On X-11 Seasonal Adjustment and Estimation of its MSE October 2009

Michail Sverchkov[1], Danny Pfeffermann[2], Stuart Scott[1]

[1] Bureau of Labor Statistics, 2 Massachusetts Avenue NE, Suite 1950, Washington, DC 20212.
[2] Department of Statistics, Hebrew University of Jerusalem, Jerusalem, 91905, Israel, and Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, S017 1BJ, UK.

Abstract

Most official seasonal adjustments are based on the X-11 method and its extensions. An important problem with the use of this method is how to estimate the mean square error (MSE) of the estimators of the seasonal effects and other components. Wolter and Monsour (1981) assumed that the estimators are unbiased and proposed an approach for variance estimation that uses the linear approximation to X-11 and accounts for the variability of the sampling errors. Pfeffermann (1994) and Bell and Kramer (1999) extend this approach (see below).

In this paper we show that the seasonal and trend components can be defined in such a way that the X-11 estimators of these components are almost unbiased at the center of the series and consequently, Pfeffermann (1994) method produces unbiased estimators for the MSE of the X-11 estimators at the center of the series but not at the two ends. We propose, therefore, bias corrections for the MSE estimates at the ends of the series. Similar bias corrections are proposed for Bell and Kramer (1999) method.

Key Words: decomposition model, bias correction, mean square error

1. Outline

We define the seasonal and trend components under which the X-11 (X-12ARIMA, X-13) estimators of the trend and the seasonal components are almost unbiased in the central part of the series.

The mean square error (MSE) of the X-11 estimators are then defined with respect to the estimation of these components over all possible realizations of the sampling errors and the irregular terms.

We investigate the behavior of the X-11 estimators of the newly defined trend and seasonal components at the two ends of the series where they are biased, and propose bias correction procedures (parametric and non-parametric).

We investigate the Bell and Kramer (1999) decomposition and propose bias correction procedures for their estimators with respect to this decomposition.

The results are illustrated by a small simulation study based on the series “Education and Health Services employment” (EDHS), obtained as part of the Current Employment Statistics program managed by the Bureau of Labor Statistics (BLS).

2. Bias, Variance and MSE of X-11 estimators and their estimation

We begin with the usual notion that an economic time series can be decomposed into a trend or trend-cycle component T_t, a seasonal component S_t, and an irregular term, I_t; $Y_t = T_t + S_t + I_t$. Here we consider for simplicity the additive decomposition but the results can be generalized to the multiplicative decomposition, $Y_t = T_t \times S_t \times I_t$, using similar considerations as in Pfeffermann *et al*. (1995). Typically, the data are obtained from a sample survey, such that the observed series, y_t, can be expressed as the sum of the population series, Y_t, plus a sampling error, ε_t. Define $e_t = I_t + \varepsilon_t$ to be the combined error. Without loss of generality we assume that the series started at time $-\infty < t_{start} < 1$ but y_t is observed only for $t = 1,...,N$, such that

$$y_t = Y_t + \varepsilon_t = T_t + S_t + e_t, \quad t = \underbrace{t_{start},...,0,}_{unobserved}\ \underbrace{1,...,N}_{y_t-observed}\ \underbrace{,N+1,...,\infty}_{unobserved}. \qquad (1)$$

Assumptions: We assume, in addition, that $S_{t_{start}}$ and $T_{t_{start}}$ are fixed (although unknown), $E(T_t + S_t)^2 < \infty$ if $t_{start} < t < \infty$; e_t and $(\mathbf{S},\mathbf{T})$ are independent where $(\mathbf{S},\mathbf{T}) \overset{def}{=} \{(S_t, T_t),\, t = t_{start},...,\infty\}$; $E(e_t) = 0$ and $Var(e_t) < \infty$.

The X-11 program applies a sequence of moving averages or linear filters to the observed series. Thus, the X-11 estimators of the trend and the seasonal components can be approximated as,

$$\hat{S}_t = \sum_{k=-N+t}^{t-1} w_{kt}^S y_{t-k}, \quad \hat{T}_t = \sum_{k=-N+t}^{t-1} w_{kt}^T y_{t-k}, \qquad (2)$$

where the filters w_{kt}^S and w_{kt}^T are defined by the X-11 program options for the given time interval $t = 1,...,N$. Moreover, at the central part of the series the filters are symmetric and time-invariant, $w_{kt}^S = w_k^S$, $w_{-k}^S = w_k^S$, for $a_S \le t \le N - a_S$, $w_{kt}^T = w_k^T$, $w_{-k}^T = w_k^T$, for $a_T \le t \le N - a_T$, where a_S, a_T are likewise defined by the X-11 program options; $2a_S + 1$ and $2a_T + 1$ correspond to the length of the filters so that $w_{kt}^T = w_k^T = 0$ if $k \notin [-a_T, a_T]$, $w_{kt}^S = w_k^S = 0$ if $k \notin [-a_S, a_S]$. Whatever t, $w_{kt}^T = w_{kt}^S = 0$ if $t - k \notin [1,...,N]$. To simplify summation indexes, we denote for a given series Z, $\sum_k w_{kt}^C Z_{t-k} = \sum_{k: w_{kt}^C \ne 0} w_{kt}^C Z_{t-k}$ and $\sum_k w_k^C Z_{t-k} = \sum_{k: w_k^C \ne 0} w_{kt}^C Z_{t-k}$, $C = S$ or T.

Remark 1. X-11 and its extensions, like X-12ARIMA and X-13 include also "non-linear" operations such as the identification and estimation of ARIMA models and the identification and gradual replacement of extreme observations. We assume that the time series under consideration is already corrected for outliers. The effects of the identification and non-linear estimation of ARIMA models are generally minor, see, e.g., Pfeffermann *et al*. (1995) and Pfeffermann *et al*. (2000).

Assuming that $t_{start} < \min(-a_S, -a_T)$, define, $S_t^{x11} = \sum_k w_k^S (T_{t-k} + S_{t-k})$, $T_t^{x11} = \sum_k w_k^T (T_{t-k} + S_{t-k})$, such that S_t^{x11} and T_t^{x11} are the outputs when applying the symmetric filters to the signal of the infinite series at time point *t*, $t = 1,...,N$. Note that (1) and the assumptions imply that $S_t^{x11} = E[\hat{S}_t \,|\, \mathbf{S},\mathbf{T}]$ for $a_S \le t \le N - a_S$, $T_t^{x11} = E[\hat{T}_t \,|\, \mathbf{S},\mathbf{T}]$ for $a_T \le t \le N - a_T$ which in turn imply the following result.

Result 1. Let $e_t^{x11} = y_t - T_t^{x11} - S_t^{x11}$. Then, X-11 decomposes the observed series into the 'X-11-trend' T_t^{x11}, the 'X-11-seasonal component' S_t^{x11}, and the 'X-11 error', e_t^{x11};

$$y_t = T_t^{x11} + S_t^{x11} + e_t^{x11}. \tag{3}$$

At the center part of the series, $\max(a_S, a_T) \le t \le N - \max(a_S, a_T)$, the X-11 estimators of the trend and the seasonal components are almost unbiased with respect to the decomposition (3).

Remark 2. The decomposition defined by (3) into a seasonal component, a trend component and an error term is clearly not unique; see, for example, the discussion in Hilmer and Tiao (1982). Bell and Kramer (1999) use a similar decomposition: they define the "target" of the seasonal adjustment as the adjusted series that would be obtained if there is no sampling error and there are sufficient data before and after the time points of interest for the application of the symmetric filter (Bell and Kramer 1999, page 15). Thus, the Bell and Cramer seasonal and trend components are defined as, $S_t^{B-K} = \sum_k w_k^S (T_{t-k} + S_{t-k} + I_{t-k})$, $T_t^{B-K} = \sum_k w_k^T (T_{t-k} + S_{t-k} + I_{t-k})$. The difference between (3) and the Bell and Kramer (1999) decomposition is therefore that the latter decomposition considers the irregular term as a part of the signal. As a result, the MSE of the X-11 estimators of the components defined by the decomposition (3) is generally higher than the MSE of the X-11 estimators of the components defined by the Bell and Kramer decomposition.

The bias, variance and MSE of the X-11 estimators with respect to decomposition (3), conditional on the true components **S, T** are obtained as follows:

$$Bias[\hat{S}_t \mid \mathbf{S,T}] = E[\hat{S}_t - S_t^{x11} \mid \mathbf{S,T}] = \sum_k (w_{kt}^S - w_k^S)(T_{t-k} + S_{t-k}) . \tag{4}$$

$$Var[\hat{S}_t \mid \mathbf{S,T}] = E\{[\sum_k w_{kt}^S y_{t-k} - E(\sum_k w_{kt}^S y_{t-k} \mid \mathbf{S,T})]^2 \mid \mathbf{S,T}\} =$$

$$E\{[\sum_k w_{kt}^S (y_{t-k} - S_{t-k} - T_{t-k})]^2 \mid \mathbf{S,T}\} \overset{By(1)}{=} E[(\sum_k w_{kt}^S e_{t-k})^2 \mid \mathbf{S,T}] . \tag{5}$$

$$MSE[\hat{S}_t] = E[(\hat{S}_t - S_t^{x11})^2 \mid \mathbf{S,T}] = Var[\hat{S}_t \mid \mathbf{S,T}] + Bias^2[\hat{S}_t \mid \mathbf{S,T}] . \tag{6}$$

Similar expressions are obtained for the trend estimator.

By (5), the variance of the X-11 estimator of the seasonal component is a linear combination of the covariances, $Cov(e_t, e_k \mid \mathbf{T,S})$, $t,k = 1,...,N$. Following Pfeffermann (1994), let $R_t = y_t - \hat{S}_t - \hat{T}_t = \sum_k a_{kt} y_{t-k}$, $a_{0t} = 1 - w_{0t}^S - w_{0t}^T$, $a_{kt} = -w_{kt}^S - w_{kt}^T$, $k \ne 0$, define the linear approximation of the X-11 irregular term. Then,

$$Var(R_t \mid \mathbf{T,S}) = E\{[\sum_k a_{kt}(y_{t-k} - E(y_{t-k} \mid \mathbf{T,S})]^2 \mid \mathbf{T,S}\} = Var(\sum_k a_{kt} e_{t-k} \mid \mathbf{T,S}) ,$$

$$Cov(R_t, R_m \mid \mathbf{T,S}) = Cov(\sum_k a_{kt} e_{t-k}, \sum_k a_{km} e_{m-k} \mid \mathbf{T,S}) = \sum_k \sum_l a_{kt} a_{lm} Cov(e_{t-k}, e_{m-l} \mid \mathbf{T,S}) . \tag{7}$$

It follows from (7) that $v_{tm} = Cov(e_t, e_m \mid \mathbf{S,T})$, $t,m = 1,...,N$, and $u_{tm} = Cov(R_t, R_m \mid \mathbf{S,T})$, $t,m = 1,...,N$, are related by the system of linear equations,

$$\mathbf{U} = D\mathbf{V} , \tag{8}$$

where the matrix D is defined by the weights a_{kt}, $t,k=1,...,N$ through (7), see Pfeffermann (1994) for details. Since the X-11 irregulars, R_t, are observed for $t=1,...,N$, (and, by the assumptions e_t is independent of the true trend and the seasonal components), $Cov(R_t, R_k \mid \mathbf{S},\mathbf{T})$ can be estimated from the observed irregulars, at least at the central part of the series, $t=t^*,...,N-t^*$ for some $t^*>0$. In theory, substituting the estimates in the matrix $\mathbf{U}$ in (8) would then allow estimating $\mathbf{V}$ by solving the equations (D is known). However, the number of equations in (8) for $t=t^*,...,N-t^*$ is smaller than the number of unknown covariates $v_{tm}=Cov(e_t,e_m \mid \mathbf{S},\mathbf{T})$, and therefore (8) can not be solved directly and the solution is very unstable. A possible way to overcome this problem is by assuming that the covariances v_{tk} are negligible (and hence set to zero) for $|t-k|>C$ for some constant C, which allows then to solve the reduced set of equations obtained from (8). See Pfeffermann (1994), Pfeffermann and Scott (1997) and Chen *et al*. (2003), for different approaches to estimation of $\mathbf{U}$ and $\mathbf{V}$.

Remark 3. Pfeffermann (1994) developed his variance estimators under the Postulate: $\sum_k a_{kt}(S_{t-k}+T_{t-k}) \cong 0$ at the center of the series. Although this assumption seems to hold approximately in practice, it is essentially impossible to test it. Note that this Postulate implies that $(\mathbf{T},\mathbf{S})=(\mathbf{T}^{x11},\mathbf{S}^{x11})$ at the center of the series, which is not generally true, see the results of the simulation study below. On the other hand, as shown above, Pfeffermann (1994) method produces consistent estimators for the variance defined by (5).

Estimation of the MSE of the X-11 estimators is complicated. The error term, e_t, can usually be assumed to be independent of the true trend and the seasonal components, and therefore the variance in (5) does not depend on the signal. On the other hand, by (4), the bias of the estimator is a function of $\mathbf{S},\mathbf{T}$ and its value depends on the particular realization of the signal. Estimating the bias requires strict model assumptions that could be hard to validate. Thus, instead of estimating the MSE given the trend and the seasonal components, we propose to estimate instead the expected MSE,

$$E\{MSE[\hat{S}_t]\}=E\{E[(\hat{S}_t-S_t^{x11})^2 \mid \mathbf{S},\mathbf{T}]\}=E\{Var[\hat{S}_t \mid \mathbf{S},\mathbf{T}]\}+E\{Bias^2[\hat{S}_t \mid \mathbf{S},\mathbf{T}]\}. \quad (9)$$

Note that $E\{MSE[\hat{S}_t]\}$ can be considered as the best predictor of $MSE[\hat{S}_t]$ under a square loss function. Assuming that the error term, e_t is independent of the true trend and the seasonal components, the first term in (9) does not depend on the signal and therefore it can be estimated by use of Pfeffermann (1994) method. Denote this estimate by $\hat{V}^{x11}$. In what follows we consider four estimators of $E\{MSE[\hat{S}_t]\}$, two parametric estimators and two non-parametric estimators.

1st parametric bootstrap estimator of $E\{MSE[\hat{S}_t]\}$.

(a) Fit a parametric model to the original series and estimate the parameters of the separate models identified for the trend, the seasonal component, the irregular term and the sampling errors (see Scott 2009 and Step 1 – Step 3 of the following simulation study for the details).

(b) Generate B series, $\mathbf{y}^b, b=1,...,B$, each of sufficient length for applying the symmetric filters to the central N time points by independently generating the four component series, and store the generated series together with the trend and the seasonal components. For each of the generated series compute the difference between X-11 estimate and X-11 signal, $D_t^b=\sum_k w_{kt}^S y_{t-k}^b-\sum_k w_k^S(T_{t-k}^b+S_{t-k}^b)$.

(c) Estimate $\hat{E}\{MSE[\hat{S}_t]\} = \frac{1}{B}\sum_{b=1}^{B}(D_t^b)^2$.

This procedure can also be used for estimating the expected MSE when estimating the original component defined by (1), $E\{MSE[\hat{S}_t]\} = E\{E[(\hat{S}_t - S_t)^2 \mid \mathbf{S,T}]\}$. In the latter case the difference D_t^b has to be defined as, $D_t^b = \sum_k w_{kt}^S y_{t-k}^b - S_t^b$.

The major disadvantage of this procedure is that it requires the identification and estimation of a four components model (trend, seasonal, irregular and sample error) which, as it shown by Scott (2009), is a very challenging task unless the series is very long. Another disadvantage of this procedure is that it is "model-dependent", when the X-11 estimator is not necessarily based on models. Some protection against possible model misspecification can be achieved by the following modification of this procedure:

(a) Fit a parametric model to the original series and estimate the parameters of the separate models identified for the trend, the seasonal component, the irregular term and the sampling errors.

(b) Generate B series, $\mathbf{y}^b, b = 1,...,B$, each of sufficient length for applying the symmetric filters to the central N time points by independently generating the four component series, and store the trend and the seasonal components. For each of the generated series compute the bias, $B_t^b = \sum_k (w_{kt}^S - w_k^S)(T_{t-k}^b + S_{t-k}^b)$.

(c) Estimate $\hat{E}\{Bias^2[\hat{S}_t \mid \mathbf{S,T}]\} = \frac{1}{B}\sum_{b=1}^{B}(B_t^b)^2$ and estimate $\hat{E}\{MSE[\hat{S}_t]\} = \hat{V}^{x11} + \frac{1}{B}\sum_{b=1}^{B}(B_t^b)^2$.

The next three estimators do not require estimation of the four component decomposition model.

1st non-parametric estimator of $E\{MSE[\hat{S}_t]\}$.

By the previous assumptions,

$$E[\sum_k w_{kt}^S y_{t-k} - \sum_k w_k^S y_{t-k}]^2 = E\{E[\sum_k (w_{kt}^S - w_k^S) y_{t-k}]^2 \mid \mathbf{T,S}\}$$
$$= E\{E[\sum_k (w_{kt}^S - w_k^S)(T_{t-k} + S_{t-k} + e_{t-k})]^2 \mid \mathbf{T,S}\}$$
$$= E\{E[\sum_k (w_{kt}^S - w_k^S)(T_{t-k} + S_{t-k})]^2 \mid \mathbf{T,S}\} + E\{E[\sum_k (w_{kt}^S - w_k^S) e_{t-k}]^2 \mid \mathbf{T,S}\}$$
$$= E\{Bias^2[\hat{S}_t \mid \mathbf{S,T}]\} + \sum_k \sum_l (w_{kt}^S - w_k^S)(w_{lt}^S - w_l^S) Cov(e_{t-k}, e_{t-l}) .$$

Therefore, one can estimate the expected square bias as,

$$\hat{E}\{Bias^2[\hat{S}_t \mid \mathbf{S,T}]\} = \max[0, \hat{E}_t^* - \sum_k \sum_l (w_{kt}^S - w_k^S)(w_{lt}^S - w_l^S) C\hat{o}v(e_{t-k}, e_{t-l})], \tag{10}$$

where $C\hat{o}v(e_{t-k}, e_{t-l})$ is estimated via (7) and (8) and $\hat{E}_t^*$ is an estimate of $E_t^* = E[\sum_k w_{kt}^S y_{t-k} - \sum_k w_k^S y_{t-k}]^2 = \sum_k \sum_l (w_{kt}^S - w_k^S)(w_{lt}^S - w_l^S) E(y_{t-k} y_{t-l})$. The latter expectation can be estimated crudely as $\hat{E}_t^* = \sum_k \sum_l (w_{kt}^S - w_k^S)(w_{lt}^S - w_l^S) \hat{E}(y_{t-k} y_{t-l})$, where $\hat{E}(y_a y_b) = \frac{1}{N - |b-a|}\sum_{k=1}^{N-|b-a|} y_k y_{k+|b-a|}$, $|b-a| \le 2a_S$. Recall that $w_{kt}^S = w_k^S = 0$ if $k \notin [-a_S, a_S]$ and therefore $(w_{kt}^S - w_k^S)(w_{lt}^S - w_l^S) = 0$ if $|k-l| > 2a_S$. Note also that $(w_{kt}^S - w_k^S)(w_{lt}^S - w_l^S)$ decreases

as $|k-l|$ increases, so that the main contribution to $\hat{E}_t^*$ comes from $\hat{E}(y_a y_b)$ with relatively small lags $|a-b|$. Actually, for long enough series, $\hat{E}_t^*$ estimates the "average" square-bias rather than the square-bias at time t: since $y_a y_b$ is an unbiased estimate for $E(y_a y_b)$, $\hat{E}(y_a y_b)$ is an unbiased estimate of $\frac{1}{N-|b-a|}\sum_{k=1}^{N-|b-a|} E(y_k y_{k+|b-a|})$, and therefore $\hat{E}_t^*$ estimates $E\{B^{-1}\sum_{b=1}^{B} Bias^2[\hat{S}_t^{(b)}\,|\,\mathbf{S,T}]\}$, where $\hat{S}_t^{(b)} = \sum_{k=-N+t}^{t-1} w_{kt}^S y_{t-k}^{(b)}$; $(y_1^{(b)},...,y_N^{(b)}) = (y_{1-b-N/2},...,y_{N-b-N/2})$.
Define the MSE estimator obtained this way as, $\hat{E}\{MSE[\hat{S}_t]\} = \hat{V}^{x11} + \hat{E}\{Bias^2[\hat{S}_t\,|\,\mathbf{S,T}]\}$.

2nd non-parametric estimator of $E\{MSE[\hat{S}_t]\}$.
Another way of estimating E_t^* is based on a result of Wildi (2005, Section 5): Under suitable regularity assumptions,

$$E_t^* = E[\sum_k w_{kt}^S y_{t-k} - \sum_k w_t^S y_{t-k}]^2 \cong \frac{2\pi}{N}\sum_{k=-[N/2]}^{[N/2]} |\Gamma(\omega_k) - \Gamma_t^*(\omega_k)|^2 \, I_{N\mathbf{Y}}(\omega_k) = \hat{E}_t^*, \qquad (11)$$

where $\omega_k = 2\pi k/N$, $I_{N\mathbf{Y}}(\omega_k)$ denotes the periodogram of the input series computed at ω_k, $\Gamma(\omega_k)$ denotes the transfer function of the symmetric filter w_k^S, $k=-\infty,...,\infty$, and $\Gamma_t^*(\omega_k)$ denotes the transfer function of the asymmetric filter w_{kt}^S, $k=-\infty,...,\infty$, see Wildi (2005) for the detailed definitions and the order of the approximation in (11). Note that the right-hand side of (11) is a function of the observed data and known filters and therefore can be calculated empirically.

2nd parametric bootstrap estimator of $E\{MSE[\hat{S}_t]\}$.
Finally, (10) can be estimated parametrically as follows:

(a) Fit a time series model to the original series $\mathbf{y} = (y_t, t=1,...,N)$, for example, an ARIMA model, and estimate the unknown model parameters.
(b) Generate B series, $\mathbf{y}^b, b=1,...,B$ from the estimated model in (a), each of sufficient length for applying the symmetric filters to the central N time points. For each generated series compute $\tilde{D}_t^b = \sum_k w_{kt}^S y_{t-k}^b - \sum_k w_k^S y_{t-k}^b$.
(c) Estimate $\tilde{E}_t^* = \frac{1}{B}\sum_{b=1}^{B}(\tilde{D}_t^b)^2$ and substitute it into (10).

The clear advantage of this procedure over the 1st parametric method is that it is based on a simpler model that does not necessarily require the identification and estimation of component models and in particular, a model for the sampling error.

Remark 4. Bell and Kramer (1999) estimate $Var(\hat{S}_t - S_t^{B-K}\,|\,\mathbf{S,T,I})$ instead of estimating $E[(\hat{S}_t - S_t^{B-K})^2\,|\,\mathbf{S,T,I}]$. Note that our bias estimation procedures can be equally applied for estimating the MSE when estimating the Bell and Kramer targets. The bias, variance and MSE of the X-11 estimators under Bell and Kramer decomposition, conditional on the true components **S, T** are as follows:

$$Bias[\hat{S}_t\,|\,\mathbf{S,T,I}] = \sum_k (w_{kt}^S - w_k^S)(T_{t-k} + S_{t-k} + I_t). \qquad (12)$$

$$Var[\hat{S}_t \mid \mathbf{S,T,I}] = E\{[\sum_k w_{kt}^S y_{t-k} - E(\sum_k w_{kt}^S y_{t-k} \mid \mathbf{S,T,I})]^2 \mid \mathbf{S,T,I}\} =$$

$$E\{[\sum_k w_{kt}^S (y_{t-k} - S_{t-k} - T_{t-k} - I_{t-k})]^2 \mid \mathbf{S,T,I}\} = E[(\sum_k w_{kt}^S \varepsilon_{t-k})^2 \mid \mathbf{S,T,I}]. \quad (13)$$

$$MSE[\hat{S}_t] = E[(\hat{S}_t - S_t^{x11})^2 \mid \mathbf{S,T,I}] = Var[\hat{S}_t \mid \mathbf{S,T,I}] + Bias^2[\hat{S}_t \mid \mathbf{S,T,I}]. \quad (14)$$

Similar expressions are obtained for the trend. The sampling error term, ε_t, can be usually assumed to be independent of the trend, seasonal and irregular components, and therefore the variance in (13) does not depend on the signal. Estimates of the autocovariances of the sampling error can often be calculated from the survey data. Therefore, the MSE of the X-11 estimator with respect to the Bell and Kramer (1999) decomposition can be obtained either by the 1st parametric bootstrap procedure, with $D_t^b = \sum_k w_{kt}^S y_{t-k}^b - \sum_k w_k^S (T_{t-k}^b + S_{t-k}^b + I_{t-k}^b)$ or $B_t^b = \sum_k (w_{kt}^S - w_k^S)(T_{t-k}^b + S_{t-k}^b + I_{t-k}^b)$, or by estimating, similarly to (10),

$$\hat{E}\{Bias^2[\hat{S}_t \mid \mathbf{S,T,I}]\} = \max[0,\ \hat{E}_t^* - \sum_k \sum_l (w_{kt}^S - w_k^S)(w_{lt}^S - w_l^S) C\hat{o}v(\varepsilon_{t-k}, \varepsilon_{t-l})\ ], \quad (15)$$

and

$$\hat{E}\{MSE[\hat{S}_t]\} = \sum_k \sum_l w_{kt}^S w_{lt}^S C\hat{o}v(\varepsilon_{t-k}, \varepsilon_{t-l}) + \hat{E}\{Bias^2[\hat{S}_t \mid \mathbf{S,T}]\}.$$

3. Simulation study

We illustrate the results of Section 2 by use of simulations. The simulations use the models fitted to the series “Education and Health Services employment” (EDHS), with observations from January, 1996 through December, 2005, (N=120). Our interest in this series is in the month-to-month change in employment. As explained in Scott, *et al*. (2004), we consider the log ratios of the EDHS series, corrected for outliers as the original series (see Remark 1).

We consider as the main objectives of the study the estimation of the trend and the seasonally adjusted (SA) series. Hence, by the decomposition (3), the “X-11 SA series” is, $A_t^{x11} = Y_t - S_t^{x11}$.

The X-11 estimate of the SA series is defined as $\hat{A}_t = y_t - \hat{S}_t = \sum_k w_{kt}^A y_{t-k}$ where $w_{0t}^A = 1 - w_{0t}^S$, $w_{kt}^A = -w_{kt}^S,\ \ k \neq 0$.

Our study consists of the following steps:

Step 1. Fit an ARIMA model to the observed series using X-12-ARIMA.

Step 2. Apply signal extraction to estimate parameters of a model for the signal by use of the REGCMPNT program (Bell, 2003), accounting for the presence of the sampling error component. (We model this component using the autocovariance estimates of the sampling errors as computed by the Bureau of Labor Statistics.)

Step 3. Decompose the signal model into component models employing the experimental software X-12-SEATS.

The models and parameter estimates used in the simulation study are as follows:

Trend- T_t; ARIMA(1,1,2) with parameters -.90, .06, -.94 and disturbance variance 0.5;
Seasonal component- S_t; ARIMA(11,0,11) model with AR-coefficients equal to 1, MA-coefficients equal to, .70, .42, .17, -.04, -.20, -.30, -.37, -.39, -.38, -.34, -.28, and disturbance variance 4.5;
Irregular component- I_t; white noise with disturbance variance 18.0;
Sampling error- ε_t; MA(1) with MA-coefficient -.15 and disturbance variance 58.68.

Step 4. Generate independently 3,000 series from the component models developed in Step 3 and add them up to form new original series $y_t^b, b=1,...,3000$. Each generated series has length *N*+96, $N=120$, so that the application of X-11 to the whole series produces values that are approximately equal to the "final" X-11 values for the central *N* points. Store the series y_t^b and their components, $b=1,...,3,000$.

Step 5. Fix the form of the X-11-ARIMA model fitted to the original series. Re-estimate the parameters of the ARIMA models for each series generated in Step 4 and compute the filter weights w_{kt}^A and w_{kt}^T, reflecting also the backcasts and forecasts produced by the estimated ARIMA model identified for the series.

Step 6. For each series generated in Step 4 and the weights obtained in Step 5 define,

$$\hat{A}_t^b = \sum_k w_{kt}^A y_{t-k}^b \,, \qquad \hat{T}_t^b = \sum_k w_{kt}^T y_{t-k}^b \,,$$

$$\tilde{A}_t^b = E(\hat{A}_t^b \mid \mathbf{T^b}, \mathbf{S^b}) = \sum_k w_{kt}^A (T_{t-k}^b + S_{t-k}^b)\,, \qquad \tilde{T}_t^b = E(\hat{T}_t^b \mid \mathbf{T^b}, \mathbf{S^b}) = \sum_k w_{kt}^T (T_{t-k}^b + S_{t-k}^b)\,,$$

$$A_t^{x11,b} = \sum_k w_k^A (T_{t-k}^b + S_{t-k}^b)\,, \qquad T_t^{x11,b} = \sum_k w_k^T (T_{t-k}^b + S_{t-k}^b)\,,$$

$$B_t^{A,b} = Bias[\hat{A}_t^b \mid \mathbf{S^b}, \mathbf{T^b}] = \sum_k (w_{kt}^A - w_k^A)(T_{t-k}^b + S_{t-k}^b)\,,$$

$$B_t^{T,b} = Bias[\hat{T}_t^b \mid \mathbf{S^b}, \mathbf{T^b}] = \sum_k (w_{kt}^T - w_k^T)(T_{t-k}^b + S_{t-k}^b)\,, \qquad t = 48,...,N+48;\ b = 1,...,3,000\,.$$

Step 7. Compute the Empirical Root MSE of the X-11 estimators with respect to the 'true components' $\mathbf{A}^b, \mathbf{T}^b$ and with respect to the 'X-11 components' $\mathbf{A}^{x11,b}, \mathbf{T}^{x11,b}$, and the Empirical Standard Deviation (SD) of the X-11 estimators.

$$RMSE_{\mathbf{T,S}}(\hat{A}_t) = \sqrt{\frac{1}{3,000}\sum_{b=1}^{3,000}(\hat{A}_t^b - A_t^b)^2}\,, \qquad RMSE_{\mathbf{T,S}}(\hat{T}_t) = \sqrt{\frac{1}{3,000}\sum_{b=1}^{3,000}(\hat{T}_t^b - T_t^b)^2}\,;$$

$$RMSE_{x11}(\hat{A}_t) = \sqrt{\frac{1}{3,000}\sum_{b=1}^{3,000}(\hat{A}_t^b - A_t^{x11,b})^2}\,, \qquad RMSE_{x11}(\hat{T}_t) = \sqrt{\frac{1}{3,000}\sum_{b=1}^{3,000}(\hat{T}_t^b - T_t^{x11,b})^2}\,;$$

$$SD(\hat{A}_t) = \sqrt{\frac{1}{3,000}\sum_{b=1}^{3,000}(\hat{A}_t^b - \tilde{A}_t^b)^2}\,, \qquad SD(\hat{T}_t) = \sqrt{\frac{1}{3,000}\sum_{b=1}^{3,000}(\hat{T}_t^b - \tilde{T}_t^b)^2}\,.$$

Step 8. For each generated series estimate the square bias, $E\{Bias^2[\hat{A}_t \mid \mathbf{S,T}]\}$, by use of the 1st non-parametric estimator, and the variances of the X-11 estimators by the method developed in Pfeffermann (1994). Denote the estimates by $\hat{B}^2(\hat{A}_t^b)$, $\hat{B}^2(\hat{T}_t^b)$; $\hat{V}(\hat{A}_t^b)$, $\hat{V}(\hat{T}_t^b)$, and define,

$$\hat{SD}(\hat{A}_t) = \sqrt{\frac{1}{3,000}\sum_{b=1}^{3,000} \hat{V}(\hat{A}_t^b)}\,, \quad \hat{SD}(\hat{T}_t) = \sqrt{\frac{1}{3,000}\sum_{b=1}^{3,000} \hat{V}(\hat{T}_t^b)}\,,$$

$$\hat{RMSE}(\hat{A}_t) = \sqrt{\frac{1}{3,000}\sum_{b=1}^{3,000} \left(\hat{V}(\hat{A}_t^b) + \hat{E}\{Bias^2[\hat{A}_t^b \,|\, \mathbf{S},\mathbf{T}]\}\right)}\,,$$

$$\hat{RMSE}(\hat{T}_t) = \sqrt{\frac{1}{3,000}\sum_{b=1}^{3,000} \left(\hat{V}(\hat{T}_t^b) + \hat{E}\{Bias^2[\hat{T}_t^b \,|\, \mathbf{S},\mathbf{T}]\}\right)}\,.$$

In addition, we computed also the RMSEs and SDs when estimating the Bell and Kramer (1999) targets: For this, Step 6 is replaced by,

$$A_t^{B-K,b} = \sum_k w_k^A (T_{t-k}^b + S_{t-k}^b + I_{t-k}^b)\,, \quad T_t^{B-K,b} = \sum_k w_k^T (T_{t-k}^b + S_{t-k}^b + I_{t-k}^b)\,,$$

$$B_t^{A,b} = Bias[\hat{A}_t^b \,|\, \mathbf{S^b},\mathbf{T^b}] = \sum_k (w_{kt}^A - w_k^A)(T_{t-k}^b + S_{t-k}^b + I_{t-k}^b)\,,$$

$$B_t^{T,b} = Bias[\hat{T}_t^b \,|\, \mathbf{S^b},\mathbf{T^b}] = \sum_k (w_{kt}^T - w_k^T)(T_{t-k}^b + S_{t-k}^b + I_{t-k}^b)\,,$$

Steps 7 and 8 remain the same.

The results of the study are summarized in Figures 1–4.

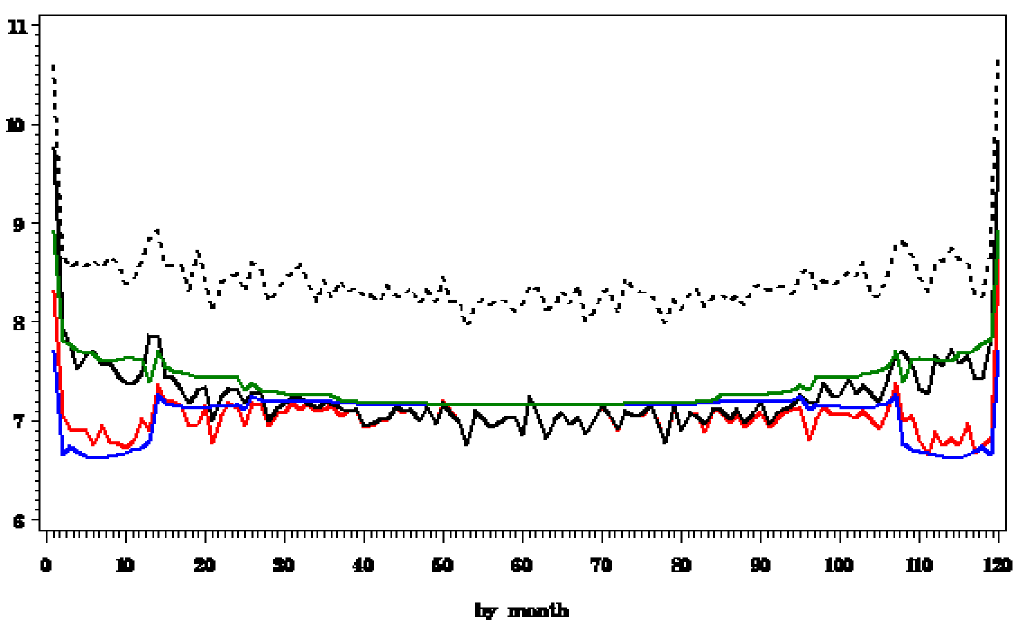

Figure 1. Empirical RMSEs, SDs along with their estimates: $RMSE_{\mathbf{T,S}}(\hat{A}_t)$ is drawn in dashed black, $RMSE_{x11}(\hat{A}_t)$ is in solid black, $SD(\hat{A}_t)$ is in red, $\hat{RMSE}(\hat{A}_t)$ is in green, and $\hat{SD}(\hat{A}_t)$ is in blue.

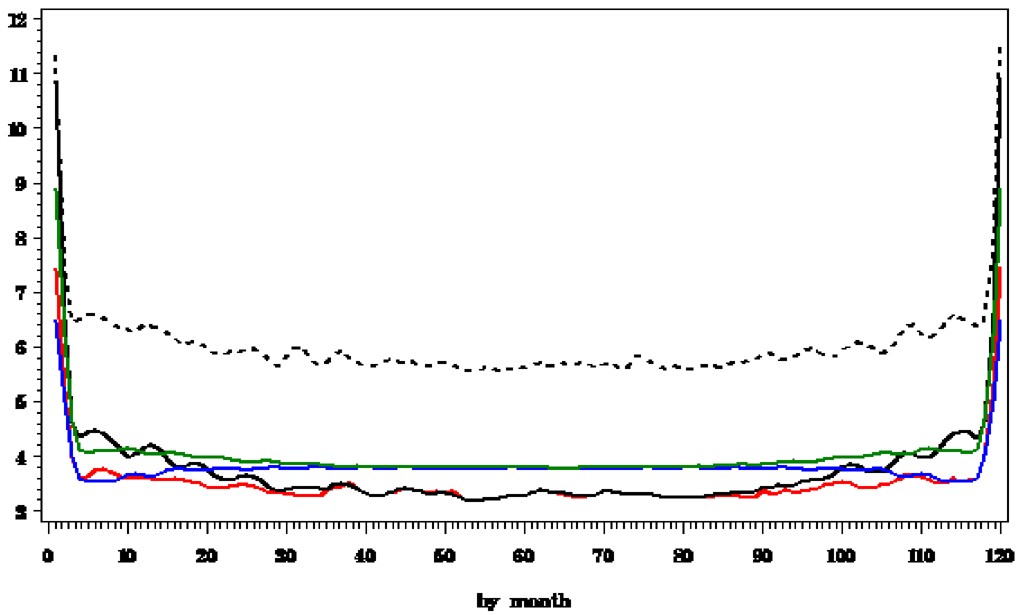

Figure 2. Empirical RMSEs, SDs along with their estimates: $RMSE_{T,S}(\hat{T}_t)$ is drawn in dashed black, $RMSE_{x11}(\hat{T}_t)$ is in solid black, $SD(\hat{T}_t)$ is in red, $R\hat{M}SE(\hat{T}_t)$ is in green, and $\hat{SD}(\hat{T}_t)$ is in blue.

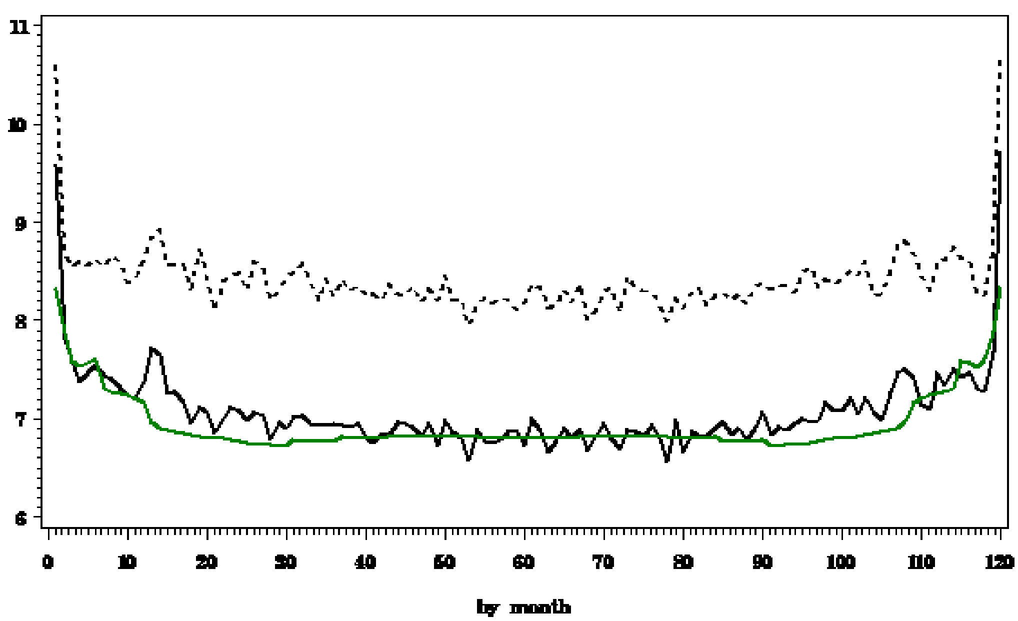

Figure 3. Empirical RMSEs along with their estimates: $RMSE_{T,S}(\hat{A}_t)$ is drawn in dashed black, $RMSE_{B-K}(\hat{A}_t)$ is in solid black, and $R\hat{M}SE_{B-K}(\hat{A}_t)$ is in green.

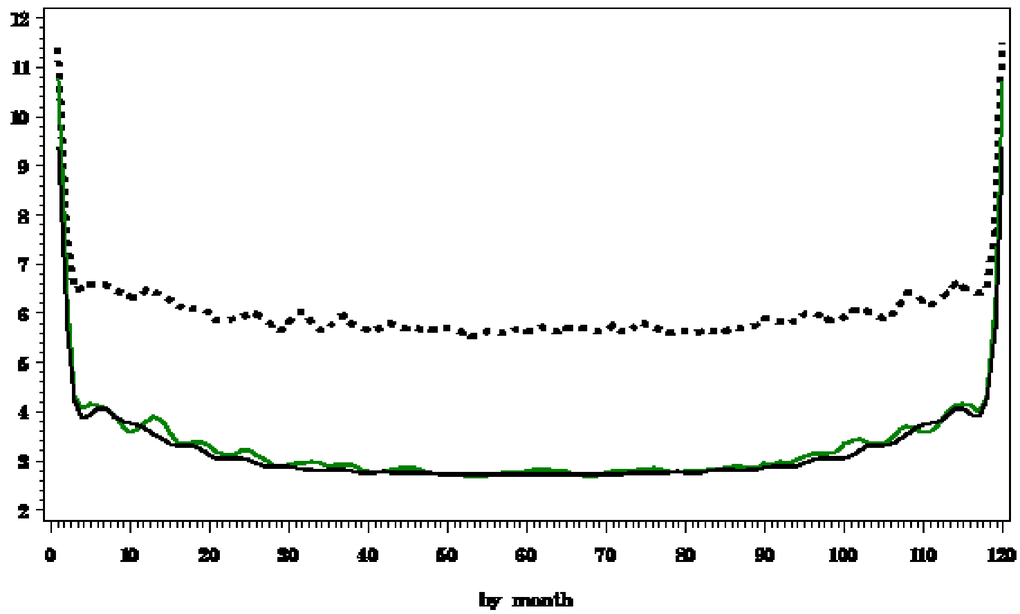

Figure 4. Empirical RMSEs along with their estimates: $RMSE_{\mathrm{T,S}}(\hat{T}_t\ )$ is drawn in dashed black, $RMSE_{B-K}(\hat{T}_t\ )$ is in solid black, and $R\hat{M}SE_{B-K}(\hat{T}_t\ )$ is in green.

<u>Conclusions from simulation study</u>

1) The Empirical RMSEs of the X-11 estimators when estimating the hypothetical components in (1) are higher than the Empirical RMSEs when estimating the "X-11 components", illustrating that the X-11 decomposition is different from the 'model-dependent' decomposition used in the simulations**.**

2) For the 5 years in the center of the series, the X-11 estimators are almost unbiased when estimating the newly defined X-11 components, but at the beginning and at the end of the series there are non-negligible biases.

3) Pfeffermann (1994) variance estimates approximates closely the empirical variances of the X-11 seasonally adjusted estimators when estimating the X-11 SA component, and overestimates slightly the empirical variance of the X-11 Trend estimator when estimating the X-11 trend.

4) The proposed RMSE estimator approximates closely the empirical RMSE when estimating the newly defined components, except for SA component at the very ends of the series. Note that in our study we extend the length of original series only by 96 time points (see Step 4) so that the application of X-11 to the whole series produces values that are *only approximately* equal to the "final" X-11 values for the central *N* points, the latter probably implies some underestimation at the very ends.

5) Conclusions 1, 2 and 4 remain correct for the Bell and Kramer decomposition.

Acknowledgements

Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics.

References

Bell, W. R., and Kramer, M. (1999), Toward Variances for X-11 Seasonal Adjustments, *Survey Methodology* 25, 13-29.

Bell, W. R. (2003), On RegComponent time series models and their applications. *Working paper*, January 2003, Statistical Research Division, U.S. Census Bureau.

Chen, Z.-G., Wong, P., Morry, M., and Fung, H. (2003), Variance Estimation for X-11 Seasonal Adjustment Procedure: Spectrum Approach and Comparison, Statistics Canada Report BSMD-2003-001E

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

Hilmer, S.C., and Tiao, G.S. (1982) An ARIMA-Model-Based Approach to Seasonal Adjustment, *Journal of the American Statistical Association*, 77, pp. 63-70.

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. (1995), Estimation of the Variances of X-11 ARIMA Seasonally Adjusted Estimators for a Multiplicative Decomposition and Heteroscedastic Variances, *International Journal of Forecasting* 11, 271-283.

Pfeffermann, D., and Scott, S. (1997), Variance Measures for X-11 Seasonally Adjusted Estimators; Some Developments with Application to Labor Force Series, *Proceedings of the ASA Section on Business & Economic Statistics*, 211-216.

Pfeffermann, D., Scott, S., and Tiller, R. (2000), Comparison of Variance Measure for Seasonally adjusted and Trend Series, *Proceedings of the 2nd International Conference on Establishment Surveys*, American Statistical Association, 755-764.

Scott, S., Sverchkov, M., and Pfeffermann, D. (2004), Variance Measures for Seasonally Adjusted Employment and Employment Change, *ASA Proceedings of the Joint Statistical Meetings*, 1328-1335.

Scott, S. (2009), On the Impact of Sampling Error on Modeling Seasonal Time Series, *ASA Proceedings of the Joint Statistical Meetings*.

Wildi, M. (2005), *Signal Extraction*, Springer

Wolter, K. M., and Monsour, N. J. (1981), On the Problem of Variance Estimation for a Deseasonalized Series, in *Current Topics in Survey Sampling*, ed. D. Krewski, R. Platek, & J. N. K. Rao, Academic Press, New York, 367-403.