A GENERAL METHOD FOR ESTIMATING THE VARIANCES OF X-11 ESTIMATORS

D. Pfeffermann, Hebrew University of Jerusalem
Mount Scopus, Jerusalem 91905

1. INTRODUCTION

Statistical bureaus throughout the world publish each month seasonally adjusted figures for numerous series. The commonly used procedure for estimating the seasonally adjusted data is the X-11 ARIMA method developed by Dagum (1980) which extends on the original census X-11 algorithm of Shiskin et al. (1967). One criticism of the X-11 procedure however is that it fails to provide estimates for the variability associated with the estimators it produces. In this article we develop a general procedure for estimating the variances of X-11 estimators which has the following properties:

1) The method accounts for both the sampling error of the survey estimators around the corresponding population means and for the error associated with the decomposition of the population means into their unobservable components.
2) The method is largely model free but it does assume that the error terms are stationary with decaying autocovariances.
3) The method can be applied to series estimated based on complex sampling designs with a possible overlap from one survey to the other.
4) The method does not require estimates for the covariances of the unadjusted survey estimators.
5) The method is very simple with negligible additional computing time.

2. COMPONENTS OF VARIANCE

There are several sources of variation which possibly contribute to the total variation of the seasonally adjusted data.

Let $\{y_t ; t=1, ..., N\}$ denote the observed series. In practice, the series $\{y_t\}$ will often consist of survey estimators in which case we define by $\{Y_t\}$ the corresponding population mean values so that we may write

$$y_t = Y_t + e_t \; ; \; E_D(e_t e_{t-k}) = \lambda_k , \quad k=0,1,... \qquad (2.1)$$

The expectation "E_D" is with respect to the design distribution of the estimators over all possible samples from the given population. Notice in (2.1) that the survey errors $\{e_t\}$ are allowed to be serially correlated. Serial correlations occur when the surveys are partially overlapping like in rotating panel surveys.

For the population means we assume the additive decomposition, i.e.

$$Y_t = T_t + S_t + \epsilon_t ; \; E_\xi(\epsilon_t)=0, \; E_\xi(\epsilon_t \epsilon_{t-k}) = \nu_k, \quad k=0,1,..., \qquad (2.2)$$

where T_t represents the trend-cycle and S_t the seasonal component. The error terms $\{\epsilon_t\}$ are the "Decomposition Errors" which are allowed to be serially correlated.

Substituting the equation (2.2) into (2.1) gives the following decomposition,

$$y_t = T_t + S_t + \epsilon_t + e_t = T_t + S_t + I_t, \qquad (2.3)$$

where the error terms $I_t = \epsilon_t + e_t$ are again stationary with the following first and second moments evaluated over the joint "Dξ" distribution

$$E(I_t) = 0, \quad E(I_t I_{t-k}) = V_k = \lambda_k + \nu_k, \; k=0,1,... \qquad (2.4)$$

The decomposition (2.3) (but without the assumption 2.4) is the decomposition postulated by the additive mode of the X-11 method. The alternative mode applicable with X-11 is the multiplicative decomposition by which the observed series is decomposed into a product of the three components. The multiplicative mode is comparable to the application of the additive mode to the logarithms of the series.

It follows from (2.3) that the variance of the observed series y_t may depend on four components of variance. These include

A) The design variance of the survey estimators,
B) The decomposition variance of the population means around the 'true' trend and seasonal components,
C) The variances of the trend and the seasonal components when considered random.

Obviously, when the time series under consideration is not the outcome of a sample survey or that it represents a 'census', the first source of variation no longer exists. The seasonally adjusted estimators are functions of the observations and so the variances of these estimators depend on the same sources of variation.

3. REVIEW OF AVAILABLE METHODS

3.1 The linear approximation to X-11

The methods reviewed in this section are only those developed for the estimation of the variances of X-11 estimators. Most of these methods (including the method proposed in this article) use the linear approximation to X-11 as established by Young (1968) and extended by Wallis (1982) and so we consider the approximation first.

The X-11 program comprises a sequence of moving averages or linear filter operations whose net effect can be represented by a single set of moving averages. Thus, if we denote by $\tilde{N}_t = y_t - \hat{S}_t$ the seasonally adjusted estimator for time t, as

$$\tilde{N}_t = \sum_{j=-(t-1)}^{N-t} w_{j,t} y_{t+j} = \mathbf{w}_t' \mathbf{y}, \quad t=1, \ldots, N \qquad (3.1)$$

with a similar representation for the other estimated components. Notice that the weights $\{w_{j,t}\}$ depend on t since for estimators at the beginning and the end of the series the adjustment is carried out using asymmetric filters. It should be emphasized also that (3.1) defines only an approximation to the actual estimates produced by X-11 because of some nonlinear operations possibly involved in the application of the procedure like the identification and gradual replacement of extreme values and the identification and fitting of ARIMA models.

3.2 Methods proposed

Wolter and Monsour (1982) consider two situations. In the first situation the population values are held fixed so that the only component of variance considered is the design variance of the estimators (equation 2.1). For known variances and covariances of the original estimators $\{y_t\}$, the variances of the seasonally adjusted estimators are obtained straightforwardly by utilizing the approximation (3.1). In the second situation the population values are considered random like in equation (2.2) but with the added assumption that the trend and the seasonal components can be modelled as fixed polynomials in time. Estimators for the variances and covariances $\{V_k\}$ of the total error terms I_t (equation 2.4) are obtained in this case from ordinary least squares residuals and these estimators are substituted for the unknown variances and covariances in the expressions for the variances of the X-11 estimators obtained from (3.1). Armstrong and Gray (1986) propose a replication based method which consists of applying X-11 to each replication and computing the variance between the replicate estimates.

The other methods proposed in the literature are "model dependent" in the sense that the variance estimators depend on stochastic structures assumed for the component series of the trend and the seasonals. The models employed use filters which are similar to the filters employed by X-11. Examples for such models can be found in Cleveland and Tiao (1976), Bell and Hillmer (1984), Burridge and Wallis (1985), and Dagum and Quenneville (1989).

3.3 Discussion

The major disadvantage of the methods described in the previous section is their lack of generality. The methods proposed for estimating the design variances assume either that the design variances and covariances of the unadjusted estimators are known or that the sample can be split to form replicates of the original sample.

The procedure proposed by Wolter and Monsour (1982) for the case where the population values are considered random is restricted to situations where the trend and the seasonal components can be approximated by fixed polynomials in time. The model dependent methods assume stochastic structures for the seasonal and the nonseasonal components series. Identifying the models for the components from the model holding for the observed series is difficult in practice since different component models can lead to the same overall model. It seems therefore that with the present state of art, the use of models for estimating the variances of the X-11 estimators is not practical as a general routine. The use of models by government bureaus was always questionable because of the possible effects of model misspecifications on the published estimators and with the added assumptions and computational complexities, the use of a model is even more problematic.

Government bureaus usually attempt to compute and publish the design variances of survey estimators but the use of this approach when estimating the variances of seasonally adjusted data is not founded. The very act of seasonally adjusting the observed series implies the use of a stochastic time series model either explicitly or implicitly (as in X-11) and by restricting to the design variances the uncertainty underlying the evolution of the population values (the series Y_t in our notation) is not taken into account.

The method described in the next section attempts to compromise between the design and model based approaches by conditioning on the population values of the trend and the seasonal components but not on the Decomposition error. Thus, the variances estimated are with respect to the joint distribution of the total irregular terms, $I_t = e_t + \epsilon_t$ (e_t is the survey error, ϵ_t is the decomposition error, see equation 2.4), with the other components of variance held fixed. This has the further practical advantage that the models generating the trend and the seasonal components need not be specified beyond what is already assumed by X-11. Moreover, under the assumption that the X-11 estimators are conditionally unbiased, the estimators of the conditional variances can be viewed also as estimators of the unconditional variances.

4. THE NEW METHOD

4.1 Alternative Definitions for the Variance

In what follows we consider the linear approximation to X-11 described in section 3.1. We describe the estimation of the variances of the seasonally adjusted data. Estimators for the variances of the other components are obtained in the very same way.

Substituting the model (2.3) into (3.1) yields the following decomposition

$$\begin{aligned} \tilde{N}_t &= \sum_{j=-(t-1)}^{N-t} w_{j,t} y_{t+j} \\ &= \sum_{j=-(t-1)}^{N-t} w_{j,t}(T_{t+j}+S_{t+j}) + \sum_{j=-(t-1)}^{N-t} w_{j,t} I_{t+j} \end{aligned} \tag{4.1}$$

Denote by $VAR(\tilde{N}_t | \mathbf{T}, \mathbf{S})$ the conditional variance of the estimator $\tilde{N}_t$ for given realizations of the trend and the seasonals. Then, by (4.1),

$$VAR(\tilde{N}_t | \boldsymbol{T}, \boldsymbol{S}) = VAR\left\{ \sum_{j=-(t-1)}^{N-t} w_{j,t} I_{t+j} \right\}. \tag{4.2}$$

The variance in (4.2) is over the joint distribution of the irregular terms $\{I_t = e_t + \epsilon_t, T = 1, \ldots, N\}$. This variance is different from the variance computed under the model dependent approach, which is taken over the distribution of the prediction error $(\tilde{N}_t - N_t) = [(y_t - \hat{S}_t) - (y_t - S_t)] = -(\hat{S}_t - S_t)$ so that it coincides with the variance of the seasonal component. Thus, if

$$\hat{S}_t = \sum_{j=-(t-1)}^{N-t} {}_{j,t} y_{t+j} \tag{4.3}$$

defines the linear approximation to the seasonal component where $\omega_{j,t} = -w_{j,t}$ for $j \neq 0$ and $\omega_{0,t} = (1 - w_{0,t})$,

$$\begin{aligned} VAR(\tilde{N}_t - N_t) | \boldsymbol{T}, \boldsymbol{S} &= VAR(\hat{S}_t - S_t) | \boldsymbol{T}, \boldsymbol{S} \\ &= VAR \sum_{j=-(t-1)}^{N-t} \omega_{j,t} I_{t+j} \end{aligned} \tag{4.4}$$

Dr. D. Binder suggested to me (private communication) that the target in the case of a survey should be the estimation of $N^*_t = Y_t - S_t$, the seasonally adjusted population value. Estimating this quantity is consistent with the common routine of estimating the population mean values Y_t. We find that $(\tilde{N}_t - N^*_t) = [(y_t - \hat{S}_t) - (Y_t - S_t)] = [e_t - (\hat{S}_t - S_t)]$ and so

$$VAR[(\tilde{N}_t - N_t^*|\boldsymbol{T},\boldsymbol{S}]$$

$$= VAR[(\hat{S}_t - S_t)|\boldsymbol{T},\boldsymbol{S}] + (1-2\tilde{\omega}_{0,t})\, VAR_D(e_t) \qquad (4.5)$$

$$-2\sum_{\substack{j=-(t-1)\\ j\neq 0}}^{N-t} \tilde{\omega}_{j,t} COV_D(e_{t+j}, e_t)$$

The computation of this variance requires in principle estimators for the design variance and autocovariances of the survey errors. Notice however that $(\tilde{N}_t - N^*_t)$ can be written as,

$$(\tilde{N}_t - N_t^*) = \sum_{j=-(t-1)}^{N-t} w^*_{j,t}[(e_{t+j}+\epsilon_{t+j})-\epsilon_t], \qquad (4.5^*)$$

where $w^*_{j,t} = -\omega_{j,t}$ for $j\neq 0$; $w^*_{0,t} = (1-\omega_{0,t})$ so that when $Cov(\epsilon_t, \epsilon_{t+k}) = 0$ for $k\neq 0$ only an estimator for the design variance is required.

It is interesting to compare the variance in (4.5) with the variance in (4.2). Assuming that the X-11 estimators of the seasonal effects are conditionally unbiased, (see the discussion in section 4.2),

$$E(\tilde{N}_t|\boldsymbol{T},\boldsymbol{S}) = E[(T_t+S_t+I_t-\hat{S}_t)|\boldsymbol{T},\boldsymbol{S}] = T_t.$$

Thus, whereas the variance in (4.5) is with respect to the distribution of $\tilde{N}_t$ around the seasonally adjusted value $N^*_t = T_t+\epsilon_t$ in the population, the variance in (4.2) is around the trend level T_t.

4.2 Estimation of the Variances

It follows from (4.2), (4.4) and (4.5) that the problem of estimating the conditional variances reduces to the problem of estimating the variances and autocovariances of the irregular terms defined in (2.4). Let

$$R_t = \sum_{j=-(t-1)}^{N-t} a_{j,t} y_{t+j};\quad (\sum_{j=-(t-1)}^{N-t} a_{j,t}=0) \qquad (4.6)$$

define the linear approximation to the residual terms produced by X-11. Substituting the equation (2.3) into the expression for R_t gives the decomposition

$$R_t = \sum_{j=-(t-1)}^{N-t} a_{j,t} y_{t+j}$$

$$= \sum_{j=-(t-1)}^{N-t} a_{j,t}(T_{t+j}+S_{t+j}) + \sum_{j=-(t-1)}^{N-t} a_{j,t} I_{t+j}. \qquad (4.7)$$

As can be seen from (4.7), the time series of the X-11 residuals is not stationary because of the use of different moving averages at different sections of the series and the contaminating effect of the means

$$M_t = \sum_{j=-(t-1)}^{N-t} a_{j,t}(T_{t+j}+S_{t+j}).$$

However, assuming that the signal $(T_t + S_t) = E(y_t)$ is a smooth function in time, the set of residuals in the centre of the series $[24 < t < (N - 24)$ for monthly series for all practical purposes] are approximately stationary. This is so because the particular weights used for the filter in the centre of the series guarantee that for sufficiently smooth signals the means M_t are close to zero and because the weighted residuals

$$E_t = \sum_{j=-(t-1)}^{N-t} a_{j,t} I_{t+j}$$

form a stationary series since the filter is fixed and the irregular terms $\{I_t\}$ are stationary. The set of weights $\{a_{j,t}\}$ used for the computation of the X-11 residuals (for example, corresponding to the case of using the 13 term Henderson moving average for the estimation of the trend levels for a monthly series) are symmetric ($a_{t-i,t} = a_{t+i,t}$ for all i) and add up to zero. The corresponding transfer function has zero power at the seasonal frequencies $\{2\pi k/12, k=1, ..., 6\}$ and at frequencies lower than $2\pi/12$ implying that the filter completely removes the trend and the seasonal components.

The requirement for the smoothness of the signal is not defined in exact mathematical terms but the condition that $M_t \approx 0$ has a clear statistical interpretation. It implies that the X-11 estimators of the signals are conditionally unbiased, i.e.

$$E[y_t|\boldsymbol{T},\boldsymbol{S}] = E[(\hat{T}_t+\hat{S}_t)|\boldsymbol{T},\boldsymbol{S}] + M_t = T_t + S_t. \qquad (4.8)$$

One could argue that the condition $E[R_t|T, S] = M_t \approx 0$ is a necessary condition for the appropriateness of the X-11 method when applied to a particular series. The closeness of the means M_t to zero is not defined in exact mathematical terms either but it will be assumed that the variance and autocovariances of the series $\{M_t\}$ are negligible compared to the corresponding variance and autocovariances of the series $\{E_t\}$.

Under this assumption, the variance and autocovariances of the X-11 residuals $\{R_t\}$ in the centre of the series can be expressed as linear combinations of the variance and autocovariances of the irregular terms $\{I_t\}$ in a straightforward manner, exploiting the relationship (4.7). Inverting the resulting equations allows to express the variance and autocovariances of the series $\{I_t\}$ as linear combinations of the variance and autocovariances of the series $\{R_t\}$. Replacing the theoretical moments of the series $\{R_t\}$ by the corresponding sample moments yields the desirable estimators for the moments of the irregular terms.

The use of this procedure has a major deficiency in that the estimators can become unstable as the number of equations increases. However, it may be assumed that the autocovariances of the irregular terms damp to zero so that after a certain cutoff C,$V_k = COV(I_t, I_{t-k}) \approx 0$. Setting $V_k = 0$ for $k > C$ reduces the number of equations required for the estimation of the variance and the remaining autocovariances of the irregular term to (C+1). Next we formulate the procedure more rigorously assuming a given value of C. We discuss the specification of C in section 4.3.

Proposition 1: For $24 < t < N-24$,

$$M_t = \sum_{j=-(t-1)}^{N-t} a_{j,t}(T_{t+j}+S_{t+j}) = 0$$

Proposition 2: For $k > C$, $V_k = COV(I_t, I_{t-k}) = 0$

Let $\mathbf{U}' = (U_0, U_1, ... U_C)$ be the $1\times(C+1)$ row vector with elements $U_k = COV(R_t, R_{t-k})$, $k=0, ..., C$ and $\mathbf{V}'$ be the $1\times(C+1)$ row vector of the covariances $\{V_k\}$ arranged in the same order. By (4.7) and proposition 1,

$$R_t = \sum_{j=-(t-1)}^{N-t} a_{j,t} I_{t+j}$$

and so by proposition 2

$U_k = a_{k0}V_0 + a_{k1}V_1 + \ldots + a_{kC}V_C = \mathbf{a'}_k\mathbf{V}$, $k=0 \ldots C$. (4.9)

The set of equations (4.9) can be written as $\mathbf{U} = A\mathbf{V}$ where A is $(C+1)\times(C+1)$, yielding in turn the set of equations $\mathbf{V} = A^{-1}\mathbf{U}$. Estimating $\mathbf{U'}$ by $\hat{\mathbf{U}}' = (\hat{U}_0,\ldots,\hat{U}_C)$ where

$U_k = \frac{1}{N-48} \sum_{t=25+k}^{N-25} (R_t - \overline{R})(R_{t-k} - \overline{R})$ gives the estimators

$$\hat{V} = A^{-1}\hat{U} \tag{4.10}$$

The procedure described by (4.10) is essentially an application of the method of moments and retains the properties of that method. In particular, under the propositions 1 and 2, the estimators $\hat{V}$ are consistent in the sense that $\operatorname{plim}_{N\to\infty} \hat{V}_k = V_k$, $k=0,\ldots,C$.

4.3 Specification of the cutoff point

The practical use of the procedure requires the specification of the cutoff point C. Often, knowledge of the sampling design implies at least an upper limit for the cutoff C. Having set an upper limit C_L for C, it would seem desirable to establish a selection procedure which can aid in choosing between plausible values of C. The following procedure has been used in the empirical examples.

1) Set C=0 and calculate $\hat{V}_0$. Estimate $\tilde{U}_1^{(0)} = a_{1,0}\hat{V}_0$ and $\tilde{U}_2^{(0)} = a_{2,0}\hat{V}_0$. Compute $D_1^{(0)} = |\hat{U}_1 - \tilde{U}_1^{(0)}|\hat{U}_0$ and $D_2^{(0)} = |\hat{U}_2 - \tilde{U}_2^{(0)}|\hat{U}_{(0)}$. If $\max[D_1^{(0)}, D_2^{(0)}] \le 0.1$, specify C = 0. Otherwise execute step 2.

2) Set C=1 and calculate $\hat{V}_0$ and $\hat{V}_1$. Estimate $\tilde{U}_2^{(1)} = a_{3,0}\hat{V}_0 + a_{3,1}\hat{V}_1$ and $\tilde{U}_3^{(1)} = a_{4,0}\hat{V}_0 + a_{4,1}\hat{V}_1$. Compute $D_2^{(1)} = |\hat{U}_2 - \tilde{U}_2^{(1)}|/\hat{U}_0$ and $D_3^{(1)} = |\hat{U}_3 - \tilde{U}_3^{(1)}|/\hat{U}_0$. If $\max[D_2^{(1)}, D_3^{(1)}] \le 0.1$, specify C = 1. Otherwise execute step 3.

3) Repeat step 2 with the value of C and the number of estimated autocovariances $\{V_k\}$ increased each time by 1. Stop the procedure for $C = q \le C_L$ such that

$$D_{q+1}^{(q)} = |\hat{U}_{q+1} - \tilde{U}_{q+1}^{(q)}|/\hat{U}_0 \le 0.1$$

and $D_{q+2}^{(q)} = |\hat{U}_{q+2} - \tilde{U}_{q+2}^{(q)}|/\hat{U}_0 \le 0.1$ where

$$\tilde{U}_g^{(q)} = a_{g,0}\hat{V}_0 + \ldots + a_{g,q}\hat{V}_q,\ g=q+1,\ q+2.$$

Specify C=q. Otherwise specify $C = C_L$.

4.4 Estimation of the Unconditional Variances

In the previous sections we focused on the estimation of the conditional variances, given the realizations of the component series of the trend and the seasonal effects.

Suppose that the trend and the seasonal effects are generated by stochastic processes and denote by $VAR(\tilde{N}_t - N_t) = VAR(\hat{S}_t - S_t)$ the unconditional variance of $(\tilde{N}_t - N_t)$. We may decompose the variance in the form

$$VAR(\tilde{N}_t - N_t) = E_1\{VAR_2[(\tilde{N}_t - N_t)\,|\,\mathbf{T,S}]\} + VAR_1\{E_2[(\tilde{N}_t - N_t)\,|\,\mathbf{T,S}]\} \tag{4.11}$$

where we use the index "2" to indicate the conditional moments given the component series and the index "1" to indicate the unconditional moments. We conclude from (4.11) that under the assumption that the trend and the seasonal components estimators are conditionally unbiased the estimators of the conditional variances can be viewed also as estimators of the unconditional variances. The same property applies for the variance of the prediction errors $(\tilde{N}_t - N^*_t)$.

5. EXAMPLES

This section contains three examples illustrating the application of the method. The first two examples use simulated series. The last example uses an actual series and compares the variance estimators obtained by the method with estimators obtained when considering only the design variances.

5.1 Simulation Results

We generated two groups of series, each composed of 100 independent monthly series of length 192. The first group of series was generated by adding white noise irregular terms to a fixed signal, i.e.,

$$Y_t^{(n)} = T_t + S_t + I_t^{(n)};$$

$$I_t^{(n)} \sim N(0,36),\ t=1,\ldots,192,\ n=1\ldots100 \tag{5.1}$$

The second group of series was generated by adding irregular terms from an AR(1) process to the same signal, i.e.,

$$Y_t^{(n)} = T_t + S_t + I_t^{(n)};$$

$$I_t^{(n)} = 0.5\, I_{t-1}^{(n)} + V_t,\ V_t,\ V_t \sim N(0,36),$$

$$t=1,\ldots,192;\ n=1,\ldots,100 \tag{5.2}$$

The signal $\{T_t + S_t\}$ was taken as the estimated signal when seasonally adjusting the series "Employed Males in Quebec". The signal values in the centre of the series along with the weighted signal

$M_t = \sum_{j=-(t-1)}^{N-t} a_{j,t}(T_{t+j} + S_{t+j})$ revealed that the series $\{M_t\}$ has indeed all its values close to zero as suggested in Section 4.2. Also, $Var(M_t) = 0.527$ compared to

$Var\{(T_t + S_t)\} = 1896.15$.

Table 1 displays the frequencies of the cutoff values C as obtained by application of the selection procedure described in Section 4.3

Table 1. Frequencies of the cutoff values for the two groups of series

	Cutoff Values					
Groups	0	1	2	3	4	5
I	63	20	11	3	2	1
II	1	34	28	23	10	4

As expected, for the first group of series the cutoff values are highly skewed to the left with 63 out of the 100 series yielding a value of C=0 (the correct cutoff in the case of a white noise). For the second group of series the cutoff values are mostly between 1 and 3 with the value of C=0 obtained in only one case. The fact that the procedure selected the value C=1 for 34 series can be explained by the fact that for the AR(1) process used in this study, $Corr(I_t, I_{t-k}) \leq 0.25$ for $k \geq 2$ which implies that the main contribution to the variance and autocovariances of the X-11 residuals originates from the variance and first autocovariance of the irregular terms.

Upon studying the percentiles P_α for α = 10, 25, 75, 90 of the empirical distribution of the estimators of the standard deviation (S.D.) of the seasonally adjusted series, along with the mean of the estimators and the true S.D. (equation 4.2), for each of the 192 months, the method yielded essentially unbiased estimators - a relative bias of 3 percent may well be attributed to sampling variations. The distribution of the estimators around the true values is approximately symmetric with well acceptable variances, considering in particular that the method is largely model free. In this respect the variances of the estimators were somewhat larger at the two ends of the series, but the CV's were very stable throughout the series.

5.2 Variance Estimation for the Series "Total Unemployed Males in Canada"

This series was chosen for illustration because it permits a comparison of the variance estimators obtained under the new method with estimators obtained when using the design based approach described in Section 3.2. The design variance and autocovariances of the unadjusted estimators are listed in Table 2. The data analyzed cover the years 1981-1989. The estimators listed in the table are average figures and the second row lists the corresponding standard deviations (SD) of the monthly estimators.

Table 2. Average and Standard Deviation of the Design Variance and Autocovariances

	Average Values	Standard Deviations
λ_0	436	51
λ_1	228	26
λ_2	148	25
λ_3	105	20
λ_4	73	6.5
λ_5	49	13
λ_6	42	7.6
λ_7	37	5.5
λ_8	37	16
λ_9	27	13
λ_{10}	25	7.4
λ_{11}	30	8.7
λ_{12}	22	-

The cutoff value selected by the selection procedure of section 4.3 is C=3. Our initial guess was C=4 and it was based on the estimators given in Table 2, implying $\hat{\lambda}_g/\hat{\lambda}_0 \leq 0.11$ for $g \geq 5$.

We studied plots of the estimators of the SD of the seasonally adjusted estimators (equation 4.2) as obtained for the two cutoff values, along with the estimators obtained by the design based approach, that is, when considering only the design variances and autocovariances. Plots of the estimators of the SD of the seasonally adjusted estimators around the population seasonally adjusted values (equation 4.5) were also studied. The average design SD of the unadjusted estimators was $\sqrt{\hat{\lambda}_0} = 20.9$.

The major conclusions were as follows:

1) The use of the two cutoff values gives quite similar results indicating that the variance estimators are not very sensitive to the choice between cutoff values which pass the goodness of fit criteria set in the stepwise procedure.
2) The use of only the design variance and autocovariances for estimating the variances of the seasonally adjusted estimators may result in considerable underestimation.
3) The design variances of the seasonally adjusted estimators are lower than the design variances of the unadjusted estimators in the centre of the series and are of a similar magnitude at the two ends of the series. Wolter and Monsour (1981) and Armstrong and Gray (1986) report similar results.
4) The variances of the seasonally adjusted estimators around the population seasonally adjusted values are smaller than the design variances of the unadjusted estimators in the centre of series but larger at the two ends. This result is explained by the particular filters used for the estimation of the seasonally adjusted values. Notice so that the variances are smaller than the variances around the trend values shown in Figure 6, i.e. $Var[(\tilde{N}_t - N^*_t) | T,S] < Var[(\tilde{N}_t - T_t) | T,S]$.

6. CONCLUDING REMARKS - EXTENSION OF THE PROCEDURE

The results presented in this article refer to the estimation of the variances of the linear approximation to X-11. As mentioned in section 3.1, the X-11 ARIMA algorithm uses several "nonlinear" operations. These include the identification and gradual replacement of extreme values and the identification and estimation of ARIMA models for augmenting the series by one and possibly two years of forecasted values. A possible way to assess the contribution to the variance implied by the nonlinear operations is by simulating replications of the original series. Work in this direction is currently under way.

ACKNOWLEDGEMENT

This study was carried out while the author was on Sabbatical leave at Statistics Canada under its Research Fellowship program. The author is grateful to M. Morry for helpful comments and suggestions and to N. Laniel for providing the computer program for calculating the linear approximation to X-11. The author is particularly indebted to P. Wong who developed the programs used for the application of the new method.

REFERENCES

Armstrong, J.B. and Gray, G.B. (1986), "Variance Estimation for Seasonally Adjusted Survey Data," Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 466-471.

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, 291-349.

Burridge, P. and Wallis, K.F. (1985), "Calculating the Variance of Seasonally Adjusted Series," Journal of the American Statistical Association, 80, 541-552.

Cleveland, W.P. and Tiao, G.C. (1976), "Decomposition of Seasonal Time Series: A model for the Census X-11 program," Journal of the American Statistical Association, 71, 581-587. Dagum, E.B. (1980), "The X-11 ARIMA Seasonal Adjustment Method," Catalogue 12-564E, Statistics Canada, Ottawa, Ontario, K1A 0T6.

Dagum, E.B. and Quenneville, B. (1989), "Variance of X-11-ARIMA Seasonally Adjusted Series," Time Series Research and Analysis Division, Statistics Canada, Ottawa, Ontario, K1A 0T6.

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.

Wallis, K.F. (1982), "Seasonal Adjustment and Revision of Current Data: Linear filters for the X-11 Method," Journal of the Royal Statistical Society, A, 145, 74-85.

Wolter, K.M. and Monsour, N.J. (1981), "On the problem of Variance Estimation for a Deseasonalized Series," in Current Topics in Survey Sampling: An International Symposium, eds. D. Krewski, R. Platek, and J.N.K. Rao: Academic Press: New York, pp. 367-403.

Young, A.H. (1968), "Linear Approximations to the Census and BLS Seasonal Adjustment Methods," Journal of the American Statistical Association, 63, 445-471.