Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Toward variances for X-11 seasonal adjustments

by William R. Bell and Matthew Kramer

June 1999

Statistics Canada Statistique Canada

Canada

Survey Methodology, June 1999
Vol. 25, No. 1, pp. 13-29
Statistics Canada, Catalogue No. 12-001

Toward variances for X-11 seasonal adjustments

William R. Bell and Matthew Kramer [1]

Abstract

We develop an approach to estimating variances for X-11 seasonal adjustments that recognizes the effects of sampling error and errors from forecast extension. In our approach, seasonal adjustment error in the central values of a sufficiently long series results only from the effect of the X-11 filtering on the sampling errors. Towards either end of the series, we also recognize the contribution to seasonal adjustment error from forecast and backcast errors. We extend the approach to produce variances of errors in X-11 trend estimates, and to recognize error in estimation of regression coefficients used to model, *e.g.*, calendar effects. In empirical results, the contribution of sampling error often dominated the seasonal adjustment variances. Trend estimate variances, however, showed large increases at the ends of series due to the effects of fore/backcast error. Nonstationarities in the sampling errors produced striking patterns in the seasonal adjustment and trend estimate variances.

Key Words: Sampling error; Forecast error; Trading-day; ARIMA model.

1. Introduction

The problem of how to obtain variances for seasonally adjusted data is long-standing (President's Committee to Appraise Employment and Unemployment Statistics 1962). Model-based methods of seasonal adjustment (see Bell and Hillmer 1984, for a discussion) use results from signal extraction theory to produce estimates and associated error variances of the seasonal and nonseasonal components. Most official seasonal adjustments, however, are made using empirical methods, most notably X-11 (Shishkin, Young and Musgrave 1967) or X-11-ARIMA (Dagum 1975). These methods are based on fixed filters, not models, and so it is not obvious how to calculate variances of the seasonal adjustment errors. Various approaches for obtaining variances for X-11 seasonal adjustments have been proposed, as summarized below.

Wolter and Monsour (1981) suggested two approaches. They recognized that many time series that are seasonally adjusted are estimates from repeated sample surveys, and thus are subject to sampling error. Their first approach accounts only for the effect of sampling error on the variance associated with seasonal adjustments. Their second approach tries to also reflect uncertainty due to stochastic time series variation in the seasonal adjustment variances. However, this second approach assumes that, apart from regression terms, the time series is stationary. This type of model is now seldom used for seasonal time series. Also, their second approach contains a conceptual error: it produces the variance of the seasonally adjusted estimate, instead of the desired variance of the error in the seasonally adjusted estimate.

Burridge and Wallis (1985) investigated use of the steady-state Kalman filter for calculation of model-based seasonal adjustment variances, and applied this approach to a model they obtained previously (Burridge and Wallis 1984) for approximating the X-11 filters. They suggested that this approach could be used to, "provide measures of the variability of the X-11 method when it is applied to data for which it is optimal," (page 551), but cautioned against doing this when the X-11 filter would be suboptimal (*i.e.*, very different from the optimal model-based filter). Hausman and Watson (1985) suggested an approach to estimating the mean squared error for X-11 when it is used in suboptimal situations. Bell and Hillmer (1984, section 4.3.4) pointed out a problem with the use of model-based approximations to X-11 for calculating seasonal adjustment variances. The problem is that X-11 filters (or any seasonal adjustment filter, for that matter) are not sufficient to uniquely determine models for the observed series and its components.

Pfeffermann (1994) developed an approach that recognizes the contributions of sampling error and irregular variation (time series variation in the irregular component) to X-11 seasonal adjustment variances. The properties of the combined error (sampling error plus irregular) are estimated using the X-11 estimated irregular. These properties are then used to estimate two types of seasonal adjustment variances. A drawback to this approach is that it relies on an assumption that the X-11 adjustment filter annihilates the seasonal component and reproduces the trend component. (Note Pfeffermann (1994, page 90), discussion surrounding equation (2.7).) Violations of this assumption in practice compromise the approach to an extent which appears difficult to assess. Thus, this assumption seems to us highly questionable and also, in any particular case, uncheckable. A second drawback is that one of the variance types proposed by Pfeffermann assumes that the X-11 seasonally adjusted series, rather than the trend estimate, is taken as an estimate of the trend. Breidt

1. William R. Bell and Matthew Kramer, Statistical Research Division, Room 3000-4, U.S. Bureau of the Census, Washington, D.C. 20233-9100, U.S.A.

(1992) and Pfeffermann, Morry and Wong (1993) further develop Pfeffermann's general approach.

The goal of this paper is the development and application of an approach to obtaining variances for X-11 seasonal adjustments accounting for two sources of error. The first error source is sampling error. The second is error that arises from the need to extend the time series with forecasts and backcasts before applying the symmetric X-11filters. These latter errors lead to seasonal adjustment revisions (Pierce 1980). Note that revisions eventually vanish as sufficient data beyond the time point being adjusted become available. Also note that a seasonally adjusted series will not contain sampling error if the corresponding unadjusted series does not. This is the case for certain economic time series, *e.g.*, export and import statistics for most countries.

Our approach assumes that the X-11 seasonal adjustment target (what we assume application of X-11 is intended to estimate) is what would result from application of the symmetric linear X-11 filter (with no forecast and backcast extension required) if the series contained no sampling error. While this definition of target might be criticized for ignoring time series variation in the underlying seasonal and nonseasonal components, we think this may be appropriate for typical users of X-11 seasonally adjusted data. Such users are most likely to be concerned about uncertainty reflected in differences between initial adjustments and final adjustments, *i.e.*, in revisions. Some of these users will also be aware that the unadjusted series consists of sample-based estimates of the true underlying population quantities, and will realize that the effects of sampling error on adjustments should also be reflected in seasonal adjustment variances.

Our development is based on use of the symmetric linear X-11 filters. We assume that the symmetric filters are applied to the series extended with minimum mean squared error forecasts and backcasts. In practice, the forecasts and backcasts are obtained from a fitted time series model. This is in the spirit of the X-11-ARIMA method of Dagum (1975), but with full forecast and backcast extension, as recommended by Geweke (1978), Pierce (1980), and Bobbitt and Otto (1990). Our results apply directly to the use of additive or log-additive X-11 (with forecast and backcast extension), and the log-additive results are assumed to apply approximately (Young 1968) to multiplicative X-11.

Section 2 of this paper develops our approach, which builds on the first approach of Wolter and Monsour (1981). The differences between the two approaches are discussed in section 2.4. Section 3 then discusses three extensions to the results of section 2. The first is to note that our approach works equally well with seasonal, trend, or irregular estimates, and that more generality is easily accommodated by allowing different filter choices for different months. The second extension produces variances of estimates of month-to-month or year-to-year change. Finally, when seasonal adjustment involves estimation of regression effects (*e.g.*, for trading-day or holiday variation), the results are extended to allow for additional variance due to error in estimating the regression parameters.

Section 4 then presents several examples illustrating the basic approach and the extensions given in section 3. One thing evident from the examples is that for time series with sampling error, our seasonal adjustment variances will often be dominated by the contribution of the sampling error. In the center of the series, our results effectively reduce to the first approach results of Wolter and Monsour. Our results do differ from those of Wolter and Monsour near the end of the series. This is important since the most recent seasonally adjusted values receive the most scrutiny. Also, the contribution of forecast and backcast error to trend estimate variances can be very large at the ends of a time series. Other results of particular interest are the effects of certain nonstationarities in the sampling errors. The examples of section 4 show that nonstationarities such as sampling error variances that change over time, or periodic independent redrawings of the sample, can yield striking changes in the pattern of the variances of seasonally adjusted data or trend estimates over time.

Section 5 provides concluding remarks.

2. Methodology

Define the observed unadjusted time series as y_t for $t = 1, \ldots, n.$ Time series that are seasonally adjusted are often estimates obtained from repeated (monthly or quarterly) sample surveys, and thus can be viewed as composed of a true underlying time series Y_t, and a series of sampling errors e_t assumed uncorrelated with Y_t. (See Bell and Hillmer 1990.) In vector notation, $\mathbf{y}_o = \mathbf{Y}_o + \boldsymbol{e}_o$, where the subscript o indicates that the time span of these vectors is the set of observed time points 1, ..., n. In certain cases y_t may arise from repeated censuses (as is typically the case for national export and import statistics, for example), in which case there is no sampling error, *i.e.*, $e_t = 0$.

The development that follows assumes that both Y_t and e_t follow known time series models. The model for Y_t will generally involve differencing, as in ARIMA (autoregressive-integrated-moving average) and ARIMA component (structural) models. The model for Y_t may be extended to include regression terms. (This will be considered in section 3.3.) The series e_t is assumed to not require differencing, but it may nonetheless exhibit certain nonstationarities, such as variances that change over time. Any such nonstationarities are assumed to be accounted for in the model for e_t. In practice, the models will be developed from observed data, as is discussed by, *e.g.*, Bell and Hillmer (1990, submitted), Binder and Dick (1989, 1990), and Tiller (1992).

In applying a symmetric X-11 filter of length $2m+1$ for seasonal adjustment with full forecast and backcast extension, the vector $\mathbf{y}_o$ needs to be augmented by m backcasts and m forecasts. The vector holding the m values of y_t prior to the observed data, and the corresponding $m \times 1$ vectors for Y_t, and e_t, are denoted $\mathbf{y}_b$, $\mathbf{Y}_b$, and $\mathbf{e}_b$. The analogous vectors of the m future values of y_t, Y_t, and e_t are denoted $\mathbf{y}_f$, $\mathbf{Y}_f$, and $\mathbf{e}_f$. Thus,

$$\begin{pmatrix} \mathbf{y}_b \\ \mathbf{y}_o \\ \mathbf{y}_f \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_b \\ \mathbf{Y}_o \\ \mathbf{Y}_f \end{pmatrix} + \begin{pmatrix} \mathbf{e}_b \\ \mathbf{e}_o \\ \mathbf{e}_f \end{pmatrix}. \tag{2.1}$$

The full vectors in (2.1), hereafter denoted as $\mathbf{y}$, $\mathbf{Y}$, and $\mathbf{e}$, have length $n + 2m$.

The backcasts and forecasts used to augment $\mathbf{y}_o$ are assumed to be minimum mean squared error (MMSE) linear predictions of $\mathbf{y}_b$ and $\mathbf{y}_f$ (using $\mathbf{y}_o$) obtained from the known time series model. (In practice, the model will be fitted to the data $\mathbf{y}_o$.) Under normality, the backcasts and forecasts are $E(\mathbf{y}_b \mid \mathbf{y}_o)$ and $E(\mathbf{y}_f \mid \mathbf{y}_o)$. The vector of observed data augmented with the backcasts and forecasts is denoted $\hat{\mathbf{y}} = (\hat{\mathbf{y}}_b', \mathbf{y}_o', \hat{\mathbf{y}}_f')'$, where $\hat{\mathbf{y}}_b = E(\mathbf{y}_b \mid \mathbf{y}_o)$ and $\hat{\mathbf{y}}_f = E(\mathbf{y}_f \mid \mathbf{y}_o)$. To simplify notation, from now on we will take expressions such as $(\hat{\mathbf{y}}_b, \mathbf{y}_o, \hat{\mathbf{y}}_f)$ to mean the column vector $(\hat{\mathbf{y}}_b', \mathbf{y}_o', \hat{\mathbf{y}}_f')'$.

Let the linear symmetric X-11 seasonal adjustment filter be written $\omega(B) = \sum_{-m}^{m} \omega_j B^j$, where B is the backshift operator and the ω_j are the filter weights ($\omega_j = \omega_{-j}$). Calculation of the ω_j is discussed by Young (1968) and Wallis (1982). Results of Bell and Monsell (1992) were used here. Application of $\omega(B)$ to the forecast and backcast extended series can be written as $\mathbf{\Omega}\hat{\mathbf{y}}$, where $\mathbf{\Omega}$ is a matrix of dimension $n \times (n + 2m)$. Each row of $\mathbf{\Omega}$ contains the filter weights $(\omega_{-m}, ..., \omega_0, ..., \omega_m)$, preceded and followed by the appropriate number of zeroes such that the center weight of the X-11 filter (ω_0) multiplies the observation being adjusted. Thus, in the first row of $\mathbf{\Omega}$ there are no preceding zeroes and $n - 1$ trailing zeroes, in the second row there is one preceding zero and $n - 2$ trailing zeroes, *etc*. For the default X-11 filter, $m = 84$. Choice of alternative seasonal or trend moving averages in X-11 changes the value of m from a low of 70 to a high of 149.

The question arises as to what $\mathbf{\Omega}\hat{\mathbf{y}}$ is estimating. As noted in the introduction, we define the "target" of the seasonal adjustment as the adjusted series that would result if there were no sampling error and there were sufficient data before and after all time points of interest for the symmetric filter to be applied. The target is thus $\omega(B)Y_t$, or in vector notation $\mathbf{\Omega}\mathbf{Y}$, and the seasonal adjustment error vector is $\mathbf{v} = \mathbf{\Omega}(\mathbf{Y} - \hat{\mathbf{y}})$. We are interested in the variance-covariance matrix $\mathrm{var}(\mathbf{v}) = \mathbf{\Omega}\,\mathrm{var}(\mathbf{Y} - \hat{\mathbf{y}})\,\mathbf{\Omega}'$. This can be easily computed once $\mathrm{var}(\mathbf{Y} - \hat{\mathbf{y}})$ is obtained. From here through section 2.3 we discuss the calculation of $\mathrm{var}(\mathbf{Y} - \hat{\mathbf{y}})$.

We start by writing $\mathbf{Y} - \hat{\mathbf{y}} = (\mathbf{y} - \mathbf{e}) - \hat{\mathbf{y}} = (\mathbf{b}, \mathbf{0}, \mathbf{f}) - \mathbf{e}$, where $\mathbf{b} = \mathbf{y}_b - \hat{\mathbf{y}}_b$ is the $m \times 1$ vector of backcast errors, and $\mathbf{f} = \mathbf{y}_f - \hat{\mathbf{y}}_f$ is the $m \times 1$ vector of forecast errors. Given the models for Y_t and e_t, we calculate $\mathrm{var}(\mathbf{Y} - \hat{\mathbf{y}})$ by separately computing $\mathrm{var}(\mathbf{e}), \mathrm{var}(\mathbf{b}, \mathbf{0}, \mathbf{f})$, and $\mathrm{cov}[(\mathbf{b}, \mathbf{0}, \mathbf{f}), \mathbf{e}]$, as discussed in sections 2.1 to 2.3. Then, $\mathrm{var}(\mathbf{Y} - \hat{\mathbf{y}})$ easily follows as $\mathrm{var}(\mathbf{b}, \mathbf{0}, \mathbf{f}) + \mathrm{var}(\mathbf{e}) - \mathrm{cov}[(\mathbf{b}, \mathbf{0}, \mathbf{f}), \mathbf{e}] - \mathrm{cov}[(\mathbf{b}, \mathbf{0}, \mathbf{f}), \mathbf{e}]'$. Thus,

$$\mathrm{var}(\mathbf{v}) = \mathbf{\Omega}\{\mathrm{var}(\mathbf{b}, \mathbf{0}, \mathbf{f}) + \mathrm{var}(\mathbf{e}) - \mathrm{cov}[(\mathbf{b}, \mathbf{0}, \mathbf{f}), \mathbf{e}] - \mathrm{cov}[(\mathbf{b}, \mathbf{0}, \mathbf{f}), \mathbf{e}]'\}\,\mathbf{\Omega}'.$$

Example – U.S. 5+ Unit Housing Starts. As the computations for each piece of $\mathrm{var}(\mathbf{Y} - \hat{\mathbf{y}})$ are explained, we illustrate the results graphically for an example series: housing starts in the U.S. for buildings of five or more units from January 1975 through November 1988 (167 observations). The original series, seasonally adjusted series, and estimated trend are shown in Figure 1. In practice, seasonal adjustment at the Census Bureau of this series uses a multiplicative decomposition with a 3×9 seasonal moving average and a 13-term Henderson trend filter. The following model for this series was developed in Bell and Hillmer (submitted):

$$\begin{aligned} y_t &= Y_t + e_t(1 - B)(1 - B^{12}) \\ Y_t &= (1 - 0.67B + 0.36B^2)(1 - 0.8753B^{12})\,a_t, \\ \sigma_a^2 &= 0.0191 \\ e_t &= (1 - 0.11B - 0.10B^2)\,b_t,\ \sigma_b^2 = 0.00714. \end{aligned} \tag{2.2}$$

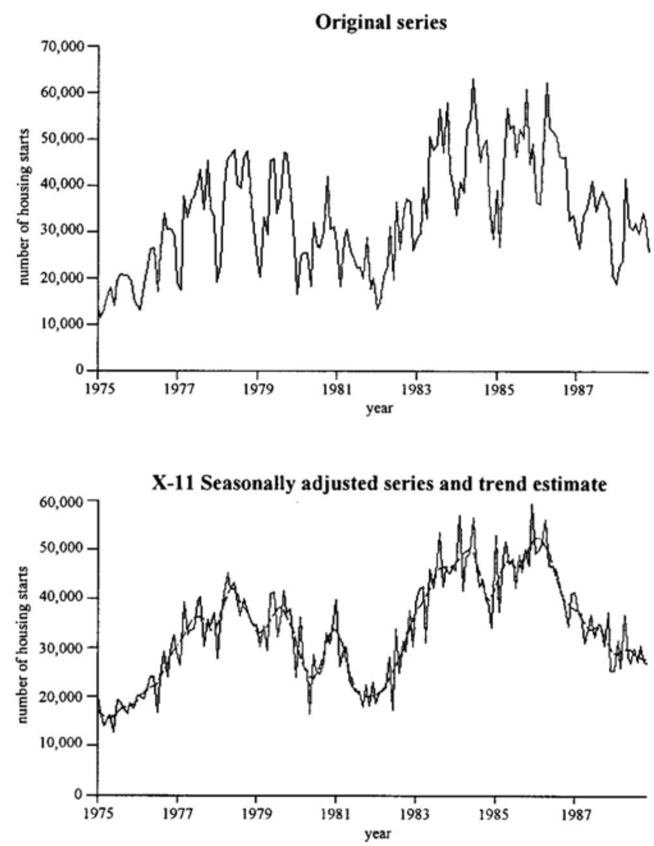

Figure 1 U.S. housing starts with five or more units. The top panel gives the original series from January 1975 through November 1988. The strong seasonality of the series is apparent from the yearly dips that typically occur during the winter months. The bottom panel gives the X-11 seasonally adjusted series (solid line) and trend estimate (dotted line) for the same period. The seasonal adjustment is multiplicative using a 3×9 seasonal moving average and 13-term Henderson trend filter

Here, y_t denotes the logarithms of the original time series (e^{y_t}), so that (2.2) implies a multiplicative decomposition for the original series $(e^{y_t} = e^{Y_t} e^{e_t})$.

2.1 Computation of var(e)

If e_t follows a stationary ARMA model, then var(**e**) can be computed from standard results, *e.g.*, McLeod (1975, 1977), Wilson (1979). If $\mathrm{var}(e_t)$ changes over time, we write $e_t = h_t \tilde{e}_t$, where $h_t^2 = \mathrm{var}(e_t)$, and $\tilde{e}_t$ has variance one and the same autocorrelation function as e_t. (See Bell and Hillmer *submitted*.) Then, writing $\mathbf{e} = \mathbf{H}\tilde{\mathbf{e}}$, where $\mathbf{H} = \mathrm{diag}(h_{1-m}, ..., h_{n+m})$, we have $\mathrm{var}(\mathbf{e}) = \mathbf{H}\,\mathrm{var}(\tilde{\mathbf{e}})\,\mathbf{H}'$. $\mathrm{Var}(\tilde{\mathbf{e}})$ is the autocorrelation matrix of $\mathbf{e}$, and it can be computed as just noted using the model for $\tilde{e}_t$.

If the sample is independently redrawn at certain times, then $\mathrm{var}(\mathbf{e})$ will be block diagonal, with blocks corresponding to the time points when each distinct sample is in effect. Each diagonal block of $\mathrm{var}(\mathbf{e})$ be computed as just discussed. These two types of nonstationarities in e_t – variance changing over time and "covariance breaks" due to independent redrawings of the sample – are those that arise in the examples of section 4.

Example – U.S. 5+ Unit Housing Starts (continued). Autocovariances for the MA(2) model for e_t given in (2.2) are easily computed. The resulting $\mathrm{var}(\mathbf{e})$ is a band matrix, with $\mathrm{var}(e_t) = 0.007298$ on the diagonal, $\mathrm{cov}(e_t, e_{t-1}) = -0.000707$ on the first sub- and super-diagonals, and $\mathrm{cov}(e_t, e_{t-2}) = -0.000714$ on the second sub- and super-diagonals. The rest of $\mathrm{var}(\mathbf{e})$ is zero. Following pre- and post multiplication by the seasonal adjustment filter matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}'$ the contribution of the sampling error to the variance of the seasonally adjusted series is constant for each observation (Figure 2). This occurs because the result of a time invariant linear filter applied to a stationary series $(\omega(B)e_t)$ is a stationary series, which has a constant variance.

2.2 Computation of var(b, 0, f)

The central n rows and n columns of $\mathrm{var}(\mathbf{b}, \mathbf{0}, \mathbf{f})$ are all zeroes. We require computation of $\mathrm{var}(\mathbf{b})$, $\mathrm{var}(\mathbf{f})$, and $\mathrm{cov}(\mathbf{b}, \mathbf{f})$ for the corner blocks of $\mathrm{var}(\mathbf{b}, \mathbf{0}, \mathbf{f})$. Although computation of variances of forecast (or backcast) errors for given models is standard in time series analysis, it is complicated here by the component representation of y_t as $Y_t + e_t$, and by differencing in the model for Y_t. Although computations for such models are often handled by the Kalman filter (Bell and Hillmer submitted; Binder and Dick 1989, 1990; Tiller 1992), this is inconvenient here since we require covariances of all distinct pairs of random variables from among the m forecast and m backcast errors. We instead use a direct matrix approach due to Bell and Hillmer (1988).

Assume that the differencing operator required to render Y_t stationary is $\delta(B)$, which is of degree d. Since e_t is assumed not to require differencing, $\delta(B)$ is also the differencing operator required by y_t. Define $\delta(B)\,y_t = w_t$, thus $w_t = \delta(B)Y_t + \delta(B)e_t$. We introduce the matrix $\boldsymbol{\Delta}$, corresponding to $\delta(B)$, defined such that $\boldsymbol{\Delta}\mathbf{y} = \mathbf{w}$ is the vector of differenced $\mathbf{y}$. The vector $\mathbf{w} = (\mathbf{w}_b, \mathbf{w}_o, \mathbf{w}_f)$, which is of length $n + 2m - d$, is partitioned so that $\mathbf{w}_b$ and $\mathbf{w}_f$ are $m \times 1$ vectors, and $\mathbf{w}_o$ is the $n - d$ vector of differenced observed data. Thus, $\boldsymbol{\Delta}$ has dimensions $(n + 2m - d) \times (n + 2m)$. Note that, because d observations are lost in differencing, $\mathbf{w}_b$ and $\mathbf{w}_o$ start d time points later than $\mathbf{y}_b$ and $\mathbf{y}_o$, respectively. That is, $\mathbf{y}_b$ and $\mathbf{y}_o$ start at time points $1 - m$ and 1, but $\mathbf{w}_b$ and $\mathbf{w}_o$ start at time points $1 - m + d$ and $d + 1$.

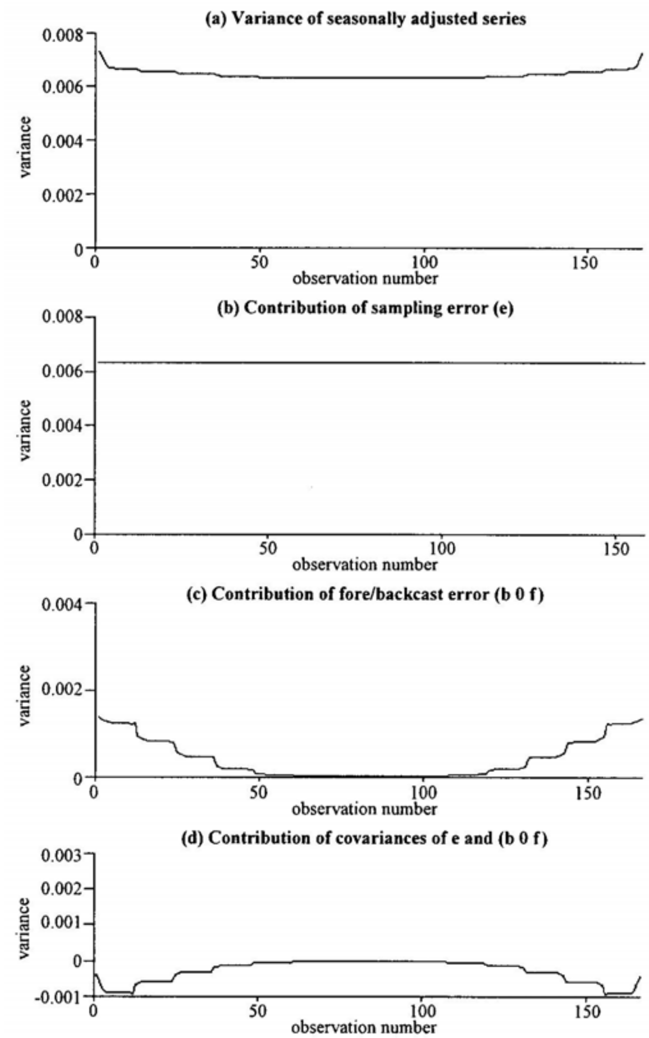

Figure 2 U.S. 5+ units housing starts: Variance decomposition after X-11 Seasonal Adjustment. The top panel gives the variance of the seasonally adjusted series as the total of the three components. The second panel give the contribution of sampling error (**e**), which is the largest component and constant across the series. The third panel gives the contribution of back/forecast error (**b 0 f**), which is zero in the middle of the series, where no back/forecasts are needed, but increases towards either end of the series as more back/forecasts are used. The bottom panel is the sum of the two covariance terms (cov(**e**, (**b 0 f**)) + cov((**b 0 f**), **e**)), which tend to offset the contribution from back/forecast error

Define $\mathbf{u} = (u_{1-m+d}, \ldots, u_{n+m})' = \mathbf{\Delta Y}$. The time series u_t is stationary. Since $\mathbf{w} = \mathbf{u} + \mathbf{\Delta e}$, with $\mathbf{u}$ and $\mathbf{e}$ uncorrelated with each other, $\mathrm{var}(\mathbf{w}) = \mathrm{var}(\mathbf{u}) + \mathbf{\Delta}\,\mathrm{var}(\mathbf{e})\mathbf{\Delta}'$. We partition $\mathrm{var}(\mathbf{w})$ as

$$\mathrm{var}(\mathbf{w}) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{pmatrix},$$

where Σ_{11} is $\mathrm{var}(\mathbf{w}_b)$, Σ_{12} is $\mathrm{cov}(\mathbf{w}_b, \mathbf{w}_o)$, *etc.*

Since $\mathbf{y}$, when differenced to $\mathbf{w}$ using $\delta(B)$, has lost d data values, $\mathbf{y}$ cannot be obtained from $\mathbf{w}$ without also knowing a sequence of d "starting values". Consider obtaining $\mathbf{y}_f$ from $\mathbf{w}_f$ and starting values $\mathbf{y}_* = (y_{n+1-d}, \ldots, y_n)'$. Theorem 1 in Bell (1984a) can be used to show that

$$\mathbf{y}_f = \mathbf{A}\mathbf{y}_* + \mathbf{C}\mathbf{w}_f \tag{2.3}$$

for matrices $\mathbf{A}$ and $\mathbf{C}$ determined by $\delta(B)$. The rows of the $m \times m$ matrix $\mathbf{C}$ consist of the coefficients of $\xi(B) = 1 + \xi_1 B + \xi_2 B^2 + \ldots = \delta(B)^{-1}$ in the form

$$\mathrm{C} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ \xi_1 & 1 & & 0 & 0 \\ \xi_2 & \xi_1 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & 1 & 0 \\ \xi_{m-1} & \xi_{m-2} & \cdots & \xi_1 & 1 \end{pmatrix}.$$

$\mathbf{A}$ is an $m \times d$ matrix which accounts for the effect of the d starting values in $\mathbf{y}_*$ on $\mathbf{y}_f$. The exact form of $\mathbf{A}$ is given in Bell (1984a) and, since it will exactly cancel in our application, it will not be given here. In (2.3) $\mathbf{y}_*$ is known since it is part of $\mathbf{y}_o$, the observed data. Thus, from (2.3) the MMSE forecast of $\mathbf{y}_f$ is $\hat{\mathbf{y}}_f = \mathbf{A}\mathbf{y}_* + \mathbf{C}\hat{\mathbf{w}}_f$, where $\hat{\mathbf{w}}_f$ is the MMSE forecast of $\mathbf{w}_f$. Therefore, $\mathbf{f} = \mathbf{y}_f - \hat{\mathbf{y}}_f = \mathbf{A}\mathbf{y}_* + \mathbf{C}\mathbf{w}_f - (\mathbf{A}\mathbf{y}_* + \mathbf{C}\hat{\mathbf{w}}_f) = \mathbf{C}(\mathbf{w}_f - \hat{\mathbf{w}}_f)$, and $\mathrm{var}(\mathbf{f}) = \mathbf{C}\,\mathrm{var}(\mathbf{w}_f - \hat{\mathbf{w}}_f)\mathbf{C}'$.

Under Assumption A of Bell (1984a), which leads to the standard results for forecasting nonstationary series (as in, *e.g.*, Box and Jenkins 1976, Chapter 5), $\hat{\mathbf{w}}_f = \Sigma_{32}\Sigma_{22}^{-1}\mathbf{w}_o$. Note that this uses only the differenced data $\mathbf{w}_o$ in forecasting $\mathbf{w}_f$. Then, from standard results on linear prediction, $\mathrm{var}(\mathbf{w}_f - \hat{\mathbf{w}}_f) = \Sigma_{33} - \Sigma_{32}\Sigma_{22}^{-1}\Sigma_{32}'$. Thus, $\mathrm{var}(\mathbf{f}) = \mathbf{C}(\Sigma_{33} - \Sigma_{32}\Sigma_{22}^{-1}\Sigma_{32}')\mathbf{C}'$.

To obtain $\mathrm{var}(\mathbf{b})$ and $\mathrm{cov}(\mathbf{b}, \mathbf{f})$ we note that results obtained by Bell (1984a, page 651) imply similar calculations hold for the backcast errors $\mathbf{b}$. In fact, it can be shown that $\mathbf{b} = (-1)^r\mathbf{C}'(\mathbf{w}_b - \hat{\mathbf{w}}_b)$, where $\hat{\mathbf{w}}_b$ is the MMSE backcast of $\mathbf{w}_b$, and r is the number of times $(1 - B)$ appears in the polynomial $\delta(B)$. (The appearance of $\mathbf{C}'$ in this expression instead of $\mathbf{C}$ stems from the indexing of $\mathbf{w}_b$ and $\hat{\mathbf{w}}_b$ forward through time although the backcasting process proceeds backwards through time.) Thus, $\mathrm{var}(\mathbf{b}) = \mathbf{C}'\,\mathrm{var}(\mathbf{w}_b - \hat{\mathbf{w}}_b)\,\mathbf{C} = \mathbf{C}'(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}')\mathbf{C}$. Similarly $\mathrm{cov}(\mathbf{f}, \mathbf{b}) = (-1)^r\,\mathbf{C}(\Sigma_{31} - \Sigma_{32}\Sigma_{22}^{-1}\Sigma_{12})\mathbf{C}$. In practice, to avoid inverting Σ_{22}, $\mathrm{var}(\mathbf{f})$, $\mathrm{var}(\mathbf{b})$, and $\mathrm{cov}(\mathbf{f}, \mathbf{b})$ can be computed using the Cholesky decomposition of Σ_{22}. (See Appendix A.)

Example – U.S. 5+ Unit Housing Starts (continued). The contribution to seasonal adjustment variance from $\mathrm{var}(\mathbf{b}, \mathbf{0}, \mathbf{f})$ is shown in Figure 2. This is zero or essentially zero for observations in the middle of the series, where no or few fore/backcasts need be made to apply the symmetric adjustment filter. Towards the ends of the series, the contribution of fore/backcast error becomes more substantial since an increasing number of observations need to be fore/backcast to apply the filter. The jumps in the graph occur when an additional fore/backcasted observation is multiplied by a weight in the adjustment filter that is a multiple of the seasonal period, since these weights have the greatest magnitude (Bell and Monsell 1992). Note that the contributions from $\mathrm{var}(\mathbf{b}, \mathbf{0}, \mathbf{f})$ at the very ends of the series are smaller than the contributions from $\mathrm{var}(\mathbf{e})$, but are not negligible.

2.3 Computation of Cov[(b 0 f), e]

To compute $\mathrm{cov}(\mathbf{f}, \mathbf{e})$, we first note from results of the preceding section that $\mathbf{f} = \mathbf{y}_f - \hat{\mathbf{y}}_f = \mathbf{C}(\mathbf{w}_f - \hat{\mathbf{w}}_f) = \mathbf{C}(\mathbf{w}_f - \Sigma_{32}\Sigma_{22}^{-1}\mathbf{w}_o) = \mathbf{C}[0|-\Sigma_{32}\Sigma_{22}^{-1}|\mathbf{I}_m]\,\mathbf{w} = \mathbf{C}[0|-\Sigma_{32}\Sigma_{22}^{-1}\Sigma|\mathbf{I}_m]\,\mathbf{\Delta y}$. Since $\mathrm{cov}(\mathbf{y}, \mathbf{e}) = \mathrm{cov}(\mathbf{Y} + \mathbf{e}, \mathbf{e}) = 0 + \mathrm{var}(\mathbf{e})$, we see that $\mathrm{cov}(\mathbf{f}, \mathbf{e}) = \mathbf{C}[0\,|-\Sigma_{32}\Sigma_{22}^{-1}\,|\,\mathbf{I}_m]\mathbf{\Delta}\,\mathrm{var}(\mathbf{e})$. $\mathrm{Cov}(\mathbf{b}, \mathbf{e})$ is computed in an analogous fashion by noting that, $\mathbf{b} = (-1)^r\mathbf{C}'(\mathbf{w}_b - \hat{\mathbf{w}}_b) = (-1)^r\mathbf{C}'(\mathbf{w}_b - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{w}_o) = (-1)^r\mathbf{C}'[\mathbf{I}_m\,|-\Sigma_{12}\Sigma_{22}^{-1}\,|\,0]\,\mathbf{\Delta y}$, so that $\mathrm{cov}(\mathbf{b}, \mathbf{e}) = (-1)^r\mathbf{C}'[\mathbf{I}_m\,|-\Sigma_{12}\Sigma_{22}^{-1}\,|\,0]\mathbf{\Delta}\,\mathrm{var}(\mathbf{e})$.

Example – U.S. 5+ Unit Housing Starts (continued). Figure 2 shows that the contribution of $\mathrm{cov}[(\mathbf{b}, \mathbf{0}, \mathbf{f}), \mathbf{e}]$ is zero or near zero in the middle of the series, but it becomes increasingly negative towards the ends of the series, in a pattern similar, though opposite in sign and of smaller magnitude, to that of $\mathrm{var}(\mathbf{b}, \mathbf{0}, \mathbf{f})$. At the very ends of the series, however, the pattern reverses and the covariance increases. The elements of $\mathrm{cov}[(\mathbf{b}, \mathbf{0}, \mathbf{f}), \mathbf{e}]$, are mainly positive, so its contribution to the seasonal adjustment variance is negative because $\mathrm{cov}[(\mathbf{b}, \mathbf{0}, \mathbf{f}), \mathbf{e}]$ and its transpose are subtracted from $\mathrm{var}(\mathbf{e}) + \mathrm{var}(\mathbf{b}, \mathbf{0}, \mathbf{f})$. The net effect is that subtracting $\mathbf{\Omega}\{\mathrm{cov}[(\mathbf{b}, \mathbf{0}, \mathbf{f}), \mathbf{e}] + \mathrm{cov}[(\mathbf{b}, \mathbf{0}, \mathbf{f}), \mathbf{e}]'\}\mathbf{\Omega}'$ tends to offset the effect of adding $\mathbf{\Omega}\,\mathrm{var}(\mathbf{b}, \mathbf{0}, \mathbf{f})\,\mathbf{\Omega}'$, except near the very ends of the series. Thus, the graph of the variances of the seasonally adjusted series in Figure 2 is very similar to the graph of the contribution of $\mathrm{var}(\mathbf{e})$, except near the very ends of the series. We observed this type of "cancellation effect" in several other examples, including those of section 4.

2.4 Comparison with the first approach of Wolter and Monsour

The first approach of Wolter and Monsour (1981) proposed use of $\mathbf{\Omega}_{wm}\,\mathrm{var}(\mathbf{e}_o)\mathbf{\Omega}_{wm}'$ as the variance-covariance matrix of the X-11 seasonal adjustment errors, where $\mathbf{\Omega}_{wm}$ is an $n + n$ matrix whose rows contain the X-11 linear filter weights, both symmetric and asymmetric. That is, the middle rows (rows t such that $m < t < n - m + 1$, assuming $n > 2m$) of $\mathbf{\Omega}_{wm}$ contain the X-11 symmetric filter weights, but the first and last m rows of $\mathbf{\Omega}_{wm}$ contain X-11's asymmetric filter weights. The middle rows of $\mathbf{\Omega}_{wm}$ and $\mathbf{\Omega}$ thus contain the same filter weights, but the first and last m rows do not. This means that our approach will give the same results as that of Wolter and Monsour for $m < t < n - m + 1$, that is, for time points at which the symmetric filter is being used. The results of the two approaches will differ for the first and last m time points. Since the most recent seasonally adjusted data receive the most attention, this difference is potentially important.

Wolter and Monsour also considered use of a matrix $\mathbf{\Omega}^*$ instead of $\mathbf{\Omega}_{wm}$ where $\mathbf{\Omega}^*$ is $(n+12) \times (n+12)$ to include 12 additional rows of weights corresponding to year-ahead seasonal adjustment filters. Though year-ahead adjustment was the common practice through the early 1980s, it has now mostly been replaced in the United States by concurrent adjustment (McKenzie 1984).

The differences between our approach and that of Wolter and Monsour can be viewed in two ways. One view is that since Wolter and Monsour did not consider forecast and backcast extension, their approach ignores the contribution of forecast and backcast errors to seasonal adjustment error. This contribution affects results for the first and last m time points, although the examples of section 4 show that this contribution is often small. However, in some cases it is not small, including those time series not subject to sampling error. For such series Wolter and Monsour's approach would assign zero variance to the adjustments, even though initial adjustments would be revised as new data became available.

The other way to view the differences between the approaches centers on the difference in "targets". The seasonal adjustment error under Wolter and Monsour's approach can be thought of as $\mathbf{\Omega}_{wm}(\mathbf{Y}_o - \mathbf{y}_o) = -\mathbf{\Omega}_{wm}\mathbf{e}_o$. Since this results in zero error for series with no sampling error $(\mathbf{Y}_o - \mathbf{y}_o)$, Wolter and Monsour implicitly define the seasonal adjustment target to be $\mathbf{\Omega}_{wm}\mathbf{Y}_o$. This definition of target has the undesirable property that the target value for a given time point changes as additional data are acquired, since the rows of $\mathbf{\Omega}_{wm}$ contain different filter weights. Our target value for any given time point t is always $\omega(B)Y_t$.

Example – U.S. 5+ Unit Housing Starts (continued). We compared results using our methodology with that of Wolter and Monsour's using the *default* X-11 seasonal adjustment filter although, as noted earlier, this example series is adjusted using the optional 3 × 9 seasonal moving average filter. This comparison used the default filter for convenience: asymmetric X-11 filter weights are needed to obtain results for the Wolter-Monsour approach and we were given a computer program by Nash Monsour that produced them only for the default filter. Figure 3 gives the results for both approaches. The non-constant variances over time from the Wolter-Monsour approach result from applying different filters at different time points. An interesting consequence of this is that, despite the stationarity of the sampling error, the Wolter-Monsour seasonal adjustment variance is noticeably *higher* in the middle of the series than for many time points toward (but not close to) either end of the series. This carries the implausible implication that use of less data produces estimators with lower variance. Similar behavior can be observed in several examples presented by Pfeffermann (1994).

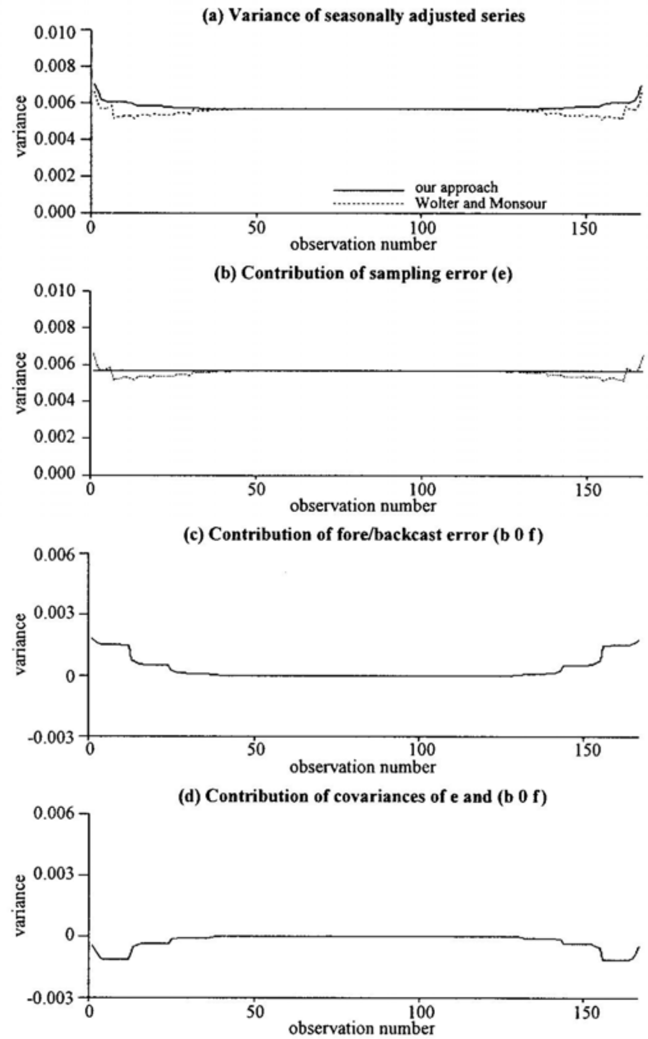

Figure 3 U.S. 5+ units housing starts: Comparison with approach of Wolter and Monsour (1981). The panel descriptions are as for Figure 2. The Wolter and Monsour approach (dotted lines) uses the asymmetric X-11 filters for the ends of the series and accounts only for sampling error. Our approaches agree in the middle of the series where there is no contribution from back/forecast error. The Wolter and Monsour variances inappropriately decrease near the ends of the series, suggesting that use of less data produces estimates with lower variances. The results here, in contrast to Figures 1, 2, 4, 5, and 6, use default X-11 filters. (See text.)

The results from using the default X-11 seasonal adjustment filter with our approach are also useful for comparing with the 3×9 seasonal moving average filter, for which results are given in Figure 2. Differences between results from using the two filters are not great. The contribution of the sampling error is somewhat lower and that of the fore/backcast error somewhat higher when using the default seasonal adjustment filter.

3. Extensions to the methodology

This section discusses three extensions to the general methodology of section 2. The first two extensions are straightforward, the third more involved.

3.1 Variances for seasonal, trend, and irregular estimates; variances with time–Varying filters

The only way the nonseasonal (seasonally adjusted) component is distinguished in the derivation of section 2 is through the filter weights placed in the matrix $\mathbf{\Omega}$. Therefore, corresponding variances for X-11 estimates of the seasonal, trend, and irregular components follow from the same expressions simply by changing the matrix $\mathbf{\Omega}$ to contain the desired filter weights. This also changes the dimension of $\mathbf{\Omega}$, since the length of the seasonal adjustment, trend, and irregular filters (for given options) differs, and the filter length determines the size of $\mathbf{\Omega}$.

A similar extension handles the case of different seasonal moving averages (MAs) selected for different months (or quarters), an option allowed by X-11. This changes the seasonal adjustment (and seasonal, trend, and irregular) filters applied in the different months. The results of section 2 also accommodate this extension through a simple modification of $\mathbf{\Omega}$. Since the rows of $\mathbf{\Omega}$ correspond to the time points being adjusted, we simply define row t of $\mathbf{\Omega}$ to contain the weights (along with sufficient zeroes) from whatever filter is being applied in month t. Some care must be taken to dimension $\mathbf{\Omega}$ appropriately if the longest selected MA is not used in the first and last months of the series.

Example – U.S. 5+ Unit Housing Starts (continued). Figure 4 shows the variance of the X-11 trend estimate, using the 3×9 seasonal MA and 13-term Henderson. The most obvious difference from the seasonal adjustment results is the substantial effect of fore/backcast error at the very ends of the series. This occurs because the largest weights of the trend filter $(\omega^{(T)}(B))$ are the center weight $(\omega_o^{(T)})$ and the adjacent weights $(\omega_1^{(T)}, \omega_2^{(T)}, \omega_3^{(T)})$ that are applied to data immediately before and after the observation being adjusted (Bell and Monsell 1992). At the very ends of the series, the weights $(\omega_1^{(T)}, \omega_2^{(T)}, \omega_3^{(T)})$ apply to fore/backcasted observations, which results in large increases in the contribution of fore/backcast error there. The result is that uncertainty about the trend increases sharply at the ends of the series. In the center of the series, however, the trend variances of Figure 4 are substantially lower than the seasonal adjustment variances of Figure 3, due to the smoothing of the sampling error by the trend filter.

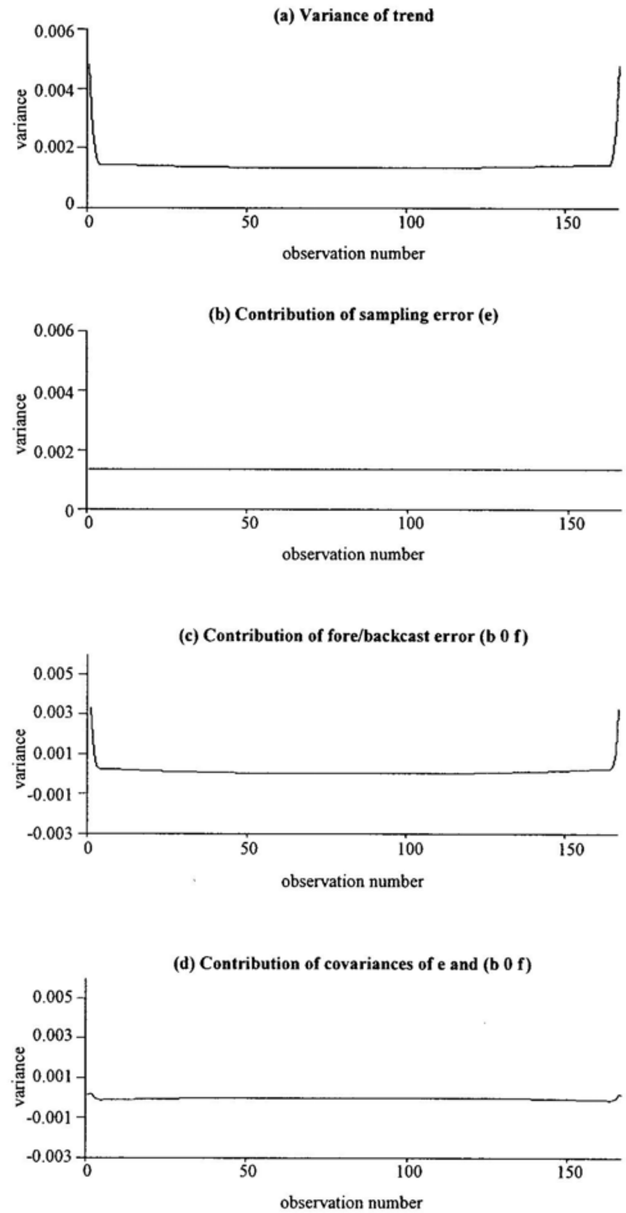

Figure 4 U.S. 5+ units housing starts: Variance decomposition of the trend estimate. The panel descriptions are as for Figure 2. Note the large jump in trend estimate variances at the ends of the series due to the contribution of back/forecast error (third panel)

3.2 Variances for seasonally adjusted month-to-month and year-to-year changes

The variances of the errors of the seasonally adjusted estimates of month-to-month change are the quantities $\mathrm{var}(v_t - v_{t-1})$, $t = 2, ..., n$. Given $\mathrm{var}(\mathbf{v})$, the complete error covariance matrix for the seasonally adjusted month-to-month changes can be calculated as $\mathbf{\Delta}_1 \, \mathrm{var}(\mathbf{v}) \mathbf{\Delta}_1'$, where

$$\mathbf{\Delta}_1 = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & \ddots & 0 & 0 \\ 0 & 0 & & -1 & 1 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}$$

is of dimension $(n-1) \times n$. The error covariance matrix for the seasonally adjusted year-to-year changes in a quarterly series is calculated similarly as $\mathbf{\Delta}_4 \, \text{var}(\mathbf{v}) \mathbf{\Delta}_4'$, where

$$\mathbf{\Delta}_4 = \begin{pmatrix} -1 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & & & & \ddots & 0 & 0 \\ 0 & 0 & \cdots & -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

is of dimension $(n-4) \times n$. The corresponding $(n-12) \times n$ matrix $\mathbf{\Delta}_{12}$ for monthly series follows a similar pattern with additional zeroes.

Variances of month-to-month or year-to-year changes in the trend are also easily obtained, as can be seen from this discussion and that of section 3.1.

Example – U.S. 5+ Unit Housing Starts (continued). We produced the standard errors for seasonally adjusted month-to-month and year-to-year changes for this series (Figure 5). Since this time series has been log transformed, standard errors can be approximately interpreted as percentages on the original (unlogged) scale. Compared to the standard errors for the seasonally adjusted series, there are slight increases in the standard errors of the month-to-month changes near the ends of the series, but the standard errors of the year-to-year changes show almost no such increase. Thus, for this series and filter, the uncertainty about month-to-month and year-to-year percent change in the seasonally adjusted data is almost constant across the series. The standard errors of the month-to-month and year-to-year changes are both about 50 percent higher than those for the seasonally adjusted series.

3.3 Variances of X-11 seasonal adjustments with estimated regression effects

Seasonal adjustment often involves the estimation of certain regression effects to account for such things as calendar variation, known interventions, and outliers (Young 1965; Cleveland and Devlin 1982; Hillmer, Bell, and Tiao 1983; Findley, Monsell, Bell, Otto, and Chen submitted). (Outlier effects are often estimated in the same way as known interventions even though inference about outliers should ideally take account of the fact that the series was searched for the most "significant" outliers.) This section shows how the results already obtained can be extended to include the contribution to seasonal adjustment error of error in estimating regression parameters. We still assume the other model parameters, which determine the covariance structures of $\mathbf{Y}$ and $\mathbf{e}$, are known. In practice these other model parameters will also be estimated, but accounting for error in estimating them is much more difficult. A Bayesian approach for doing so in the context of model-based seasonal adjustment is investigated by Bell and Otto (submitted).

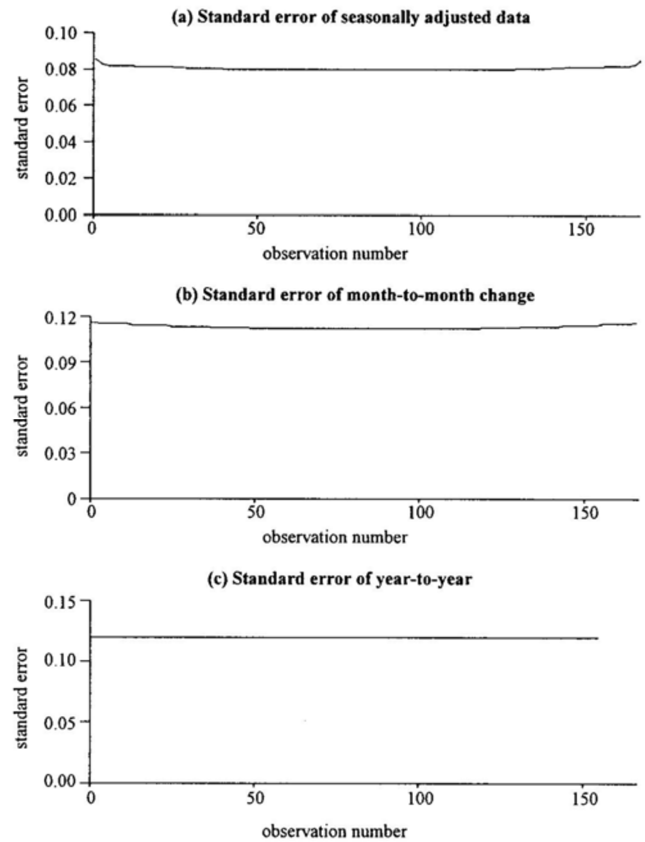

Figure 5 U.S. 5+ Units housing starts: Standard errors. These panels contrast the standard errors (not variances, as in previous figures) of the seasonally adjusted data (top panel) with the larger standard errors of seasonally adjusted month-to-month (middle panel) and year-to-year (bottom panel) change estimates

We extend the model for Y_t to include regression terms by writing $Y_t = x_t'\beta + Z_t$, where x_t is the vector of regression variables at time t, β is the vector of regression parameters, and Z_t is the series of true population quantities with regression effects removed. Extending our matrix-vector notation, we write $\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}, \mathbf{Y}_o = \mathbf{X}_o\beta + \mathbf{Z}_o$, *etc.* The regression matrix $\mathbf{X}$ can be partitioned by its rows corresponding to the backcast, observation, and forecast periods: $\mathbf{X} = (\mathbf{X}_b' \,|\, \mathbf{X}_o' \,|\, \mathbf{X}_f')'$. We assume e_t has mean zero, so its model does not involve any regression effects. We then have $\mathbf{y} = \mathbf{Y} + \mathbf{e} = (\mathbf{X}\beta + \mathbf{Z}) + \mathbf{e}$, with the usual partitioning applying. Letting z_t denote the series y_t with the regression effects removed, we have $\mathbf{z} = \mathbf{y} - \mathbf{X}\beta = \mathbf{Z} + \mathbf{e}$.

An additional partition is needed of the matrix $\mathbf{X}$ and vector β. This is because some of the regression effects in $\mathbf{x}_t'\beta$ may be assigned to the nonseasonal component while others, such as trading-day or holiday effects, may be removed as part of the seasonal adjustment. See Bell (1984b) for a discussion. Partition $\mathbf{x}_t'$ as $(\mathbf{x}_{St}'|\mathbf{x}_{Nt}')$ where $\mathbf{x}_{Nt}$ represents the regression variables assigned to the non-seasonal and $\mathbf{x}_{St}$ the variables whose effects are to be removed in the seasonal adjustment. Correspondingly partition β so $\mathbf{x}_t'\beta = \mathbf{x}_{St}'\beta_S + \mathbf{x}_{Nt}'\beta_N$ and $\mathbf{X}\beta = \mathbf{X}_S\beta_S + \mathbf{X}_N\beta_N = (\mathbf{X}_S \,|\, \mathbf{X}_N)\,(\beta_S' \,|\, \beta_N')'$. ($\mathbf{x}_{St}\beta_S$ is assigned to the "combined" seasonal component.) The matrix $\mathbf{X}$ can thus be partitioned two ways: by seasonal versus nonseasonal regression effects, and by the backcast, observation, and forecast periods. Thus we write

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{Sb} & \mathbf{X}_{Nb} \\ \mathbf{X}_{So} & \mathbf{X}_{No} \\ \mathbf{X}_{Sf} & \mathbf{X}_{Nf} \end{bmatrix}.$$

If β were known we could compute $\mathbf{z}_o = \mathbf{y}_o - \mathbf{X}_o\beta = \mathbf{Z}_o + \mathbf{e}_o$, forecast and backcast extend this series (call the extended series $\hat{\mathbf{z}}$), adjust $\hat{\mathbf{z}}$ by X-11 ($\mathbf{\Omega}\hat{\mathbf{z}}$), and add back the required regression effects $\mathbf{X}_{No}\beta_{No}$. The target of the seasonal adjustment would be $\mathbf{X}_{No}\beta_{No} + \mathbf{\Omega Z} = \mathbf{X}_{No}\beta_{No} + \mathbf{\Omega}(\mathbf{Y} - \mathbf{X}\beta)$, and the seasonal adjustment error would then be $(\mathbf{X}_{No}\beta_{No} + \mathbf{\Omega Z}) - (\mathbf{X}_{No}\beta_{No} + \mathbf{\Omega}\hat{\mathbf{z}}) = \mathbf{\Omega}(\mathbf{Z} - \hat{\mathbf{z}})$. Thus, if the regression parameters were known they would not contribute to the seasonal adjustment error, and the results already given could be used to compute $\operatorname{var}(\mathbf{\Omega}(\mathbf{Z} - \hat{\mathbf{z}}))$.

In practice, β will be estimated as part of the model fitting, say by maximum likelihood assuming normality. Given the estimates of the other model parameters, and taking these parameters as if they were known, the maximum likelihood estimate of β and its variance are given by

$$\hat{\beta} = [\mathbf{X}_o'\mathbf{\Delta}_o'\,\Sigma_{22}^{-1}\,\mathbf{\Delta}_o\mathbf{X}_0]^{-1}\,\mathbf{X}_o'\mathbf{\Delta}_o'\,\Sigma_{22}^{-1}\,\mathbf{\Delta}_o\mathbf{y}_o \qquad (3.1)$$

$$\operatorname{var}(\hat{\beta}) = [\mathbf{X}_o'\mathbf{\Delta}_o'\,\Sigma_{22}^{-1}\,\mathbf{\Delta}_o\mathbf{X}_0]^{-1}, \qquad (3.2)$$

where $\mathbf{\Delta}_o$ is of dimension $(n - d) \times n$, containing that part of the larger matrix $\mathbf{\Delta}$ which differences the observed series $\mathbf{y}_o$. The expressions (3.1) and (3.2) are generalized least squares results using the regression equation for the differenced data, $\mathbf{w}_o = \mathbf{\Delta}\,\mathbf{y}_o = (\mathbf{\Delta}\,\mathbf{X}_o)\beta + (\mathbf{u}_o + \mathbf{\Delta}\,\mathbf{e}_o)$, where the error term, $\mathbf{u}_o + \mathbf{\Delta e}_o$, has covariance matrix $\operatorname{var}(\mathbf{w}_o) = \Sigma_{22}$, which is determined by the other model parameters.

Given the estimated regression parameters $\hat{\beta}$, the seasonally adjusted series would be obtained by subtracting the estimated regression effects from the data (call the resulting series $\hat{\mathbf{z}}_o = \mathbf{y}_o - \mathbf{X}_o\hat{\beta}$), extending this series with forecasts and backcasts using the model (denote this extended series $\hat{\hat{\mathbf{z}}} = [\hat{\hat{\mathbf{z}}}_b, \hat{\mathbf{z}}_o, \hat{\hat{\mathbf{z}}}_f]$), applying X-11 to the extended series ($\mathbf{\Omega}\hat{\hat{\mathbf{z}}}$), and adding back the estimated regression effects assigned to the nonseasonal component ($\mathbf{\Omega}\hat{\hat{\mathbf{z}}} + \mathbf{X}_{No}\hat{\beta}_{No}$). The target of the seasonal adjustment is still $\mathbf{X}_{No}\beta_{No} + \mathbf{\Omega Z}$, discussed above. The seasonal adjustment error is then $\mathbf{v} = (\mathbf{X}_{No}\beta_{No} + \mathbf{\Omega Z}) - (\mathbf{\Omega}\hat{\hat{\mathbf{z}}} + \mathbf{X}_{No}\hat{\beta}_{No}) = \mathbf{X}_{No}(\beta_{No} - \hat{\beta}_{No}) + \mathbf{\Omega}(\mathbf{Z} - \hat{\hat{\mathbf{z}}})$.

The expression for $\mathbf{v}$ can be simplified by rewriting $\hat{\hat{\mathbf{z}}}$. First, let $\mathbf{G} = [\mathbf{B}' \,|\, \mathbf{I} \,|\, \mathbf{F}']'$, where $\mathbf{F}$ is the matrix that produces forecasts $\hat{\mathbf{y}}_f$ from $\mathbf{y}_o$ and $\mathbf{B}$ is the corresponding matrix that produces backcasts $\hat{\mathbf{y}}_b$ from $\mathbf{y}_o$. We will not need explicit expressions for $\mathbf{F}$ or $\mathbf{B}$. $\mathbf{G}$ applied to $\mathbf{z}_o$ produces $\hat{\mathbf{z}}$ while $\mathbf{G}$ applied to $\hat{\mathbf{z}}_o$ produces $\hat{\hat{\mathbf{z}}}$. Therefore, $\hat{\hat{\mathbf{z}}} = \hat{\mathbf{z}} - (\hat{\mathbf{z}} - \hat{\hat{\mathbf{z}}}) = \hat{\mathbf{z}} - [\mathbf{G}(\mathbf{y}_o - \mathbf{X}_o\beta) - \mathbf{G}(\mathbf{y}_o - \mathbf{X}_o\hat{\beta})] = \hat{\mathbf{z}} + \mathbf{GX}_o\,(\beta - \hat{\beta})$. Note that $\mathbf{GX}_o$ is obtained by applying the procedure for forecast and backcast extension (from the model for z_t) to each column of $\mathbf{X}_o$. The approach we used to do this is described in Appendix B. Continuing, we have

$$\begin{aligned}\mathbf{v} &= \mathbf{X}_{No}(\beta_{No} - \hat{\beta}_{No}) + \mathbf{\Omega}[(\mathbf{Z} - \hat{\mathbf{z}}) - \mathbf{GX}_o(\beta - \hat{\beta})] \\ &= \mathbf{\Omega}(\mathbf{Z} - \hat{\mathbf{z}}) + \{[\mathbf{0}|\mathbf{X}_{No}] - \mathbf{\Omega GX}_o\}(\beta - \hat{\beta}).\end{aligned}$$

Now, $\mathbf{Z} - \hat{\mathbf{z}} = \mathbf{z} - \mathbf{e} - \hat{\mathbf{z}} = [\mathbf{b} \,|\, \mathbf{0} \,|\, \mathbf{f}] - \mathbf{e}$. Note that $[\mathbf{b} \,|\, \mathbf{0} \,|\, \mathbf{f}]$, the error vector from projecting $\mathbf{z}$ on $\mathbf{z}_o$ or $\mathbf{y}_o$, is orthogonal to (uncorrelated with) $\beta - \hat{\beta}$, since $\hat{\beta}$ is a linear function of the data $\mathbf{y}_o$. Therefore, letting $\mathbf{K} = [\mathbf{0}|\mathbf{X}_{No}] - \mathbf{\Omega GX}_o$, we have the variance-covariance matrix of the seasonal adjustment error allowing for error in estimating β:

$$\begin{aligned}\operatorname{var}(\mathbf{v}) &= \mathbf{\Omega}\operatorname{var}(\mathbf{Z} - \hat{\mathbf{z}})\mathbf{\Omega}' + \mathbf{K}\operatorname{var}(\hat{\beta})\mathbf{K}' \\ &\quad + \mathbf{\Omega}\operatorname{cov}(\mathbf{e},\hat{\beta})\mathbf{K}' + \mathbf{K}\operatorname{cov}(\hat{\beta},\mathbf{e})\mathbf{\Omega}' \qquad (3.3)\end{aligned}$$

where $\operatorname{var}(\hat{\beta})$ is given by (3.2). In (3.3) $\mathbf{\Omega}\operatorname{var}(\mathbf{Z} - \hat{\mathbf{z}})\mathbf{\Omega}'$ is computed by the results of section 2, and computation of $\mathbf{K}\operatorname{var}(\hat{\beta})\mathbf{K}'$ is straightforward once $\mathbf{GX}_o$ has been computed. To compute the other two terms requires

$$\begin{aligned}&\operatorname{cov}(\hat{\beta},\mathbf{e}) \\ &= \operatorname{cov}([\mathbf{X}_o'\mathbf{\Delta}_o'\,\Sigma_{22}^{-1}\,\mathbf{\Delta}_o\mathbf{X}_o]^{-1}\,\mathbf{X}_o'\mathbf{\Delta}_o'\,\Sigma_{22}^{-1}\,\mathbf{\Delta}_o\mathbf{y}_o,\mathbf{e}) \\ &= \operatorname{cov}([\mathbf{X}_o'\mathbf{\Delta}_o'\,\Sigma_{22}^{-1}\,\mathbf{\Delta}_o\mathbf{X}_o]^{-1}\,\mathbf{X}_o'\mathbf{\Delta}_o'\,\Sigma_{22}^{-1}[\mathbf{u}_o + \mathbf{\Delta}_o\mathbf{e}_o],\mathbf{e}) \\ &= [\mathbf{X}_o'\mathbf{\Delta}_o'\,\Sigma_{22}^{-1}\,\mathbf{\Delta}_o\mathbf{X}_o]^{-1} \\ &\qquad \mathbf{X}_o'\mathbf{\Delta}_o'\,\Sigma_{22}^{-1}\,\mathbf{\Delta}_o[\mathbf{0}_{n\times m}|\mathbf{I}_{n\times n}|\mathbf{0}_{n\times m}]\operatorname{var}(\mathbf{e}). \qquad (3.4)\end{aligned}$$

Note that $[\mathbf{0}_{n\times m} \,|\, \mathbf{I}_{n\times n} \,|\mathbf{0}_{n\times m}]\operatorname{var}(\mathbf{e}) = [\operatorname{cov}(\mathbf{e}_o,\,\mathbf{e}_b) \,|\, \operatorname{var}(\mathbf{e}_o) \,|\, \operatorname{cov}(\mathbf{e}_o,\,\mathbf{e}_f)]$ is the middle n rows of $\operatorname{var}(\mathbf{e})$. Using (3.4) and the aforementioned results, (3.3) can be computed. We can compare the resulting diagonal elements of $\operatorname{var}(\mathbf{v})$ with those of the sum of the last three terms in (3.3), to see if allowing for the error due to estimating the regression parameters is important.

There is an important qualification to make about the results of this section. Since the first term on the right hand side of (3.3), $\mathbf{\Omega}\operatorname{var}(\mathbf{Z} - \hat{\mathbf{z}})\mathbf{\Omega}'$, is the seasonal adjustment variance we would get by ignoring error in estimating the regression parameters, it is tempting to interpret the sum of the last three terms in (3.3) as the contribution to seasonal adjustment variance of error due to estimating regression parameters. Unfortunately, this sum is not itself a variance (it can in fact be written as $\text{var}(\mathbf{K}\hat{\beta}+\mathbf{\Omega e})-\text{var}(\mathbf{\Omega e})$), and so it can actually be negative. When this happens the seasonal adjustment variances that allow for error due to estimating regression parameters are actually lower than those that ignore this error. We were in fact able to achieve such a result by artificially modifying model parameters in the following example with trading-day variables (though, as in the results shown, the effects were quite small). This situation contrasts with comparable results for model-based approaches which express the seasonal adjustment error as the sum of two orthogonal terms: the error when all parameters are known, plus the contribution to error from estimating regression parameters. The seasonal adjustment variance in this case is thus the sum of the variances of these two terms, and so the "regression contribution" is always nonnegative. This result is analogous to $\mathbf{\Omega}\,\text{var}(\mathbf{Z}-\hat{\mathbf{z}})\mathbf{\Omega}' + \mathbf{K}\,\text{var}(\hat{\beta})\mathbf{K}'$ in (3.3). The problem in (3.3) is that the X-11 estimate $\mathbf{\Omega z}$ is not an optimal (MMSE) estimator of the target $\mathbf{\Omega Z}$, hence the error $\mathbf{\Omega}(\mathbf{Z}-\hat{\mathbf{z}})$ is correlated with $\hat{\beta}$ through the sampling error $\mathbf{e}$, leading to the two covariance terms in (3.3). This situation results partly from our choice of target $(\mathbf{X}\beta+\mathbf{\Omega Z})$ and partly from the fact that X-11 cannot be assumed to produce an optimal estimator of anything (note comments related to this in the Introduction).

Example – U.S. 5+ Unit Housing Starts (continued). We use the same example to illustrate the contribution to seasonal adjustment error of adding trading-day variables (Bell and Hillmer 1983), although the corresponding regression coefficients were not statistically significant when estimated with this series. Figure 6a shows the results. In this illustration, the lowest line is the "contribution" to the seasonal adjustment variance from estimating the trading-day effects (but see remarks above). When added to the original estimate of variance (dotted line), we obtain the variance of the seasonally and trading-day adjusted series, allowing for error in estimating the trading-day coefficients (top solid line). We see that, for this example, the increase in variance due to including estimated trading-day effects in the model is slight. Figure 6b gives results for the trend filter. Here the contribution to trend uncertainty due to estimating the trading-day coefficients is certainly negligible.

The contribution to seasonal adjustment variance of adding three additive outlier variables and one level shift variable is illustrated in Figure 6c. These regression variables were identified as potential outlier effects using the Regarima program (produced by the Time Series Staff at the U.S. Census Bureau) with a critical *t*-statistic of 2.5. Regarima uses an outlier detection methodology similar to those discussed in Bell (1983) and Chang, Tiao, and Chen (1988). The contributions of the additive outliers appear as three spikes while that of the level shift is a single smaller hump in the middle of the series. In comparison to the trading-day regression variables, the effect of these outlier variables is mainly local but much stronger. In particular, there is additional uncertainty about seasonal adjustments for observations considered additive outliers.

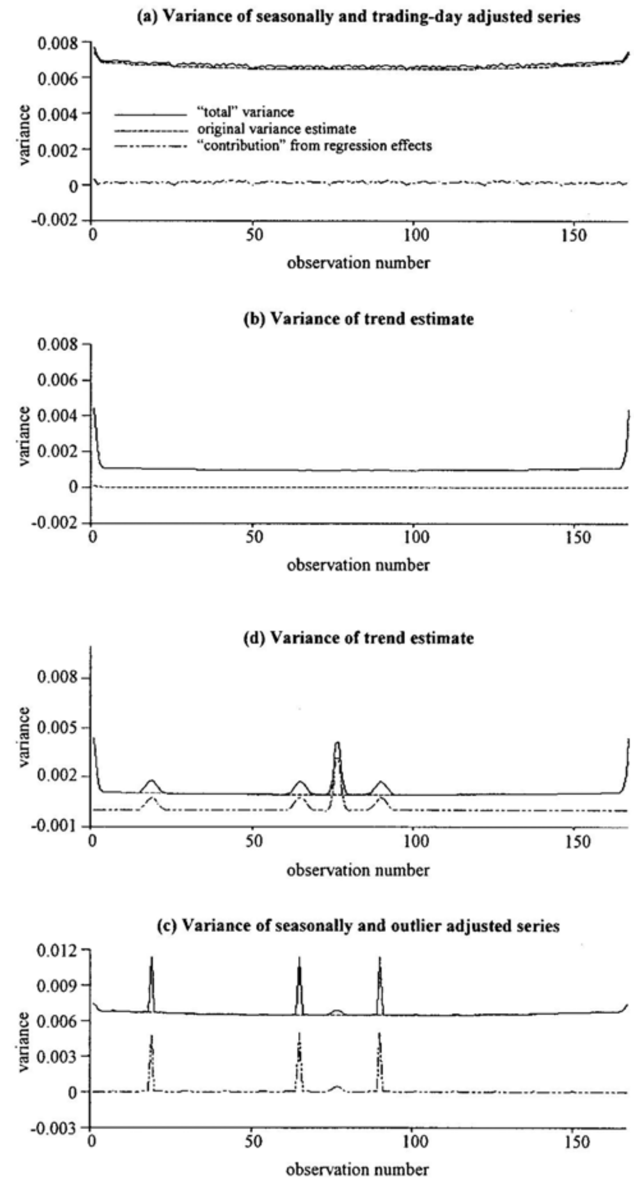

Figure 6 U.S. 5+ Units housing starts: Including the "Contribution" from regression effects in the variance estimates. The top panel shows both the original variances from the first panel of Figure 2 (dotted curve) and the variances allowing for additional uncertainty due to estimating trading-day regression effects (top solid curve). The regression contribution is also shown (bottom solid curve). The second panel shows the corresponding results for the variances of the trend estimates. Note that the regression contribution to the seasonal adjustment variances is small, and to the trend estimate variances it is essentially zero. The third and fourth panels show analogous results when the trading-day regression effects are replaced by three additive outliers and a level shift. Notice that these have important local effects on the seasonal adjustment and trend estimate variances

Results for the trend filter (Figure 6d) differ in that uncertainty is much greater around the observation where a level shift was detected, which approaches the level of uncertainty at the ends of the series. A level shift is considered part of the trend, so an estimated level shift effect would first be subtracted from the series (in $\mathbf{X}\hat{\beta}$), and then added back following application of the X-11 trend filter. (This is analogous to the treatment of regression effects assigned to the nonseasonal or seasonal components in seasonal adjustment as discussed above.) In contrast, since both additive outliers and level shifts are considered part of the nonseasonal component, all four effects were added back as part of the seasonal adjustment when producing results for Figures 6a and 6b.

Actually, these sorts of results for outliers should only be regarded as crude approximations, since they treat the time of occurrence and types of outliers as known, leaving only the magnitude of the effects to be estimated. Ideally, one would like to recognize that the series was searched for significant outliers, but this is much more difficult.

4. Examples

We illustrate our approach using several additional economic time series whose sampling errors follow different models. The models used for these example series are taken from previous work as noted.

4.1 Retail sales of department stores

Department store sales are estimated in the Census Bureau's monthly retail trade survey. Essentially all sales come from department store chains, all of which are included in the survey, hence, there is virtually no sampling error in the estimates. Thus, the variance of the X-11 seasonal adjustment comes only from fore/backcast error and from error in estimating regression effects. (Note that the Wolter-Monsour seasonal adjustment variance would be zero for this series.) The model used for this series (Bell and Wilcox 1993), for the period August 1972 through March 1989 for the logs of the observations, is $(1-B)(1-B^{12})$ $[Y_t - x_t'\beta] = (1-0.53B)(1-0.52B^{12})\, a_t$ with $\sigma_a^2 = 4.32\times10^{-4}$, where x_t includes variables to account for trading-day and Easter holiday effects, and $Y_t = y_t$ is the log of the original series divided by length-of-month factors. In adjusting the series at the Census Bureau, the default X-11 adjustment filter and 13-term Henderson trend filter are used.

Figure 7a shows the standard errors for the seasonally adjusted data over time, with and without the contribution of regression effects. Unlike the 5+ units housing starts series, there are marked increases in the standard errors of seasonally adjusted data at the ends of series, due entirely to fore/backcast error. The contribution to the standard error due to estimating regression effects is also more pronounced for this series. An interesting feature in Figure 7 is the sets of small downward projecting spikes that occur one year apart in triplets. These occur at non-leap year Februaries, for which there is no trading-day effect (the trading-day regression variables are all zero). There is still a small regression contribution to seasonal adjustment error at these time points since the adjustment averages in these contributions from adjacent time points. (Dips at non-leap year Februaries are also visible on close inspection of Figure 6a.) In addition, for some years, the error in estimating the Easter effect produces a noticeable upward projecting spike involving the two months March and April.

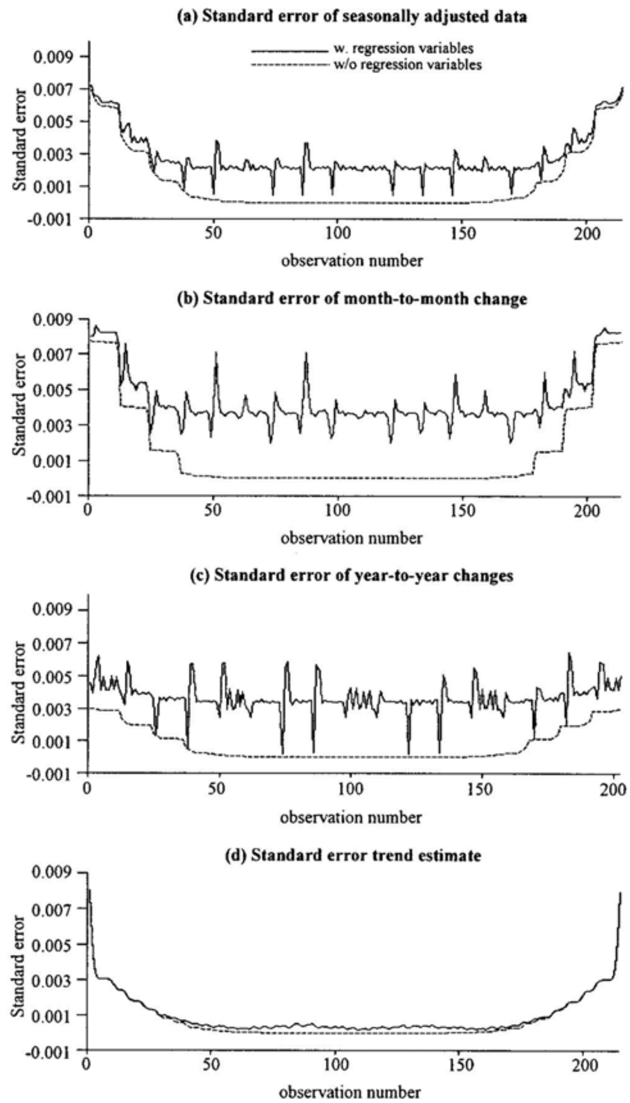

Figure 7 U.S. Department stores, with trading-day and easter effects. This series has no sampling error. The four panels give standard errors with and without the contribution from estimating regression effects. For the seasonally adjusted data and corresponding month-to-month and year-to-year changes (first three panels), the "contribution" from estimating regression effects is substantial and erratic in the middle of the series (where it is the sole contributor) but, at either end, diminishes for reasons explained in the text. The regression contribution to the trend estimate standard errors is small

The regression relative contribution to the seasonal adjustment standard errors diminishes towards the ends of the series. This results from two factors: (1) the magnitude of the regression contribution to $\mathrm{var}(v_t)$ decreases somewhat towards the ends of the series, and, more importantly, (2) $\mathrm{var}(Z_t - \hat{Z}_t)$ increases dramatically towards the ends of the series, diminishing the *relative* contribution to $\mathrm{var}(v_t)$ due to regression (and this is further accentuated when square roots are taken).

The pattern of the standard errors of seasonally adjusted month-to-month changes (Figure 7b) is similar to that for the standard error of the seasonally adjusted data (Figure 7a). The regression contribution is slightly larger than it is for the seasonally adjusted data. Standard errors of year-to-year changes (Figure 7c) follow similar patterns but the regression contribution is considerably larger than it is for the month-to-month changes, and it remains important at the ends of the series.

A similar set of calculations was performed using the default X-11 trend filter, and results for the standard errors of the trend estimates, with and without the regression contribution, are depicted in Figure 7d. The patterns over time of these standard errors are similar to the corresponding figures for the 5+ units housing starts series, but the standard errors are much smaller due to the absence of sampling error. The regression contribution is small.

The standard errors for all plots in Figure 7 are small – none exceed 0.8 percent. For this series, the regression contribution is small and probably ignorable near the very ends of the series, for all but the year-to-year changes. However, in the middle of the series, the sole contributor to standard errors is that due to the regression effects.

4.2 Teenage unemployment

The Bureau of Labor Statistics (BLS) publishes the monthly time series of number of U.S. unemployed teenagers estimated from the Current Population Survey (CPS). Data from January 1972 to December 1983 (n = 144) were used by Bell and Hillmer (submitted) to develop a model for this series. The sampling error variance h_t^2 changes over time, so is nonstationary. The sampling error model they developed is

$$e_t = h_t \tilde{e}_t \text{ where } (1 - 0.6B)\tilde{e}_t = (1 - 0.3B)b_t, \quad (4.1)$$

with $\sigma_b^2 = 0.87671$ so that $\mathrm{var}(\tilde{e}_t) = 1$. CPS sampling error variances can be approximated by generalized variance functions (Wolter 1985, Chapter 5; Hanson 1968). The generalized variance function Bell and Hillmer used for the teenage unemployment series is

$$h_t^2 = 1.971 y_t - (1.53 \times 10^{-5}) y_t^2, \quad (4.2)$$

where y_t is the estimate of the number in thousands of unemployed teenagers at time t. The estimated model for the signal component Y_t is

$$(1 - B)(1 - B^{12}) Y_t = (1 - 0.27B)(1 - 0.68B^{12})\, a_t, \quad (4.3)$$

with $\sigma_a^2 = 4{,}294$. There are no regression effects in the model and the series is not transformed. BLS uses the default X-11 seasonal adjustment filter (so $m = 84$).

In applying the methods of this paper to this example, problems arise from the fact that the (estimated) sampling error variance h_t^2 depends on the estimate y_t through the generalized variance function (4.2). In the backcast and forecast periods y_t is unknown. To obtain h_t^2 in these periods we forecast and backcast y_t using a simple ARIMA(0 1 1)(0 1 1)$_{12}$ model for y_t (not for Y_t, as in (4.3)). The resulting 84 forecasts and backcasts were then used in (4.2) to produce h_t^2 in the forecast and backcast periods. More refined treatments are possible, such as using the component model given by (4.1) and (4.3) to forecast y_t.

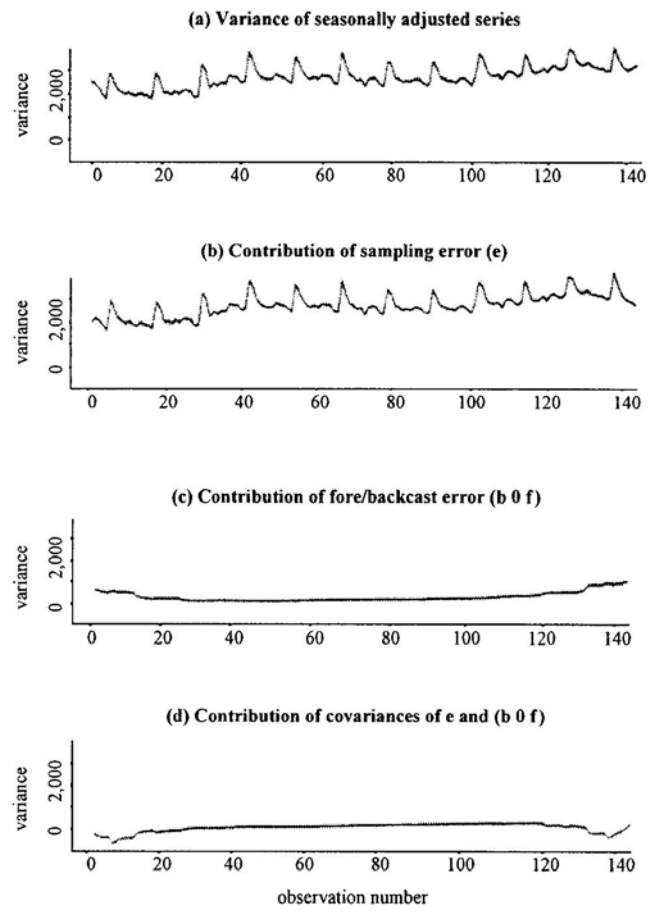

Figure 8 Teenage unemployment, with default X-11 options. The panel descriptions are as for Figure 2. The seasonal pattern of the sampling error variance contribution (second panel) results from its dependence on the level of the series through a generalized variance function (see text)

The seasonal adjustment variance for this series (Figure 8a) is dominated at most times t by the sampling error contribution (Figure 8b). This is because, while the contribution of $\mathrm{var}(\mathbf{b}, \mathbf{0}, \mathbf{f})$ is substantial for this series (Figure 8c), it tends to be offset by the contribution of $\mathrm{cov}[(\mathbf{b}, \mathbf{0}, \mathbf{f}), \mathbf{e}] + \mathrm{cov}[(\mathbf{b}, \mathbf{0}, \mathbf{f}), \mathbf{e}]'$ (Figure 8d), except at the first and last few time points. The patterns of variances of seasonally adjusted month-to-month changes and year-to-year changes (not shown) are similar to that of Figure 8a. The variances of the month-to-month changes are slightly larger than those of the adjusted series, those of the year-to-year changes are larger still.

4.3 Retail sales of drinking places

Retail sales of drinking places are estimated in the Census Bureau's monthly retail trade survey. In this survey, (noncertainty) sample cases are independently redrawn approximately every 5 years, so the covariance matrix of the sampling errors is block diagonal. Bell and Hillmer (1990) developed the following model for the sampling error of the logged series within a given sample:

$$(1 - 0.75B - 0.66B^3 + 0.50B^4)(1 - 0.71B^{12})\, e_t = (1 + 0.13B)\, b_t, \quad (4.4)$$

with $\sigma_b^2 = 9.301 \times 10^{-5}$. For time points t and j in different samples, $\mathrm{cov}(e_t, e_j) = 0$. Bell and Hillmer developed a model for the signal component of the logged series using unbenchmarked estimates from September 1977 to December 1986. We shall instead use the following model fit by Bell and Wilcox (1993) using additional data through October 1989:

$$(1 - B)(1 - B^{12})[Y_t - X_t'\beta] = (1 - 0.23B)(1 - 0.88B^{12})\, a_t,$$

where X_t contains trading-day regression variables, and $\sigma_a^2 = 4.16 \times 10^{-4}$.

In seasonally adjusting this series, the default X-11 filters are used. The contribution of error due to estimating regression parameters is small for this series, and so is not included in the results to follow. Since the contribution of sampling error overwhelms the contributions from fore/backcast error and $\mathrm{cov}[(\mathbf{b}, \mathbf{0}, \mathbf{f}), \mathbf{e}] + \mathrm{cov}[(\mathbf{b}, \mathbf{0}, \mathbf{f}), \mathbf{e}]'$, we also do not illustrate these separate variance contributions. Figure 9a gives the standard error of the seasonally adjusted data (shown over 232 observations to better illustrate the pattern, with vertical lines indicating sample redraws) and Figure 9b the standard error of seasonally adjusted month-to-month changes.

Note the strong pattern in Figure 9a, b due to the redrawing of the sample every five years. In particular, this produces a large spike in the standard error of seasonally adjusted month-to-month changes (Figure 9b) when the sample is redrawn. Similar jumps in standard deviations of year-to-year changes occur for the first year of a new sample. We also found similar patterns for other series from the retail trade survey using models from Bell and Wilcox (1993).

The preceding discussion and results ignored certain aspects of how estimation for the retail trade survey is actually carried out. In fact, to avoid large increases in variances of change estimates around the sample redraw, such as those reflected in Figure 9b, simple modifications are made to estimates in a newly introduced sample to make their level consistent with that from the old sample. The simplest version of the modification is as follows. Let $z_{(\mathrm{old})t} (= \exp(y_{(\mathrm{old})t}))$ denote estimates from the old sample, and $z_{(\mathrm{new})t}$ unmodified estimates from the new sample. Assume that the old sample provides estimates for $t \le \tau$, and that the new sample is to provide estimates for $t > \tau$. To provide overlap data for the modification, the new sample is begun one month early, so that both $z_{(\mathrm{old})\tau}$ and $z_{(\mathrm{new})\tau}$ are available. The modified new sample estimates are defined as $z'_{(\mathrm{new})t} = z_{(\mathrm{new})t}(z_{(\mathrm{old})\tau}/z_{(\mathrm{new})\tau})$ for $t \ge \tau$. This modification is carried out each time a new sample is introduced. In terms of the corresponding logged estimates y_t, the modification is $y'_{(\mathrm{new})t} = y_{(\mathrm{new})t} + (y_{(\mathrm{old})\tau} - y_{(\mathrm{new})\tau})$. Since the modification to y_t is linear, it is easy to account for its effects on the seasonal adjustment variance calculations here. The month-to-month change at time $\tau + 1$ before the modification (and without seasonal adjustment) is $y_{(\mathrm{new})\tau+1} - y_{(\mathrm{old})\tau}$. Note that this change has a large variance since $y_{(\mathrm{new})\tau+1}$ and $y_{(\mathrm{old})\tau}$ come from different, independent samples. After modification, this change is $y_{(\mathrm{new})\tau+1} - y_{(\mathrm{new})\tau}$, which has a much lower variance due to strong positive correlation between $y_{(\mathrm{new})\tau+1}$ and $y_{(\mathrm{new})\tau}$ (arising from the the sampling error model (4.4)). Unadjusted month-to-month change estimates for time points other than $\tau + 1$ are unaltered by the modification.

Figure 9d shows that modifying new sample estimates eliminates the large increases in the standard deviation of seasonally adjusted month-to-month changes at the transitions to new samples. Similar effects were seen for year-to-year changes over a one year period. The price paid for this improvement is a steadily increasing error in the level estimates (Figure 9c) following introduction of new samples. This occurs because the modification introduces a transient error into the level estimates that persists throughout the new sample. Thus, the modification trades off worse accuracy of level estimates for improvements in change estimates. (Figure 9c shows no increase for the first five years because we assume the estimates there are not modified to agree with those from a previous sample.) Moreover, the strong patterns in Figure 9a occur because the sampling errors from unmodified estimates in adjacent samples are uncorrelated. On the other hand, sampling errors in the modified estimates are fairly strongly correlated between adjacent samples. The effect of this, after applying the seasonal adjustment filter, is a much different pattern (almost no pattern) in the first five years of Figure 9c, and slight oscillations around the linear increase thereafter.

The standard errors for the X-11 trend estimates and changes (not shown) look like smoothed versions of those shown in Figure 9.

In practice, final estimates from the retail trade survey are even more complicated than what was just described and illustrated. First, more than one month of overlapping

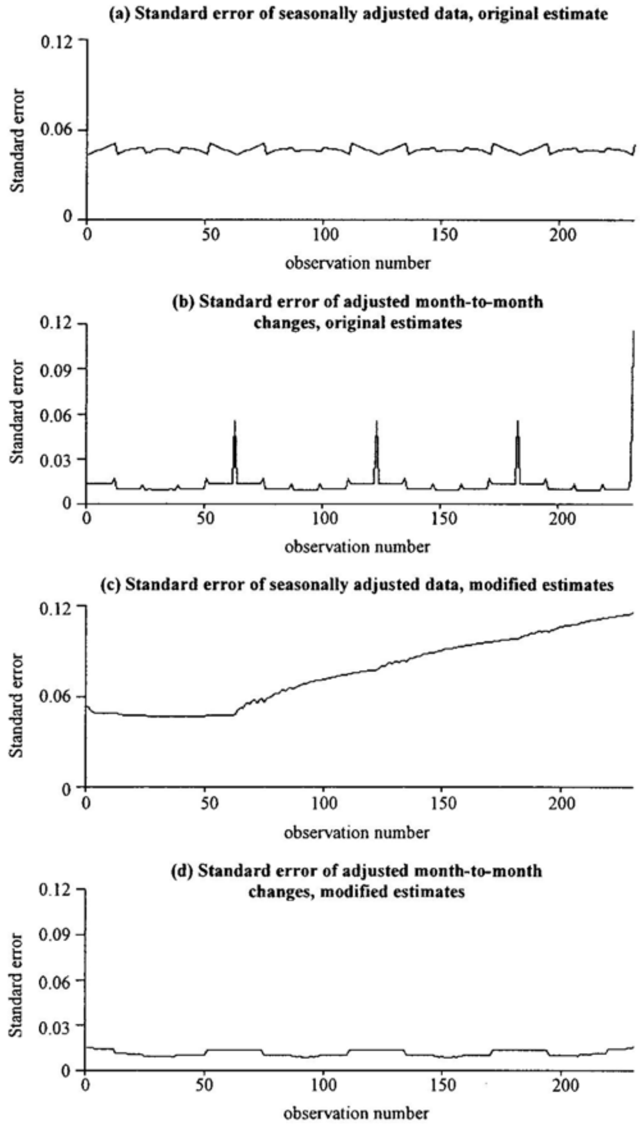

Figure 9 Retail sales of drinking places: Samples redrawn every five years. The top panel shows the standard error of the seasonally adjusted data and the second panel the standard error of the corresponding month-to-month changes. The strong pattern results from independently drawing a new sample every five years (at the dotted vertical lines). For month-to-month changes, this produces large increases in standard errors at the time of the sample redraw. To eliminate this problem, a new sample is drawn to overlap with the previous sample for one or more months and the new sample's estimates are modified using data from the overlap to make them consistent in level with estimates from the previous sample (see text). This eliminates the increases in standard errors of change estimates when the sample is redrawn (fourth panel), but introduces a transient error into the modified level estimates, whose effects accumulate over time (third panel)

data are collected and may be used to modify level estimates when a new sample is introduced. More importantly, monthly estimates are benchmarked to agree with annual totals obtained from the more accurate annual retail trade survey or five year economic census. Benchmarking should thus alleviate the problem of level variances increasing over time seen in Figure 9c. However, since benchmarking imposes linear sum constraints on the original (unlogged) estimates, its effects on seasonal adjustment variances are difficult to investigate under the approach developed here, and we have not done so. (We have used a model for unbenchmarked data to avoid this problem.) Durbin and Quenneville (1995) develop a model-based approach to benchmarking that accounts for the nonlinearities that such benchmark constraints impose on logged data.

5. Conclusions

This paper presented an approach to the long-standing problem of obtaining variances for X-11 seasonal adjustments. Our goal was the development and application of an approach to obtain variances accounting for two sources of error. The first error source is sampling error (e_t), which arises because we do not observe the true series, Y_t, but instead observe estimates $y_t = Y_t + e_t$ from a repeated survey. The second error source results from the need to extend the observed series with forecasts and backcasts to apply the symmetric X-11 filters. This second error source leads to seasonal adjustment revisions. To account for these two sources of error, we defined the seasonal adjustment variance as the variance of the error in using the X-11 adjustment to estimate a specific target. This target, $\omega(B)Y_t$, is what would result from applying the symmetric, linear X-11 filter, $\omega(B)$, to the true series if its values were available far enough into the future and past for the symmetric filter to be used. (The application to additive X-11 with fore/backcast extension is immediate, and log-additive X-11 is taken as an approximation to multiplicative X-11.)

Our approach was also applied to produce variances of X-11 trend estimates, and to produce variances of month-to-month and year-to-year changes in both the seasonally adjusted data and trend estimates. A further extension was made to allow for error in estimating regression parameters (*e.g.*, to model calendar effects), though this was more involved and had some limitations.

The variances we obtain ignore uncertainty due to time series variation in the seasonal and nonseasonal components. We argued in section 2 that this may be appropriate for typical users of X-11 seasonally adjusted data. If one desires to account for this time series variation, however, we suggest that consideration be given to model-based approaches to seasonal adjustment, since time series models provide a means to explicitly account for variation in all the components. Alternatively, Pfeffermann (1994) developed an approach to X-11 seasonal adjustments that attempts to account for irregular variation and sampling error.

Our approach builds on the first approach suggested by Wolter and Monsour (1981), by accounting for the contribution of forecast and backcast error that was ignored by them. An alternative view of the difference between our approach and theirs is that we define a consistent seasonal adjustment target, whereas, in using X-11's asymmetric filters, Wolter and Monsour implicitly used targets that change over time. Because of this, our approach avoids the unrealistic feature of seasonal adjustment variances that decrease towards the ends of the series, which can be seen in results of Wolter and Monsour, and also of Pfeffermann.

In the empirical results presented, the contribution of sampling error often dominated the seasonal adjustment variances. This is partly because sampling error was often large relative to fore/backcast error, and partly because the contribution of fore/backcast error tended to be offset by the contribution of the covariance of fore/backcast error with the sampling error. On the other hand, empirical results for trend estimate variances showed large increases at the ends of series due to the effects of fore/backcast error. Since the largest contribution of fore/backcast error occurs at the ends of the series, and variances for the most recent seasonal adjustments and trend estimates are of the most interest, one should not ignore the contribution of fore/backcast error.

The relative contribution to our variances of error in estimating trading-day or holiday regression coefficients tended to be small, unless the series had no sampling error. Error due to estimating additive outlier and level shift effects was substantial around the time point of the outlier. The effects of AOs were large on seasonal adjustment variances; the effects of LSs were large on trend estimate variances.

Nonstationarities in the sampling errors produced interesting patterns in the seasonal adjustment and trend estimate variances. Two types of sampling error nonstationarities were examined. Seasonal patterns in sampling error variances produced corresponding seasonal patterns in seasonal adjustment variances. Independent redrawings of the sample, which yield sampling errors correlated within but not across samples, produced erratic patterns in seasonal adjustment and trend estimate variances over time within a sample. These patterns approximately repeat across different samples if the samples remain in force for approximately equal time spans.

Computations for the examples shown (given the fitted models, which were obtained from the references cited) were done by programming the expressions of Sections 2 and 3 in the S+ programming statistical language. The resulting computer code is available on request.

Acknowledgements

We thank Nash Monsour for providing the FORTRAN program to calculate default X-11 asymmetric filter weights. In addition, he patiently explained how estimates from the monthly retail trade surveys are modified when new samples are drawn. This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributed to the authors and do not necessarily reflect those of the Census Bureau.

Appendix A

Several expressions to be calculated in this paper are of the general form

$$\mathbf{A}\Sigma^{-1}\mathbf{B} \tag{A.1}$$

where Σ is a positive definite matrix, and $\mathbf{A}$ and $\mathbf{B}$ are conformable to Σ. Let $\Sigma = \mathbf{LL}'$ be the Cholesky decomposition of Σ. Then $\mathbf{A}\Sigma^{-1}\mathbf{B} = \mathbf{A}(\mathbf{L}^{-1})'\mathbf{L}^{-1}\mathbf{B}$ and (A.1) can be computed as follows:

(1) Solve $\mathbf{LQ}_1 = \mathbf{B}$ for $\mathbf{Q}_1$
(2) Solve $\mathbf{LQ}_2 = \mathbf{A}'$ for $\mathbf{Q}_2$
(3) Compute $\mathbf{A}\Sigma^{-1}\mathbf{B} = \mathbf{Q}_2\mathbf{Q}_1$.

(1) and (2) can be solved efficiently since $\mathbf{L}$ is lower triangular.

Appendix B

Two steps are required to obtain $\mathbf{GX}_o$, used in section 3.3. The first step produces "forecast" and "backcast" extension of the differenced regression variables. The second step uses these results and the difference equation to produce forecast and backcast extension of the original (undifferenced) regression variables.

Let $\mathbf{R}_o = \mathbf{\Delta}_o\mathbf{X}_o$, where $\mathbf{\Delta}_o$ is that part of the matrix $\mathbf{\Delta}$ which differences the observed series $\mathbf{y}_o$. Analogous to the computation of $\hat{\mathbf{w}}_f$ and $\hat{\mathbf{w}}_b$ in section 2.2, forecast extensions of the differenced regression variables are calculated as $\mathbf{R}_f = \Sigma_{32}\Sigma_{22}^{-1}\mathbf{R}_o$ and backcast extensions as $\mathbf{R}_b = \Sigma_{12}\Sigma_{22}^{-1}\mathbf{R}_o$. $\mathbf{R}_f$ and $\mathbf{R}_b$ are of the form (A.1) and can be computed by the technique given above.

For the second step, let x_t denote any one of the regression variables in $\mathbf{X}$. Let the required forecast extensions be denoted $\hat{x}_{n+l}$ for $l = 1, 2, ..., m$. Let the differencing operator in the model be $\delta(B) = 1 - \delta_1 B - \ldots - \delta_d B^d$, and let $\hat{r}_{n+l}$ be the forecast extension of $\delta(B)x_t = r_t$ at time $n + l$ ($\hat{r}_{n+l}$ is an element of $\mathbf{R}_f$). The $\hat{x}_{n+l}$ are calculated iteratively as

$$\hat{x}_{n+l} = \delta_1\hat{x}_{n+l-1} + \ldots + \delta_d\hat{x}_{n+l-d} + \hat{r}_{n+l}, \text{ for } l = 1, ..., m,$$

where $\hat{x}_{n+j} = x_{n+j}$ if $j \le 0$.

The required backcast extensions of x_t are denoted $\hat{x}_{1+l}$ for $l = 1, ..., m$. These are also obtained recursively from the difference equation $\delta(B)\hat{x}_t = \hat{r}_t$ by solving for $\hat{x}_{1+l}$ in the expression

$$\hat{x}_{d+1-l} = \delta_1\hat{x}_{d-l} + \ldots + \delta_d\hat{x}_{1-l} + \hat{r}_{d+1+l}$$

and substituting previously computed backcasts as needed. Thus,

$$\hat{x}_{1+l} = \delta_d^{-1}(\hat{x}_{d+1-l} - \delta_1\hat{x}_{d-l} - \ldots - \delta_{d-1}\hat{x}_{2-l} - \hat{r}_{d+1-l}),$$

$$\text{for } l = 1, ..., m,$$

where $\hat{x}_{1-j} = x_{1-j}$ for $j \le 0$.

References

Bell, W.R. (1983). A computer program for detecting outliers in time series. *Proceedings of the Business and Economic Statistics Section*, American Statistical Association, 634-639.

Bell, W.R. (1984a). Signal extraction for nonstationary time series. *Annals of Statistics*, 12, 646-664.

Bell, W.R. (1984b). Seasonal Decomposition of Deterministic Effects, Research Report Number 84/01. Statistical Research Division, Bureau of the Census.

Bell, W.R., and Hillmer, S.C. (1983). Modeling time series with calendar variation. *Journal of the American Statistical Association*, 78, 526-534.

Bell, W.R., and Hillmer, S.C. (1984). Issues involved with the seasonal adjustment of economic time series (with discussion). *Journal of Business and Economic Statistics*, 2, 291-320.

Bell, W.R., and Hillmer, S.C. (1988). A Matrix Approach to Signal Extraction and Likelihood Evaluation for ARIMA Component Time Series Models. Research Report Number 88/22, Statistical Research Division, Bureau of the Census.

Bell, W.R., and Hillmer, S.C. (1990). The time series approach to estimation for repeated surveys. *Survey Methodology*, 16, 195-215.

Bell, W.R., and Hillmer, S.C. (submitted). Applying time series models in survey estimation.

Bell, W.R., and Monsell, B.C. (1992). X-11 Symmetric Linear Filters and Their Transfer Functions. Research Report 92/15, Statistical Research Division, Bureau of the Census.

Bell, W.R., and Otto, M.C. (submitted). Bayesian Assessment of Uncertainty in Seasonal Adjustment with Sampling Error Present.

Bell, W.R., and Wilcox, D.W. (1993). The effect of sampling error on the time series behavior of consumption data. *Journal of Econometrics*, 55, 235-265.

Binder, D.A., and Dick, J.P. (1989). Modelling and estimation for repeated surveys. *Survey Methodology*, 15, 29-45.

Binder, D.A., and Dick, J.P. (1990). A method for the analysis of seasonal ARIMA models. *Survey Methodology*, 16, 239-253.

Bobbitt, L., and Otto, M.C. (1990). Effects of forecasts on the revisions of seasonally adjusted values using the X-11 seasonal adjustment procedure. *Proceedings of the Business and Economics Statistics Section*, American Statistical Association, 449-454.

Box, G.E.P., and Jenkins, G.M. (1976). *Time Series Analysis*: *Forecasting and Control*. San Francisco: Holden Day.

Breidt, F.J. (1992). Variance estimation in the frequency domain for seasonally adjusted time series. *Proceedings of the Business and Economic Statistics Section*, American Statistical Association, 337-342.

Burridge, P., and Wallis, K.F. (1984). Unobserved components models for seasonal adjustment filters. *Journal of Business and Economic Statistics*, 2, 350-359.

Burridge, P., and Wallis, K.F. (1985). Calculating the variance of seasonally adjusted series. *Journal of the American Statistical Association*, 80, 541-552.

Chang, I., Tiao, G.C. and Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics*, 30, 193-204.

Cleveland, W.S., and Devlin, S.J. (1982). Calendar effects in monthly time series: Modeling and adjustment. *Journal of the American Statistical Association*, 77, 520-528.

Dagum, E.B. (1975). Seasonal factor forecasts from ARIMA models. *Proceedings of the 40th Session of the International Statistical Institute*, Warsaw, Poland, 206-219.

Durbin, J., and Quenneville, B. (1995). Benchmarking monthly time series with structural time series models. *Proceedings of the Survey Methods Section*, *23rd Annual Meeting of the Statistical Society of Canada*, Montreal, Canada, 13-18.

Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C. and Chen, B. (1998). New capabilities and methods of the X-12-ARIMA seasonal adjustment program. *Journal of Business and Economic Statistics*, 16, 127-152.

Geweke, J. (1978). Revision of Seasonally Adjusted Time Series. SSRI Report No. 7822, Department of Economics, University of Wisconsin.

Hanson, R.H. (1968). *The Current Population Survey*: *Design and Methodology*. Technical Paper 40, U.S. Census Bureau, Washington, D.C., Government Printing Office.

Hausman, J.A., and Watson, M.W. (1985). Errors in variables and seasonal adjustment procedures. *Journal of the American Statistical Association*, 80, 531-540.

Hillmer, S.C., Bell, W.R. and Tiao, G.C. (1983). Modeling considerations in the seasonal adjustment of economic time series. *Applied Time Series Analysis of Economic Data*, (Ed. A. Zellner), U.S. Department of Commerce, Bureau of the Census, 74-100.

McKenzie, S.K. (1984). Concurrent seasonal adjustment with Census X-11. *Journal of Business and Economic Statistics*, 2, 235-249.

McLeod, I. (1975). Derivation of the theoretical autocovariance function of autoregressive-moving average time series. *Applied Statistics*, 24, 255-256.

McLeod, I. (1977). Correction to derivation of the theoretical autocovariance function of autoregressive-moving average time series. *Applied Statistics*, 26, 194.

Pfeffermann, D. (1994). A general method for estimating the variances of X-11 seasonally adjusted estimators. *Journal of Time Series Analysis*, 15, 85-116.

Pfeffermann, D., Morry, M. and Wong, P. (1993). Estimation of the Variances of X-11-ARIMA Seasonally Adjusted Estimators for a Multiplicative Decomposition and Heteroscedastic Variances. Working Paper METH-93-005, Time Series Research and Analysis Division, Statistics Canada, Ottawa, Canada.

Pierce, D.A. (1980). Data revisions with moving average seasonal adjustment procedures. *Journal of Econometrics*, 14, 95-114.

President's Committee to Appraise Employment and Unemployment Statistics (1962). Measuring Employment and Unemployment, Washington, D.C.: U.S. Government Printing Office.

Shiskin, J., Young, A.H. and Musgrave, J.C. (1967). The X-11 Variant of the Census Method II Seasonal Adjustment Program. Technical Paper No. 15, U.S. Department of Commerce, Bureau of Economic Analysis.

Tiller, R.B. (1992). Time series modeling of sample survey data from the U.S. current population survey. *Journal of Official Statistics*, 8, 149-166.

Wallis, K.F. (1982). Seasonal adjustment and revision of current data: linear filters for the X-11 method. *Journal of the Royal Statistical Society*, Series A, 145, 74-85.

Wilson, G.T. (1979). Some efficient computational procedures for high order ARIMA models. *Journal of Statistical Computation and Simulation*, 8, 301-309.

Wolter, K.M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.

Wolter, K.M., and Monsour, N.J. (1981). On the problem of variance estimation for a deseasonalized Series. In *Current Topics in Survey Sampling*, (Eds. D. Krewski, R. Platek, and J.N.K. Rao), New York: Academic Press, 367-403.

Young, A. (1965). Estimating Trading-day Variation in Monthly Economic Time Series. Technical Paper 12, Bureau of the Census.

Young, A.H. (1968). Linear approximations to the Census and BLS seasonal adjustment methods. *Journal of the American Statistical Association*, 63, 445-471.