{-11}$. Our specification (9)-(12) is essentially the model used by Harvey and Todd (1983). They allowed a random walk intercept term to enter (11), but for their applications, point estimates suggested that the intercept was a constant representing the average drift in the data. Since building permits do not contain a drift, this component is absent from our model. A detailed discussion of the specification and its relation to other models can be found in the Harvey and Todd paper and in Harvey (1989).[2]

The model (9)-(12) was estimated by maximum likelihood using a diffuse prior on the initial values of τ_t and s_t.[3] The results are summarized in the first row of Table 3. The point estimates imply that the standard deviation of the one-step-ahead forecast error for y_t is 9.5%. Most of the uncertainty arises from the nonseasonal component; the standard deviation of the seasonal innovation is 0.48%. The model fits the data reasonably well; point estimates are broadly in accord with an unconstrained ARIMA model for y_t. The normalized innovations from the estimated model suggest some modest autocorrelation; estimated autocorrelation coefficients exceeded .2 in absolute value at lags 10 and 13.

The second model allows the variance of the innovation in s_t, denoted ϵ_t^s in equation (12), to have a month-specific variance. Thus for example, February's seasonal innovation may have small variance while August's seasonal innovation may have a large variance. The unrestricted seasonally heteroskedastic model yielded a maximized log likelihood value of 543.6 compared to 539.0 for the homoskedastic model. While the difference in the likelihood values is not statistically significant, the point estimates from the heteroskedastic model suggested that the months could be separated into three groups. The point estimates of the variance of ϵ_t^s for August and December were large; the variances were smaller, but still markedly different from zero, for March, May, September and November. For the other months, the variances were essentially zero. These point estimates suggested a specification in which August and December had a common variance, March, May, September and November another, and the variance of ϵ_t^s was constrained to equal zero for the other months. The point estimates for this model are reported in the second row of Table 3.

This restricted seasonally heteroskedastic model fits the data nearly as well as the general seasonally heteroskedastic model. The log likelihood falls from 543.6 in the unrestricted model to 543.1 in the restricted model, even though 10 fewer parameters have been estimated.[4] The restricted heteroskedastic model appears to fit the data much better than the homoskedastic model. The improvement in the log likelihood is 4.1, while only one additional parameter has been estimated.

In the next section we will use both the homoskedastic model and the restricted seasonally heteroskedastic model together with the models for the preliminary and revised data errors to construct and evaluate seasonal adjustment procedures.

5. Seasonal Adjustment

In this section we answer three questions. First, what is the standard error of the estimated value of n_t using Census X-11? Second, how much improvement can be expected over X-11 from the use of optimal model based procedures? Finally, how much of the mean square error in the seasonally adjusted estimates can be attributed to the preliminary and revised measurement error, and how much is inherent in the underlying data generation process?

To answer the first of these questions we use the linear approximation to Census X-11 presented in Wallis (1974). This allows us to calculate the mean square error of the seasonally adjusted data using standard linear time series methods. Wallis derives a symmetric 82-term

TABLE 3: Unobserved Components Model

	σ_τ	σ_i	σ_s	σ_s^1	σ_s^2	Log Likelihood
Model 1	.0690 (.0044)	.0431 (.0053)	.0048 (.0016)	--	--	539.00
Model 2	.0693 (.0044)	.0396 (.0054)	--	.0041 (.0029)	.0196 (.0070)	543.13

Notes: Model 1 is the homoskedastic UCARIMA model and Model 2 is the heteroskedastic UCARIMA model. σ_s^1 is the standard deviation of the seasonal innovation for August and December. σ_s^2 is the standard deviation of the seasonal innovation for March, May, September and November. The sample period used was 1960:8-1989:12.

moving average filter to approximate X-11. We write his filter as:

$$(13) \qquad X11(L) = \sum_{i=-82}^{82} a_{|i|} L^i$$

Wallis' two-sided filter approximates the historical X-11 filter, the filter used to adjust historical data.

For concurrent seasonal adjustment—seasonally adjusting the most currently available data—X-11 cannot be used since it requires future and past data. Instead, an alternative filter, X-11 ARIMA, is used. This procedure replaces the future values of the series, necessary for X11(L), with forecasts constructed from an ARIMA model. We write the X-11 ARIMA filter as:

$$(14) \quad X11A(L) = \sum_{i=0}^{\infty} b_i L_i$$

where the filter weights b_i can be calculated as a function of the historical X-11 weights, a_i in equation (13), and the parameters of the ARIMA process used to form the forecasts of the series. In this paper we formed X11A(L) using the ARIMA process for y_t implied by the homoskedastic model estimated in the last section. This ARIMA model will closely approximate any well specified model for y_t constructed from the historical time series.

Since both X11(L) and X11A(L) are time invariant linear filters, the mean square error of the seasonally adjusted data can be calculated using standard calculations. Letting $\hat{n}_t$ denote the historical value of the X-11 seasonally adjusted data,

$$(15) \quad \hat{n}_t = X11(L)y_t = X11(L)n_t + X11(L)s_t .$$

Thus,

$$(16) \quad n_t - \hat{n}_t = [1-X11(L)]n_t - X11(L)s_t .$$

Noting that 1-X11(L) contains the factor (1-L) and X11(L) contains the factor $(1+L+ \ldots +L^{11})$, the mean square error of $\hat{n}_t$ can be calculated directly from (16).

The calculation for the currently adjusted value is slightly different, since the filter is not applied to y_t, but to a combination of y_t^p, y_t^r, and y_t. In particular, if we denote the current (X-11 ARIMA) nonseasonal estimate by $\hat{n}_t^a$

$$(17) \quad \begin{aligned} \hat{n}_t^a &= b_0 y_t^p + b_1 y_{t-1}^r + \sum_{i=2}^{\infty} b_i y_{t-i} \\ &= X11A(L)y_t + b_0 u_t^p + b_1 u_{t-1}^r \end{aligned}$$

so that

$$(18) \quad \begin{aligned} n_t - \hat{n}_t^a &= (1-X11A(L))n_t - X11A(L)s_t \\ &\quad - b_0 u_t^p - b_1 u_{t-1}^r . \end{aligned}$$

The argument in Watson (1987, Appendix B) implies that 1-X11A(L) contains the factor (1-L) and X11A(L) contains the factor $(1+L+ \ldots +L^{11})$, so that the root mean square error of $\hat{n}_t^a$ can be calculated directly from (18).

It is possible to construct more efficient estimates of the nonseasonal component using optimal model based seasonal adjustment methods. It is a straightforward exercise to write the joint $\{y_t^p, y_t^r, y_t\}$ process in state-space form. Optimal estimates of the nonseasonal component, n_t, can then be formed using the Kalman filter (for concurrent seasonal adjustment) and Kalman smoother (for historical seasonal adjustment). The root mean square error associated with the filtered and smoothed estimates are calculated as a byproduct.

Table 4 summarizes our results on the root mean square error of the seasonally adjusted data. Panel A shows the results for the homoskedastic model and Panels B and C show the results for the heteroskedastic model. Looking first at Panel A, the seasonally adjusted preliminary data has a root mean square error of 4.9% and the annual growth rate has a RMSE of

6.4%.[5] Using optimal filters these values fall to 2.7% and 1.9%, respectively. The official procedures, X-11 and X-11 ARIMA, are far from optimal. The relative efficiency of X-11 and X-11 ARIMA ranges from a high of 34% to a low of 1%.

Panel B shows results for the heteroskedastic model averaged over the year. The results are similar to the results for the homoskedastic model. Panel C presents month-by-month results for the heteroskedastic model. Variation in the root mean square error of the optimal concurrent estimates range from a high of 3.6% in August and December to a low of 2.2% in February and July. Interestingly, while the X-11 RMSEs are considerably larger than the optimal estimates, they vary much less over the year.

6. Concluding Remarks

Seasonal adjustment techniques arose from a need to isolate trend and cyclical movements in economic time series. For this purpose the minimum mean square error extraction methods discussed here are significant improvements over the existing official seasonal adjustment methods. For other purposes, minimum mean square error adjustment procedures may not be an improvement. As an extreme example, as Sargent (1978) and Miron (1986) argue, seasonal variation in economic time series may contain useful information about economic behavior that should be used when estimating an econometric model. When using seasonally adjusted data to estimate linear relations between variables, the arguments in Wallis (1974) suggest that the use of a generic filter, like X-11, is preferred to model-based filters. Finally, in a similar context, Sims (1974) argues that seasonal misspecification in economic models can be mitigated by removing seasonality with a narrow band pass filter and imposing smoothness constraints on estimated transfer functions.

There are important lessons in this paper for users of seasonally *unadjusted* data as well as seasonally adjusted data. The model of the preliminary and revised measurement error, with the composite model for y_t can be used to form improved estimates of the seasonally *unadjusted* data. Given the model, the Kalman filter can be used to eliminate some preliminary and revised error from the data. As an example, the preliminary data, y_t^p, has a root mean square error of 2.1%. By optimally using the time series properties of y_t, y_t^r, y_t^p, the Kalman filter can produce an estimate with a root mean square error of 1.7%. While admittedly not dramatic for this data series, this improvement could be much larger for other series.

TABLE 4: Root Mean Square Error Seasonally Adjusted Data

A. Homoskedastic Model

	n_t	n_t-n_{t-1}	n_t-n_{t-12}
Historical filters			
Optimal	.018	.023	.007
X-11	.047	.060	.062
Concurrent filters			
Preliminary data			
Optimal	.027	.035	.019
X11-ARIMA	.049	.060	.064
Final data			
Optimal	.022	.031	.007
X11-ARIMA	.046	.058	.062

B. Heteroskedastic Model (Average over 12 months)

	n_t	n_t-n_{t-1}	n_t-n_{t-12}
Historical filters			
Optimal	.018	.024	.008
X-11	.045	.057	.058
Concurrent filters			
Preliminary data			
Optimal	.027	.035	.021
X11-ARIMA	.048	.061	.062
Final data			
Optimal	.022	.032	.008
X11-ARIMA	.046	.059	.060

C. Heteroskesdastic Model RMSE of Estimated n_t Monthly Values

	Historical Filter		Concurrent Filter (Preliminary Data)	
	Optimal	X11	Optimal	X11-ARIMA
January	.023	.044	.033	.047
February	.015	.044	.022	.047
March	.016	.044	.025	.047
April	.016	.044	.024	.047
May	.016	.044	.025	.047
June	.016	.044	.024	.047
July	.015	.046	.022	.052
August	.023	.046	.036	.053
September	.024	.044	.033	.047
October	.016	.044	.025	.047
November	.016	.046	.025	.052
December	.024	.046	.036	.053

Notes: The rmse for the concurrent optimal estimate in the homoskesdastic model using preliminary data varies over the year. This variation occurs because of the timing of the release of final data. In the table average values over the year are reported.

References

Bell, W.R. and Hillmer, S.C (1990), "The Time Series Approach to Estimation for Periodic Surveys," manuscript, Bureau of the Census.

Bell, W.R. and Wilcox, D.W. (1990), "The Effect of Sampling Error on the Time Series Behavior of Consumption Data," manuscript, Bureau of the Census.

Cleveland, W.P. and Tiao, G.C. (1976), "Decomposition of Seasonal Time Series: A Model for the Census X-11 Program," Journal of the American Statistical Association, 71, 581-587.

Engle, R.F. (1978), "Estimating Structural Models of Seasonality," in *Seasonal Analysis of Economic Time Series*, edited by A. Zellner, U.S. Bureau of the Census, Economic Research Report ER-1.

Grether, D.M. and M. Nerlove (1970), "Some Properties of 'Optimal' Seasonal Adjustment," *Econometrica*, 38, 686-703.

Hansen, L.P. and Sargent, T.J. (1990). "Disguised Periodicity as a Source of Seasonality", manuscript, Hoover Institution, Stanford University.

Harvey, A.C. (1989), *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 and Economic Statistics*, 1, 299-306.

Hausman, J.A. and Watson, M.W. (1985), "Seasonal Adjustment with Measurement Error Present," *Journal of the American Statistical Association*, 80, 531-540.

Kalman, R.E. (1960), "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, 83, 95-108.

Miron, J.A. (1986), "Seasonal Fluctuations and the Life Cycle-Permanent Income Model of Consumption," *Journal of Political Economy*, 94, 1258-1279.

Osborn, D.R. and Smith, J.P. (1989), "The Performance of Periodic Autoregressive Models in Forecasting Seasonal U.K. Consumption," *Journal of Business and Economic Statistics*, 7, 117-128.

Tiao, G.C. and Grupe, M.R. (1980), "Hidden Periodic Autoregressive-Moving Average Models in Time Series Data," *Biometrika*, 67, 365-373.

Sargent, T.J. (1978), "Comment on 'Seasonal Adjustment and Multiple Time Series Analysis'," in *Seasonal Analysis of Economic Time Series*, edited by A. Zellner, U.S. Bureau of the Census, Economic Research Report ER-1.

Sims, C.A. (1974), "Seasonality in Regression," *Journal of the American Statistical Association*, 69, 618-626.

Stock, J.H. and Watson, M.W. (1989), "A Simple MLE of Cointegrating Vectors in Higher Order Integrated Systems," NBER Technical Working Paper No. 83.

Wallis. K.F. (1974), "Seasonal Adjustment and Relations Between Variables," *Journal of the American Statistical Association*, 69, 18-31.

Weiner, N. (1950), *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Cambridge and New York: Massachusetts Institute of Technology Press and John Wiley and Sons, Inc.

Whittle, P. (1963), *Prediction and Regulation by Least Squares*, London: The English University Press, Ltd.

Wilcox, David W. (1989), "What Do We Know About Consumption," manuscript, Federal Reserve Board.

Wolak, F.A. (1986), "Testing Nonlinear Inequality Constraints in the Maximum Likelihood Model," Technical Report 25, Stanford University Econometric Workshop, Department of Economics.

Notes

We thank Ruth Judson for excellent research assistance and Linda Hoyle and Jesse Pollock of the Census Bureau for useful conversations concerning the construction of the data. This research was supported by the National Science Foundation through grants SES-89-10601 and SES-86-18769.

1. See Current Construction Reports, Series C20: *Housing Starts*.

2. Our specification abstracts from trading day variation in the data. In our model, regular seasonal trading day variation will be captured by the seasonal component. Other variation, such as the timing of Easter, will presumably be captured by the nonseasonal component. A useful and interesting extension of our results would incorporate this additional component.

3. An excellent discussion of estimation methods for UCARIMA models can be found in Harvey (1989).

4. Because the restricted model constrains some of the variances to zero, one suspects that the asymptotic χ^2 distribution of the likelihood ratio statistic will not obtain. Two complications arise. The first is the problem of testing a parameter on the boundary of the parameter space, a problem considered in detail in Wolak (1986). The second is that, because of the unit roots in the seasonal autoregressive operator, the elimination of some the shock effectively eliminates a seasonal stochastic trend. This suggests that problems analogous to testing for a unit moving average coefficient will affect the distribution of the likelihood ratio statistics.

5. When calculating the RMSE for the monthly or annual growth rate it is important to note that the estimates of n_t, n_{t-1} and n_{t-12} are calculated using different filters. All are constructed using data through time t, so that the estimate of n_{t-1} is formed using one future value (y_t) and the estimate of n_{t-12} is formed used twelve future values ($y_{t-11}, y_{t-10}, \ldots, y_t$).