

s Statistique a Canada

Methodology Branch

Time Series Research and Analysis Division

Direction de la méthodologie

Division de la recherche et de l'analyse des chroniques

11-614 no.90-12 c. 2

Canada



WORKING PAPER NO. TSRA 90-012

TIME SERIES RESEARCH & ANALYSIS DIVISION METHODOLOGY BRANCH



C. 2

CALCULATION OF PREDICTION INTERVALS FOR SEASONALLY ADJUSTED LABOUR FORCE SERIES

by

Estela Bee Dagum and Benoit Quenneville

This is a preliminary version. Do not quote without authors' permission. Comments are welcome.



Calculation of Prediction Intervals for Seasonally Adjusted Labour Force Series by

Estela Bee Dagum and Benoit Quenneville

Time Series Research and Analysis Division 7-E R.H. Coats Bldg. Statistics Canada Ottawa (Ontario) Canada KIA OT6 (613) 951-9876

July 1990



ABSTRACT

- I -

This paper describes a state-space approach for the approximation of the mean square errors (MSE) of the X-11-ARIMA seasonally adjusted estimates and their changes over time. The trend-cycle, seasonality and irregular unobserved-components are assumed to follow simple stochastic models of the class found by Cleveland and Tiao (1976) and Burridge and Wallis (1984) that approximates well the Census X-11 variant filters. Similar models have been applied by Kitagawa and Gersch (1984). The seasonal and trend-cycle values from X-11-ARIMA are used to obtain an initial value of the mean of the state vector and initial estimates of the variances of both the observation noise and the noise processes of the unobserved-components models (UCM). These initial values of the variances are used to obtain maximum likelihood estimates (MLE) by the method of scoring. The only other estimate required by the fixed-interval smoothing algorithm, the initial state covariance matrix, is set to be a large multiple of the identity matrix. The Kalman filter and the fixed interval smoother are applied to the original series to obtain the estimates of the UCM as well as their corresponding MSE. Finally, the MSE of the X-11-ARIMA estimates are approximated by the MSE of the UCM estimates if their differences are not significant.

Key words: Maximum likelihood estimation, Kalman filter, fixed interval test, smoother, seasonally adjusted, month-to-month change and ratio.

RESUME

Cet article présente une méthode pour l'approximation de la variance des données corrigées de leurs variations saisonnières par le progiciel X-ll-ARMMI.

La méthode utilise une formulation de vecteur d'état et le filtrage de Kalman. Des modèles stochastiques sont définis pour les composantes cyclo-tendancielle, saisonnière et irrégulière de la série chronologique. Ces modèles sont ajustés à la décomposition du X-ll-ARMMI pour obtenir les estimés des paramètres nécessaires au filtrage de Kalman. Le filtrage de Kalman et le lissage sur un intervalle fixe sont appliqués sur la série originale pour obtenir des estimés "stochastiques" des composantes de la série ainsi que leurs variances. Finalement, la variance de l'estimé "X-ll-ARMMI" est approximée par celle de l'estimé stochastique si leur différence n'est pas significative.



1. INTRODUCTION.

The need for the development of standard errors of seasonally adjusted data as published by statistical bureaus has a long standing. The President's Committee to Appraise Employment and Unemployment Statistics (1962) recommended: "that estimates of the standard errors of seasonally adjusted data be prepared and published as soon as the technical problems have been solved". Seventeen years later, the National Commission on Employment and Unemployment Statistics (1979) reemphasized the importance of standard errors for seasonally adjusted series and urged the Census Bureau to undertake research to develop them. In response to this goal, Wolter and Monsour (1981) developed a procedure based on the linear filters of the Method II-X-11-variant (Shiskin, Young and Musgrave, 1967) to calculate the variance of seasonally adjusted data. These authors considered two situations, one, where the components were assumed as deterministic and thus only the sample variability contributes to the variance of the seasonally adjusted value; and, two, where the components are assumed to be stochastic processes and the nonstationary part of the time series is removed by fitting a polynomial in time. This procedure offered a simplified approximation to the variance of the X-ll estimates given the two assumptions on the kind of variability that affected the data and the fact that the linear filters themselves are an approximation of what the method really does to actual series. With the same kind of reasoning, Burridge and Wallis (1984) developed unobserved-components models of the ARIMA type that approximate the seasonal adjustment filters used by the X-ll variant and derived measures of variance using the Kalman filter (Burridge and Wallis, 1985). Similarly, measures of the asymptotic variance could be calculated from the ARIMA model developed by Cleveland and Tiao (1976) as an approximation of the symmetric filters of the X-11 variant. Hillmer (1985) made a major contribution for computing variances of the components estimates from model based procedures such as Hillmer and Tiao (1982) and Burman (1980); and generalized Pierce (1980) results for the revision of current seasonally adjusted data. Hillmer (1985) calculated the total variance as the sum of the conditional asymptotic variance (from the case in which a doubly infinite realization is

available) and the variance from the forecasts and backcasts values that are needed to replace the missing observations from the future and the past when dealing with actual series.

The studies concerned with measures of variance of seasonally adjusted data by the X-ll-variant approached the problem from the viewpoint of its linear filters. These linear filters, however, are approximations of what the method really does under the assumptions of: (1) additive decomposition, (2) no treatment of extreme values, (3) no trading-day variations and (4) only the filters of the <u>default</u> option are applied to estimate the seasonal and trend-cycle components.

The main purpose of this paper is to present a new procedure that approximates the mean square errors (MSE) of the unobserved-components and their changes as <u>really estimated from actual data</u> by the X-ll-ARIMA method (Dagum, 1980) which is applied by most statistical bureaus, with or without the ARIMA extrapolations.

Section 2 introduces the models assumed for the unobserved-components, trend-cycle, seasonality and irregular and discusses the relationship between the models and the various filters of the X-11-ARIMA method. Section 3 gives a brief description of the Kalman filter and fixed interval smoother. Section 4 presents the procedure followed to obtain maximum likelihood estimates of the signal to noise ratios and the observation noise variance. Section 5 gives the variances of the X-11-ARIMA seasonally adjusted values and of their month-to-month changes for the additive, log-additive and multiplicative decompositions. Section 6 analyses the results for two seasonally adjusted series, one additively and the other multiplicatively. Section 7 gives the conclusions.

- 2 -

2. THE X-11-ARIMA METHOD AND THE MODELS FOR THE UNOBSERVED-COMPONENTS.

The X-ll-ARIMA seasonal adjustment method assumes that a series Y_t can be decomposed into trend-cycle C_t , seasonality S_t and irregular variations I_t , either in an additive manner:

$$Y_t = C_t + S_t + I_t,$$
 (2.1)

a multiplicative manner:

$$Y_{t} = C_{t}S_{t}I_{t}$$
(2.2)

or, a logarithmic manner:

$$\log Y_t = \log C_t + \log S_t + \log I_t.$$
(2.3)

This method is based on moving averages or linear smoothing filters implying that the time series components are stochastic and thus, cannot be closely approximated by simple functions of time over the entire range of the series. The X-ll-ARIMA method consists of extending the original series at each end with extrapolated values from seasonal ARIMA models and then seasonally adjusting the extended series with a combination of the X-ll filters and the ARIMA model extrapolation filters.

The models proposed here to estimate the MSE of the X-11-ARIMA seasonally adjusted values (for levels and changes) are variants of those found by Cleveland and Tiao (1976) and Burridge and Wallis (1984) that approximate closely the X-11 seasonal adjustment filters. Similar models have been also used by Kitagawa and Gersch (1984) in their seasonal adjustment method.

The basic unobserved-components model has the form:

$$Y_t = \mu_t + \gamma_t + \epsilon_t , t=1, \dots, T$$
 (2.4)

where μ_t , γ_t and ϵ_t are the trend-cycle , seasonal and irregular components respectively.

The trend is here assumed to follow a second order stochastically perturbed difference equation:

$$(1-B)^{2}\mu_{t} = \eta_{t}, \ \eta_{t} \sim N(0, \sigma_{\eta}^{2})$$
(2.5)

or equivalently:

$$\mu_{t} = 2\mu_{t-1} - \mu_{t-2} + \eta_{t}, \qquad (2.6)$$

where η_t is an independently identically distributed (i.i.d.) sequence and B denotes the backsift operator $(B\mu_t = \mu_{t-1})$.

The model for the seasonal component is defined by:

$$\gamma_{t} = \sum_{i=1}^{s-1} \gamma_{t-j} + \omega_{t}, \ \omega_{t} \sim N(0, \sigma_{\omega}^{2})$$
(2.7)

where ω_t is an i.i.d. sequence and s is the number of "seasons" in the year. The seasonal pattern is thus slowly changing but by a process that ensures that the sum of the seasonal components over any s consecutive time periods has an <u>expected</u> value of zero and a variance that remains constant over time.

The disturbances η_t and ω_t are independent of each other and of the irregular component $\epsilon_t \sim i.i.d. N(0, \sigma^2)$.

It is straightforward in the Kalman filter and related state-space smoothing algorithm to add additional components models for trading-day, both deterministic and stochastic (Dagum and Quenneville(1990)), outliers, intervention analysis or explanatory (regression) variables (Harvey (1984)) and autocorrelated sampling error (Pfeffermann and Friedman (1988)). These are not discussed here as we limit ourselves to the trend-cycle, seasonal and irregular components that form the basic structural model.

Models (2.5) and (2.7) have the same autoregressive operators as the models given by Cleveland and Tiao (1976) and Burridge and Wallis (1984) but not the moving average operators. There are several reasons why we limited our models to be purely autoregressive. First, the moving average operators of Burridge and Wallis (1984) models change for each X-ll asymmetric filter and the moving average for the symmetric filter is different from that given by Cleveland and Tiao (1976). Second, Burridge and Wallis (1984) and Cleveland and Tiao (1976) models were constructed for the default option of the X-ll filters but non-standard options are often applied by Statistics Canada and other statistical bureaus for the seasonal adjustment of their series. Third, the asymmetric filters of X-ll-ARIMA change not only depending on its position in time but with the ARIMA model used for the extrapolations. Fourth, it is shown by Burridge and Wallis (1984) that a very simple model such as:

$$(1-B)\mu_{t} = \eta_{t},$$
 (2.8)

$$\gamma_{t} = -\Sigma \gamma_{t-j} + \omega_{t}, \qquad (2.9)$$

$$i=1$$

with appropriately chosen innovation variances accounts for 97.1% of the

s-1

- 5 -

total variations in the weights of the symmetric seasonal adjustment filter.

3. KALMAN FILTER AND FIXED INTERVAL SMOOTHER.

In this context and for the case of monthly observations, the state space model consists of a measurement equation:

$$Y_t = \underline{z}' \underline{\alpha}_t + \epsilon_t \tag{3.1}$$

and a transition equation:

$$\underline{\alpha}_{t} = C\underline{\alpha}_{t-1} + D\underline{r}_{t} \tag{3.2}$$

where

is a fixed vector.

$$\underline{\alpha}_{t} = (\mu_{t} \ \mu_{t-1} \ \gamma_{t} \ \gamma_{t-1} \ \cdots \ \gamma_{t-10}) \tag{3.4}$$

is the state vector and

		2	-1	0	0	0	0	0	0	0	0	0	0	0		1	0	
		1	0	0	0	0	0	0	0	0	0	0	0	0		0	0	
		0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1		0	1	
		0	0	1	0	0	0	0	0	0	0	0	0	0		0	0	
		0	0	0	1	0	0	0	0	0	0	0	0	0		0	0	
		0	0	0	0	1	0	0	0	0	0	0	0	0		0	0	
С	-	0	0	0	0	0	1	0	0	0	0	0	0	0	D =	0	0	
		0	0	0	0	0	0	1	0	0	0	0	0	0		0	0	
		0	0	0	0	0	0	0	1	0	0	0	0	0		0	0	
		0	0	0	0	0	0	0	0	1	0	0	0	0		0	0	
		0	0	0	0	0	0	0	0	0	1	0	0	0		0	0	
		0	0	0	0	0	0	0	0	0	0	1	0	0		0	0	
		0	0	0	0	0	0	0	0	0	0	0	1	0		0	0	

are fixed matrices. The signal variance vector

$$\underline{\tau}_{t} = (\eta_{t} \,\omega_{t}) \tag{3.6}$$

is normally distributed with mean zero and covariance matrix $\sigma^2 Q$,

$$Q = diag(\sigma_n^2/\sigma_n^2 \sigma_n^2/\sigma_n^2)$$

and σ^2 is the variance of ϵ_t , i.i.d. N(0, σ^2) independently of $\underline{\tau}_t$.

Let \underline{a}_t be the minimum mean square estimate (MMSE) of $\underline{\alpha}_t$ and $\sigma^2 P_t$ its covariance matrix, i.e. $\sigma^2 P_t = E[\underline{a}_t - \underline{\alpha}_t][\underline{a}_t - \underline{\alpha}_t]'$. The MMSE of $\underline{\alpha}_{t+1}$ given \underline{a}_t and P_t is then given by:

 $a_{t+1|t} = Ca_t$

(3.7)

(3.5)

with MSE matrix:

$P_{t+1 t} = CP_tC' + DQD'.$	(3.8)
Once Y_{t+1} becomes available, the estimate of $\underline{\alpha}_{t+1}$ can be updated as	follows:
$\underline{a}_{t+1} = \underline{a}_{t+1 t} + \underline{K}_{t+1} \underline{v}_{t+1}$	(3.9)
$P_{t+1} = (I - \underline{K}_{t+1} \underline{z}') P_{t+1 t}$	(3.10)
$v_{t+1} = Y_{t+1} - \underline{z}' \underline{a}_{t+1 t}$	(3.11)
$\underline{K}_{t+1} - \underline{P}_{t+1 t\underline{Z}/f_{t+1}}$	(3.12)
$f_{t+1} = \underline{z}' P_{t+1 t\underline{z}} + 1.$	(3.13)

Starting values <u>a</u>₀ and P₀ and knowledge of the covariance matrix Q are needed to implement the Kalman filter given by (3.7) to (3.13).

The Kalman filter yields the MMSE of $\underline{\alpha}_t$ given the information available up to time t. However, once all the observations are available, a better estimator can be obtained. One of the techniques for computing such an estimator is the fixed interval smoother. The fixed interval smoother is a set of recursions which start with the Kalman filter estimates \underline{a}_T and P_T , and works backwards. If $\underline{a}_{t|T}$ and $\sigma^2 P_{t|T}$ denote the smoothed estimates and its covariance matrix, the smoothing equations are given by:

$$\underline{a}_{t|T} = \underline{a}_{t} + \mathbf{P}^{*}_{t}(\underline{a}_{t+1|T} - \mathbf{C}_{\underline{a}_{t}})$$
(3.14)
with

$$P_{t|T} = P_t + P_t^*(P_{t+1|T} - P_{t+1|t})P_t^*$$
(3.15)

where

 $P_{t}^{*} = P_{t}C'(P_{t+1|t})^{-1}.$ (3.16)

4. ESTIMATION OF a0, PO AND Q.

4.1 Additive variant.

Maximum likelihood estimates (MLE) of the signal to noise ratios, σ_{η}^2/σ^2 and $\sigma_{\omega}^2/\sigma^2$, are obtained by the method of scoring on the concentrated log-likelihood function. The MLE of σ^2 is obtained analytically, conditional on the estimates of the signal to noise ratios. Using the prediction error decomposition (Harvey (1981a)), the likelihood function L can be written in the form:

$$\log(L) = -T/2 \log(2\pi) -T/2 \log(\sigma^2) -1/2 \sum_{t=1}^{T} \log(f_t) -1/2 \sigma^{-2} \sum_{t=1}^{T} v_t^2 / f_t \quad (4.1)$$

where T is the number of observatins and v_t and f_t are defined by (3.11) and (3.13). Differentiation of (4.1) with respect to σ leads to the MLE of σ^2 , given by:

$$\hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} v_t^2 / f_t.$$
 (4.2)

The scalar parameter, σ^2 , may be concentrated out of the log-likelihood function leaving the concentrated log-likelihood function:

$$\log(Lc) = -T/2 \log(2\pi) -T/2 \log(\sigma^2) -1/2 \sum_{t=1}^{1} \log(f_t) -T/2.$$
(4.3)

Numerical optimization has to be carried out with respect to the signal to noise ratios to maximize equation (4.3). While Kitagawa and Gersch (1984) use a very simple grid search in their paper, we here apply a more accurate procedure based on the method of scoring. A brief description of this method follows.

Let $\underline{x}' = (x_1, x_2) = (\sigma_{\eta}^2 / \sigma^2, \sigma_{\omega}^2 / \sigma^2)$ be the vector of unknown parameters to be estimated. The method of scoring (Harvey (1981b)) is the iterative scheme:

 $\underline{x}^{(i)} = \underline{x}^{(i-1)} + I^{-1}(\underline{x}^{(i-1)}) \operatorname{DlogL}(\underline{x}^{(i-1)})$ (4.4)

where $\underline{x}^{(i)}$ is the estimate of \underline{x} at the i-th iteration, $I(\underline{x})$ is the Fisher information matrix evaluated at \underline{x} and $DlogL(\underline{x})$ is the matrix of derivatives of the log-likelihood function evaluated at \underline{x} .

From (4.1) the elements of the matrix DlogL are:

$$\frac{\partial}{\partial x_{i}} = \frac{T}{\Sigma} \frac{\partial}{\partial p_{t}} \frac{\partial}{\partial p_{t}}, \quad i=1,2. \quad (4.5)$$

The density of the innovation process (3.11) is here symbolized by p_t to simplify the notation. The derivative of the log-likelihood of the t-th innovation is:

$$\frac{\partial}{\partial x_i} \log(p_t) = -.5f_t \frac{\partial}{\partial x_i} t_t - \sigma f_t v_t \frac{\partial}{\partial x_i} t_t + .5\sigma f_t v_t \frac{\partial}{\partial x_i} t_t.$$
(4.6)

Here, the derivatives of f_t and v_t are computed recursively from the Kalman filter equations (3.7) to (3.13) starting with the derivatives of \underline{a}_0 and \underline{P}_0 equal to zero.

The (i,j)-element of the information matrix is given by:

$$I(i,j) = E(\frac{\partial}{\partial x_i} \log(L))(\frac{\partial}{\partial x_i} \log(L))$$
(4.7)

$$-\sum_{t=1}^{T} E\left(\frac{\partial}{\partial x_{i}}\log(p_{t})\right) \left(\frac{\partial}{\partial x_{j}}\log(p_{t})\right).$$
(4.8)

Since the innovations v_t are independent and normally distributed with mean zero and variance $\sigma^2 f_t$ it follows that

$$E \left(\frac{\partial}{\partial x_{i}}\log(p_{t})\right) \left(\frac{\partial}{\partial x_{j}}\log(p_{t})\right) = .5f_{t}\frac{\partial}{\partial x_{i}}f_{t}\frac{\partial}{\partial x_{i}}f_{t} + .5\sigma f_{t}\frac{\partial}{\partial x_{i}}v_{t}\frac{\partial}{\partial x_{i}}v_{t} \quad (4.9)$$

from which (4.8) is easily derived. These results agree with those of Engle and Watson (1981) for the case of univariate observations.

The above discussion applies when the log-likelihood (4.1) is maximized. In our context, it is the concentrated log-likelihood (4.3) that has to be maximized. However it can be easily verify that:

$$\frac{\partial}{\partial \mathbf{x}_{i}} \log(\mathbf{L}c) = \frac{\partial}{\partial \mathbf{x}_{i}} \log(\mathbf{L}(\sigma^{2}))$$
(4.10)

where the right hand side of (4.10) is the derivative of the log-likelihood $\hat{\rho}_{=\sigma^2}$.

For values not near the boundary of zero an assymptotic t-statistics can be constructed for each parameters, namely:

$$t_i = x_i / MSE(x_i)^{1/2}$$
 (4.11)

Here an estimate of the MSE matrix of the MLE of \underline{x} is provided by the inverse of the last information matrix obtained in the iterative procedure (4.4).

Starting values \underline{x}^0 are needed to initiate the iterative scheme (4.4). These are obtained via simple moment estimates from the X-ll-ARIMA decomposition. Hence, $\underline{x}^0 = (\overline{\sigma_\eta}^2/\overline{\sigma^2}, \overline{\sigma_\omega}^2/\overline{\sigma^2})$ where:

$$\overline{\sigma_{\eta}}^{2} = (T-2)^{-1} \sum_{t=3}^{T} (c_{t}-2c_{t-1}+c_{t-2})^{2}, \qquad (4.12)$$

$$\tilde{\sigma}_{\omega}^{2} = (T-11)^{-1} \sum_{t=12}^{T} \omega_{t}^{2}$$
(4.13)

and

where

$$\omega_{t} = \sum_{j=1}^{12} s_{t-j} \qquad (4.14)$$

$$\tilde{\sigma}^2 = T^{-1} \sum_{t=1}^{T} (Y_t - c_t - s_t)^2.$$
(4.15)

Here ct and st denotes the estimates of Ct and St from X-11-ARIMA.

Finally the iterative process (4.4) is stopped when the relative change in the value of the log-likelihood function between two succesive iterations is smaller than a prespecified constant (ex:.001.).

The estimate of the initial state vector $\underline{\alpha}_0$ requires knowledge of $(\mu_0, \mu_{-1}, \gamma_0, \gamma_{-1}, \ldots, \gamma_{-10})$. Since c_{-1} and s_{-10} to s_0 are not readily available from X-11-ARIMA the first eleven month of data are used to estimate $\underline{\alpha}_{11}$ by $\underline{a}_{11}=(c_{11}, c_{10}, s_{11}, s_{10}, \ldots, s_1)'$ and the Kalman filter is started at time t-12. This ensure that both the UCM and the X-11-ARIMA estimates start at the same point. The initial covariance matrix P₁₁ is taken to be kI₁₃ where k is a large constant and I₁₃ is the identity matrix of order 13.

4.2 Logarithmic decomposition.

For the logarithmic decompositon all the calculations are done in the log metric. That is, the estimates (4.12) and (4.14) are obtained with $log(c_t)$ and $log(s_t)$ respectively, the estimate for the irregular are obtained with $log(Y_t/c_ts_t)$. The initial state vector is $a_{11} - (log(c_{11}), log(c_{10}), log(s_{11}), \ldots, log(s_1))$ ' and the Kalman filter uses $log(Y_t)$ instead of Y_t .

4.3 Multiplicative decomposition.

For the multiplicative decomposition, the preliminary estimates of the signal to noise ratio for the trend and the first two elements of \underline{a}_{11} are obtained as in the logarihmic decomposition.

The preliminary estimation of the signal to noise ratio for the seasonality is obtained as follows. First, in the estimate (4.14) we use s_t-1 instead of $log(s_t)$. This is due to the fact that in the

multiplicative version of X-11-ARIMA, the seasonal factors are constrained to have an arithmetic mean instead of a geometric mean equal to 1. The estimate of the noise variance (4.15) is obtained with (Y_t/s_tc_t) -1. Finally the ratio of the two estimates gives the preliminary estimate of the signal to noise ratio for the seasonal. Since log(x) is approximately equals to x-1 for x close to 1, the two estimates of the observation noise variance are very close. The last eleven elements of <u>All</u> are estimated by $(s_{11}-1,\ldots,s_1-1)$ and the Kalman filter uses $log(Y_t)$ instead of Y_t .

5. VARIANCES OF X-11-ARIMA SEASONALLY ADJUSTED VALUES AND OF THEIR MONTH TO MONTH CHANGES.

5.1 Additive Decomposition.

The Kalman filter and fixed interval smoother as described by equations (3.7) to (3.16) are applied using the initial state vector, covariance matrix of the initial state vector, the MLE of the signal to noise variances $(\sigma_{\eta}^2/\sigma^2, \sigma_{\omega}^2/\sigma^2)$ and the MLE of σ^2 . The X-11-ARIMA seasonally adjusted data is given by Y_t -st and the seasonally adjusted estimate form $\hat{\gamma}_t$ the UCM is Y_t - $\hat{\gamma}_t$, where $\hat{\gamma}_t$ denotes the estimate of γ_t . The MSE is defined by:

 $MSE(\gamma_{t}) = E[(Y_{t} - \gamma_{t}) - (Y_{t} - \gamma_{t})]^{2} = E[\gamma_{t} - \gamma_{t}]^{2}$ (5.1)

and is given by $\sigma^2 P_{t|T}(3,3)$, where $P_{t|T}(3,3)$ is the third row, third column element of $P_{t|T}$.

In the analysis of seasonally adjusted data, comparisons of month-to-month changes are often done to assess the direction and magnitude of the short-term trend.

The method discussed here allows the estimation of the MSE of changes between any two months included in the state vector. The change in the UCM seasonally adjusted data betwen month t and t-a, for a-1,.,10 is given by:

 $(Y_{t}-\gamma_{t}) - (Y_{t-a}-\gamma_{t-a})$ (5.2)

with MSE:

$$E[((Y_{t-\gamma_{t}}) - (Y_{t-a-\gamma_{t-a}})) - ((Y_{t-\gamma_{t}}) - (Y_{t-a-\gamma_{t-a}}))]^{2} = E[(\gamma_{t-a-\gamma_{t-a}}) - (\gamma_{t-\gamma_{t}})]^{2} = C$$

$$E(\gamma_{t-a}-\gamma_{t-a})^2 + E(\gamma_t-\gamma_t)^2 - 2E[(\gamma_{t-a}-\gamma_{t-a})(\gamma_t-\gamma_t)]$$
(5.3)

which is:

$$\sigma^{2}(\mathbf{P}_{t|T}(3+a,3+a)+\mathbf{P}_{t|T}(3,3)-2\mathbf{P}_{t|T}(3,3+a))$$
(5.4)

for the smoothed UCM estimate.

5.2 Logarithmic and multiplicative decompositions.

In the logarithmic and multiplicative decompositions the estimates from the models are obtained in the log metric, so for practical purposes, it is necessary to make a transformation back to the original metric. Denoting by Y_t/Γ_t the seasonally adjusted estimate in the original metric, Granger and Newbold(1976) show that the estimates that minimizes the MSE is:

$$Y_{t}/\Gamma_{t} = \exp\{(\log Y_{t} - \gamma_{t}) + .5\sigma^{2}P_{t|T}(3,3)\}$$
(5.5)

with MSE:

$$MSE(Y_{t}/\Gamma_{t}) = \exp\{2(\log Y_{t}-\gamma_{t}) + \sigma^{2}P_{t|T}(3,3)\}$$

$$[\exp\{\sigma^{2}P_{t|T}(3,3)\} - 1].$$
(5.6)

Similar transformations are applied for month to month comparisons.

6. APPLICATIONS.

The seasonal adjustment of actual data presents problems that require special attention, particularly, the identification and replacement of extreme values; the use of ARIMA extrapolations to reduce revisions of the current seasonally adjusted estimate; and the use of concurrent or year-ahead seasonal factors to obtain a current seasonally adjusted value. These problems have been taken into consideration for the estimation of the UCM following the same procedure of X-11-ARIMA when applicable.

The method discussed here has been tested with a large sample of series from Canada and the United States with very good results. For illustration purposes two cases are shown here. Canada Total, Unemployed Male Aged 25 and Over (CA-UM), for the period January 1975 to December 1985, is used to illustrate the additive decomposition and U.S. Total, Nonagricultural Employed Male Aged 20 and Over (US-EM), for the same time period, is used to illustrate the multiplicative decomposition.

Example 1: CANADA TOTAL OF UNEMPLOYED MALE AGED 25 AND OVER (CA-UM),

The official X-11-ARIMA decomposition for this series is of the additive type with one year of forecasts from an ARIMA $(0,1,2)(0,1,1)_{12}$ model. Table 1 gives the results of the MLE iterative procedure for the signal to noise The starting value of the vector $(\sigma_n^2/\sigma^2, \sigma_\omega^2/\sigma^2)$ is (.6563,.3907) ratios. with the matrix of the derivatives of the concentrated log-likelihood (4.5) and (4.6) given in the second column and the information matrix (4.8) given in the third and fourth columns. The initial value of log(Lc) (constant terms are not included) is -390.5. Finally, the initial estimate of the noise variance σ^2 is 16.855. At the 6-th iteration the relative increase in the values of log(Lc) is less than .001 and the procedure is stopped. The final estimates of the vector of the signal to noise ratios $(\sigma_n^2/\sigma^2, \sigma_\omega^2/\sigma^2)$ is (2.5605,.1151) and the final estimate of σ^2 is 10.7422. The values of the derivatives Dlog(Lc) indicate that log(Lc) is relatively flat at the final estimates as compare to its value at the initial estimates.

Given the estimates of the signal to noise ratios, the UCM estimates are calculated and compared with the X-11-ARIMA estimates. Figure 1A.1 shows the original series and the X-11-ARIMA seasonally adjusted series. Figure 1A.2 indicates how close the X-11-ARIMA seasonally adjusted values are to the smoothed seasonally adjusted UCM estimates. Figure 1A.3 gives the 95% predictive interval of the seasonally adjusted X-11-ARIMA series. Figure 1A.4 shows how small are the relative differences (in percentage) between the smoothed seasonally adjusted UCM and the seasonally adjusted X-11-ARIMA values (the relative difference is calculated as: 100 (UCM - X-11-ARIMA)/X-11-ARIMA). The correlation coefficient between the seasonal factors produced by the X-11-ARIMA and the UCM methods is .99848. This clearly indicates that their linear relationship is very strong and in the same direction. To asses whether or not the difference in the seasonal factors of the two methods is significant, we perform a basic statistical analysis on their relative differences. The results of Table 4 indicate that the relative differences are in fact very small. Figure 1A.5 shows the MSE's of the smoothed seasonally adjusted UCM estimates. The graph of the smoothed MSE's versus time has a concave shape with jumps every year. The MSE's are the smallest in the middle of the series which agrees with the

results obtained by Wolter and Monsour (1981).

All figures 1B refer to the month-to-month changes instead of levels as discussed above. Figure 1B.2 gives the 95% predictive interval. Values falling above (below) the zero line indicate positive (negative) changes in the seasonally adjusted series. Particularly, the period from September 1981 till December 1982 stands out with the only exceptions of October and November 1981 and January 1982. (May 1981 till December 1982 corresponds to the deep Canadian recession). Figure 1B.3 shows the MSE's of the month-to-month changes of the seasonally adjusted values.

Figures 1C.1 to 1C.3 are the same as 1B.1 to 1B.3 but for the changes between period t and t-2 instead t and t-1. The May 1981 to December 1982 recession period is clearer in this case than in figure 1B.2.

One of the main reasons for seasonally adjusting series is to get a clearer signal of the short-term trend. Consequently, it is important to assess if a change of direction in a <u>current</u> seasonally adjusted value indicates the presence of a true turning point.

The month-to-month changes of the seasonally adjusted data for the whole period 1975-1985 were different from zero and positive in May 1981 and from September 1981 till December 1982 with the exceptions of October, November 1981 and January 1982. Using the series from January 1975 till May 1981 and adding one month at a time, we wanted to identify how long it would take to the method discussed here to detect these changes of direction using <u>current</u> seasonally adjusted figures.

Table 3A provides the 95% predictive interval constructed around the month-to-month changes. It can be seen that the change from April to May 1981 is significantly different from zero and remains so when more data are added to the series. For five out of eight month-to-month changes, the current seasonally adjusted values are good estimators of the corresponding "historical" values obtained when the series ends in December 1985. For the months of June, July and October the historical 95% predictive intervals give a different signal than the current and the first five revisions. Given the amount of irregularity in the UM series, we applied the Month for Cyclical Dominance (MCD) measure of X-11-ARIMA as an indicator of the length of the month-span where the contribution of the cyclical variations surpasses that of the irregulars. For the UM series the MCD is equal to 2 indicating that to assess the short term trend, comparisons must be made between the current seasonally adjusted values and 2 months before.

Table 3B shows the predictive intervals for the 2-months span changes of the UM series. The results clearly indicate that these changes are significantly different from zero and positive since June 1981 with only two exceptions, August-June and November-September 1981. Furthermore, seven out of the eight months analysed give the same trend direction as the "historical" estimates.

Tables 3A and 3B also indicate that there is no need to revise the current seasonally adjusted data during the current year to obtain better estimates of the short term trend. These results conform with those given by Dagum (1982.a and 1982.b) concerning the revisions of the seasonal adjustment filters of X-11-ARIMA.

Example 2: AMERICAIN NONAGRICULTURAL TOTAL EMPLOYED MALE AGED 20 AND OVER (US-EM).

The official X-11-ARIMA decomposition for this series is of the multiplicative type with one year of forecasts from an ARIMA $(0,1,2)(0,1,1)_{12}$ model on the log-transformed data. Table 2 and figures 2 are the equivalent of Tables 1 and figures 1 of the CA-UM series. Here, the calculations of the estimates of the signal to noise ratios are done with the data in the log metric. In this case, the correlation coefficient between the seasonal factors is .99799. As in the CA-UM case, the results of Table 4 show that the relative differences are very small.

7. CONCLUSIONS.

This study has introduced a method that calculates MSE's of seasonally adjusted values given by the X-11-ARIMA computer package. The method basically consists of fitting simple stochastic models to the X-11-ARIMA estimates to obtain the initial state vector and signal to noise ratios. Maximum likelihood estimates are then obtained using the method of scoring. The models assumed for the unobserved-components belong to the class found by Cleveland and Tiao (1976) and Burridge and Wallis (1984) that approximates well the default option of the Census X-11 filters. These models have also been used by Kitagawa and Gersch (1984) for developing a seasonal adjustment method.

The Kalman filter and smoother are applied to the original series to

obtain estimates and corresponding MSE's of the unobserved-components models (UCM).

This method has been tested with a large sample of series from Canada and United State and produced very good results. For illustrative purposes two series are discussed, namely the Canada Total of Unemployed Male, aged 25 and Over (1975 to 1985) additively seasonally adjusted and the U.S. Employed Male, aged 20 and Over (1975-1985) multiplicatively seasonally adjusted.

BIBLIOGRAPHY

- Burman, J.P., (1980), "Seasonal Adjustment by Signal Extraction", <u>JRSS-A</u>, 143, 321-337.
- Burridge, P. and Wallis, K.F., (1984), "Unobserved-Components Models for Seasonal Adjustment Filters", JBES, 2, 350-359.
- Burridge, P. and Wallis, K.F., (1985), "Calculating the Variance of Seasonally Adjusted Series", <u>JASA</u>, 80, 541-552.
- Cleveland, W.P. and Tiao, G.C., (1976), "Decompositon of Seasonal Time Series: A Model for the Census X-11 Program", JASA, 71, 581-587.
- Dagum, E.B., (1980), "The X-11-ARIMA Seasonal Adjustment Method", <u>Statistics</u> <u>Canada</u>, Catalogue No. 12-564E.
- Dagum,E.B.,(1982a), "Revision of Time Varying Seasonal Filters", Journal of Forecasting, 1, 173-187.
- Dagum, E.B., (1982b), "The Effects of Asymmetric Filters on Seasonal Factor Revisions", JASA, 77, 732-738.
- Dagum,E.B. and Quenneville,B.,(1990), "Dynamic Linear Models for Time Series Components", Presented at the Conference on Seasonality and Econometric Models, C.R.D.E. Montréal University, Marsh 11-12, 1990.
- Engle, R. and Watson, M., (1981), "A One-Factor Multivariate Time Series Model of the Metropolitan Wage Rates", <u>JASA</u>, 76, 774-781.
- Granger, C.W.J. and Newbold, P., (1976), "Forcasting Transformed Series", JRSS-B, 38, 189-203.
- Harvey, A.C., (1981a), Time Series Models, Oxford, Philip Allan.
- Harvey, A.C., (1981b), <u>The Econometric Analysis of Time Series</u>, Oxford, Philip Allan.
- Kitagawa,G. and Gersch,W., (1984), "A Smoothness Priors-State Space Modeling of Time Series With Trend and Seasonality", <u>JASA</u>, 79, 378-389.
- Hillmer, S.C. and Tiao, G.C., (1982), "An ARIMA-Model-Based Approach to Seasonal Adjustment", JASA, 77, 63-70.
- Hillmer, S.C., (1985), "Measures of Variability for Model-Based Seasonal Adjustment Procedures", JBES, 3, 60-68.

- 16 -

National Commission on Employment and Unemployment Statistics,(1979), <u>Counting the Labor Force</u>, Washington, D.C., U.S. Government Printing Office.

- Pfefferman, D. and Friedman, D., (1988), "Estimation od Population means Composed of Trend and Seasonal Components Using Data from Repeated Surveys", Research paper, <u>Hebrew University</u>.
- Pierce, D.A., (1980), "Data Revisions with Moving Average Seasonal Adjustment Procedures", Journal of Econometrics, 14, 93-114.
- President's Committee to Appraise Employment and Unemployment Statistics, (1962), <u>Measuring Employment and Unemployment</u>, 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 11 Seasonal Adjustment Program", Technical Paper No. 15, <u>U.S. Department of Commerce</u>, Bureau of Economic Analysis.
- Wolter,K.M. and Monsour,N.J., (1981), "On the Problem of Variance Estimation for a Deseasonalized Series", <u>Current Topics in Survey Sampling</u>, Academic Press, 367-403.

TABLE 1

iter.	X	Dlog(Ic)	Infol	Info2	log(Lc)	σ2
1	0.6563	23.0126	45.6511	-8.3246	-390.5	16.8555
	0.3907	-16.4130	-8.3246	73.4508		
2	1.1294	7.9834	15.7843	-3.6832	-381.4	14.2136
	0.2209	-12.6483	-3.6832	149.3000		
3	1.6183	3.2588	7.9475	-0.9249	-378.1	12.7694
	0.1483	-7.1623	-0.9249	239.9000		
4	2.0250	1.5975	5.2440	0.3890	-377.0	11.8726
	0.1200	-2.0816	0.3890	306.2000		
5	2.3302	0.9382	4.0627	0.9565	-376.6	11.2365
	0.1128	0.9849	0.9565	325.7000		
6	2,5605	0.6379	3.4278	1.1995	-376.4	10.7422
	0.1151	2.3210	1.1995	313.4000		

Canada Total Unemployed Male - Aged 25 and Over. Results of the MLE iterative procedure

TABLE 2

Americain Total Employed Male - Aged 20 and Over. Results of the MLE iterative procedure

iter.	X	Dlog(Lc)	Infol	Info2	log(Lc)	<i>o</i> 2
1	0.6713	19.0286	41.1345	-2.8237	730.7	8.4E-7
	0.3145	-19.6394	-2.8237	84.8538		
2	1.1191	5.6092	15.1192	2.9928	739.3	7.8E-7
	0.0979	-7.0544	2.9928	356.5000		
3	1.4926	2.6819	8.9297	4.1770	740.8	7.1E-7
	0.0750	7.0391	4.1770	506.5000		
4	1.7876	1.8205	6.4329	4.2188	742.5	6.4E-7
	0.0865	6.5982	4.2188	417.2000		
5	2.0621	1.3397	4.9646	4.0366	742.0	5.9E-7
	0.0995	5.5047	4.0366	342.8000		
6	2.3214	1.0348	4.0058	3.7925	742.4	5.5E-7
	0.1125	4.6146	3.7925	287.7000		

TABLE 3A

Canada Total Unemployed Male - Aged 25 and Over. 95% Confidence Intervals for (Yt-st)-(Yt-1-st-1) Jul. 81 Aug. 81 Sep. 81 Oct. 81 Nov. 81 Dec. 81 date May 81 Jun. 81 Mary 81 (9.28, 19.66) Jun. 81 (10.84, 21.14) (1.06, 11.40) Jul.81 (13.35,23.00) (1.19,10.91) (-0.63, 9.15) Aug. 81 (12.00, 22.29) (0.18, 10.41) (-0.24, 10.04) (-14.65, -4.31) Sep. 81 (10.39, 21.54) (2.79, 13.95) (-0.99, 10.18) (-17.45, -6.22) (25.78, 37.07) Oct.81 (10.58,22.18) (2.89,14.48) (-1.16,10.43) (-18.62,-7.01) (23.80,35.46) (0.48,12.22) Nov. 81 (10.02, 22.48) (3.56, 16.02) (-3.16, 9.30) (-18.61, -6.15) (22.50, 34.98) (3.00, 15.52) (-18.43, -5.83)Dec.81 (9.50,22.14) (4.12,16.75) (-2.54,10.09) (-18.58,-5.95) (24.04,36.67) (1.24,13.89) (-17.74,-5.05) (40.19,52.95) May 82 (9.21, 20.38) (6.16, 17.50) (-1.23, 10.14) (-16.35, -4.98) (25.85, 37.22) (3.82, 15.19) (-15.38, -4.02) (42.72, 54.06) Dec 85 (11.00,22.85) (-2.87, 8.97) (1.83,13.67) (-16.25,-4.40) (30.01,41.87) (-3.77, 8.11) (-20.26,-8.36) (42.20,54.08)

TABLE 3B

Canada Total Unemployed Male - Aged 25 and Over. 95% Confidence Intervals for $(Y_t-s_t)-(Y_{t-2}-s_{t-2})$

date	Mar. to May 81	Apr. to Jun. 81	May. to Jul. 81	Jun. to Aug. 81	Jul. to Sep. 81	Aug. to Oct. 81	Sep. to Nov. 81	Oct. to Dec. 81
Mey 81	(-2.01, 9.38)		2					
Jun. 81	(-0.76.10.56)	(16, 51, 27, 93)						
Jul.81	(1.42, 12.57)	(18.65,29.80)	(4.68, 15.94)					
Aug. 81	(0.23, 11.59)	(16.82,28.15)	(4.52,15.86)	(-10.32, 1.15)				
Sep.81	(-0.42,11.39)	(18.44,30.23)	(7.08,18.85)	(-13.14, -1.34)	(13.62,25.55)			
Oct.81	(0.47, 12.76)	(18.94,31.20)	(7.21.19.44)	(-14.29, -2.07)	(10.68, 22.94)	(29.79.42.17)		
Nov. 81	(-0.33, 12.38)	(19.70, 32.37)	(6.53, 19.18)	(-15.63, -2.99)	(10.03.22.68)	(31.66.44.34)	(- 9.25, 3.52)	
Dec. 81	(1.20.14.05)	(19.84.32.66)	(7.81.20.61)	(-14.882.09)	(11.70.24.49)	(31.52.44.33)	(-10.25, 2.59)	(28,71.41.64
May 82	(-6.47, 5.85)	(20.45, 32, 77)	(10.04,22.49)	(-12.51, 0.09)	(14.57.27.17)	(34.74.47.34)	(- 6.49, 6.11)	(32.40.44.98
Dec 85	(3.12,17.55)	(12.78,27.17)	(3.62,17.98)	(- 9.76, 4.60)	(18.43, 32.79)	(30.91,45.30)	(-19.35, -4.92)	(26.59.41.08



TABLE 4

Summary Statistics on the Relative Differences.

Statistics	CA-U	M(1) US-EM(2)	
N	132	132	
Mean	00017	0000143	
Std. Dev. (2)	.00802	.00053	
T-ratio	24446	31068	
Prob> T	.80726	.75625	
D:Normal(3)	.07380	.04567	
Prob>D	.078	>.15	
Min	01757	00122	
Max	.01921	.00127	
SSQ(4)	.00843	.0000367	

(1): Canada Total of Unemployed Male Aged 25 and Over

(2): Americain Nonagricultural Total Employed Male Aged 20 and Over

(3): Kolmogorov-Smirnov test for normality assumption.(4): Sum of Square of the Relative Differences.



1

.



- 22 -FIGURES 1B CANADA TOTAL OF UNEMPLOYED MALE - AGED 25 AND OVER MONIH TO MONIH CHANGES OF SEASONALLY ADJUSTED VALUES (T AND T-1)





FIGURE 1B.2: CHANGE IN X-11-ARIMA SEASONALLY ADJUSTED: 95% FREDICTIVE INTERVAL



FIGURE 1B.3: MSE OF THE SMOOTHED UCM ESTIMATES VERSUS TIME

FIGURE 1B.1

•

- 23 -FIGURES 1C CANADA TOTAL OF UNEMPLOYED MALE - AGED 25 AND OVER MONTH TO MONTH CHANGES OF SEASONALLY ADJUSTED VALUES (T AND T-2)



1





FIGURE 1C.2 CHANGE IN X-11-ARIMA SEASONALLY ADJUSTED ...: 95% PREDICTIVE INTERVAL



FIGURE 1C.3 ____: MSE OF THE SMOOTHED UCM ESTIMATES VERSUS TIME





FIGURE 2B.3 ____: MSE OF THE SMOOTHED UCM ESTIMATES VERSUS TIME

JANTE.

34879

11100

AND D

DATE

JANES

JANES.

JAIRA

JANES

JANES

34876

10077

3

- 26 -FIGURES 2C AMERICAIN TOTAL OF EMPLOYED MALE - AGED 20 AND OVER MONTH TO MONTH RATIOS OF SEASONALLY ADJUSTED VALUES (T AND T-2)





FIGURE 2C.1 ...: RATIO IN X-11-ARIMA SEASONALLY ADJUSTED ...: RATIO IN SMOOTHED UCM





FIGURE 2C.3 —: MSE OF THE SMOOTHED UCM ESTIMATES VERSUS TIME

•

		DAT	E DUE	
	ta			
		1		
1ª				



Ca 005

74839 c.2

4

1

,