Methodology Branch

Time Series Research and Analysis Division

Direction de la méthodologie

Division de la recherche et de l'analyse des chroniques

Statistics Canada Statistique Canada

Canada

[redacted]

WORKING PAPER NO. METH-93-005
TIME SERIES RESEARCH & ANALYSIS DIVISION
METHODOLOGY

ESTIMATION OF THE VARIANCES OF X-11-ARIMA SEASONALLY ADJUSTED ESTIMATORS FOR A MULTIPLICATIVE DECOMPOSITION AND HETEROSCEDASTIC VARIANCES

by

Danny Pfeffermann, Marietta Morry and Paul Wong

ABSTRACT

This paper provides an extension to the method of estimating the variances of the X11ARIMA seasonally adjusted estimators under the assumption of linearity described in Pfeffermann (1993). These extensions include the case where the variance and the autocovariances of the survey errors are time dependent, the case of multiplicative decomposition and the case of variance estimation in the presence of the additional non-linear operations of the X11ARIMA method such as the identification and replacement of extreme values and the identification and estimation of the X11ARIMA models used for extrapolation.

RÉSUMÉ

Cet article porte sur la variance des données désaisonnalisées par la méthode X-11-ARIMA. Nous généralisons la méthode de Pfeffermann (1993). Cette méthode fait l'hypothèse que le X-11-ARIMA fait une décomposition additive et qu'il n y a pas d'opération non-linéaire lors de la désaisonnalisation.

La méthode généralisée contient les cas pour lesquels la méthode X-11-ARIMA utilise soit une décomposition multiplicative, soit des opérations non-linéaires, ou bien soit des données provenant d'un sondage avec échantillonnage aléatoire. Les opératiens non-linéaires du X-11-ARIMA sont l'identification et l'estimation du modèle d'extrapolation ARIMA, et l'identification et l'ajustement des valeurs extrêmes. La variance des erreurs d'échantillonnage et leurs autocorrélations peuvent dépendre du temps dans les cas où les données non-désaisonnalisées proviennent d'un sondage.

1. INTRODUCTION

In a recent article, Danny Pfeffermann (1993), abbreviated hereafter as DP, proposes a new method for estimating the variances of X-11 ARIMA seasonally adjusted estimators (SAE). The method consists of the following stages

1) Decompose the observed time series $\{y_t, t=1 \cdots N\}$ as

$$y_t = (T_t + S_t + I_t) + \epsilon_t = T_t + S_t + e_t \tag{1.1}$$

where T_t and S_t are correspondingly the trend level and the seasonal effect, I_t is the irregular component and ϵ_t is the survey error when the population mean value $Y_t = (T_t + S_t + I_t)$ is estimated from a survey. Most business and economic time series analyzed routinely consist of survey estimates.

2) Express the variances of the SAE as functions of the variance and autocovariances of the compound error terms, $e_t = I_t + \epsilon_t$.

3) Express the variance and autocovariances of the series $\{e_t\}$ as linear combinations of the variance and autocovariances of the X-11 residual series, $\{R_t = (y_t - \hat{T}_t - \hat{S}_t)\}$.

4) Estimate the variance and autocovariances of the series $\{e_t\}$ by replacing in 3 the theoretical variance and autocovariances of the series $\{R_t\}$ by standard sample estimates.

5) Substitute the estimates obtained in 4 in the expressions derived in 2 to obtain the estimates of the variances of the SAE.

The actual application of the method involves the specification of a cutoff value C such that for $k > C$, $V_k = \text{COV}(e_t, e_{t-k}) \cong 0$. Such a specification is needed in order to limit the number of equations in 3 and hence secure the stability of the estimates. Plausible values of C can often be determined from knowledge of the sampling design used to collect the data. Alternatively, a stepwise algorithm described in DP can be used. For a given cutoff C, the equations in stages 2 and 3 are fixed for any given linear filters of X-11 ARIMA so that the application of the method is straightforward.

The objectives of the present article are as follows:

A- Extend the method to the case of a multiplicative decomposition - Many of the time series deseasonalized in practice are decomposed as products rather than as sums of the trend, seasonal and error components so that

$$y_t = T_t^* \times S_t^* \times I_t^* \times \epsilon_t^* = T_t^* \times S_t^* \times e_t^* \tag{1.2}$$

where S_t^*, I_t^* and e_t^* are percentage measurements. The multiplicative decomposition implies that the seasonal effect $S_t = (y_t - T_t^* \times e_t^*)$ is proportional to the trend level which is often more realistic.

B- Extend the method to the case where the variance and autocovariances of the survey errors are time dependent - A common feature to many of the time series collected from surveys is that the variance and autocovariances of the survey errors, ϵ_t, (often the dominant component of the compound error terms, e_t) change over time. Such changes may occur as a result of changes in the sampling design, and in particular from increasing or decreasing the sample size, or because the variances change with the level of the series. The latter case is implicit in the multiplicative decomposition since

$\epsilon_t = y_t - Y_t = T_t^* \times S_t^* \times I_t^* \; (\epsilon_t^* - 1)$. Estimators for the variances of the survey errors are routinely produced by statistical agencies but they are seldom used in the analysis of time series.

C- Study the effect of the nonlinear options of X-11 ARIMA on the variances and variance estimators of the SAE - The method described in DP assumes that the SAE are derived using the linear filters of X-11 ARIMA. As well known, the X-11 ARIMA program contains also some nonlinear options, beyond the use of the multiplicative mode. These options include the identification and gradual replacement of extreme values and the identification and estimation of ARIMA models used for the extrapolation of data at the beginning and the end of the series. With this option the extrapolated data are added to the observed series so that the X-11 procedure is applied to the augmented series.

The two extensions to the method and the effects of the nonlinear options of X-11 ARIMA are studied separately in the next four sections.

2. ESTIMATION OF THE VARIANCES OF THE SAE FOR A MULTIPLICATIVE DECOMPOSITION

The multiplicative decomposition is defined by equation (1.2). The use of the multiplicative mode of X-11 ARIMA is known to yield very similar results to the use of a "log-additive decomposition" (LAD) by which the additive mode is applied to the logarithms $\tilde{y}_t = \log(y_t)$ of the series. The SAE for the original series are computed under this mode as $\hat{N}_t^* = (y_t / \hat{S}_t^*) = [y_t / \exp(\hat{s}_t)]$ where $s_t = \log(S_t^*)$. (We applied the two decomposition methods to the series "Employed Women in Nova Scotia, 1980-1989" which is deseasonalysed routinely using a multiplicative decomposition. The estimates obtained for the seasonal component and the residual terms by the two methods are almost identical. We use this series for illustrating the estimation of the variances of the SAE.)

Let $\hat{n}_t = (\tilde{y}_t - \hat{s}_t) = \sum_{k=-(t-1)}^{N-t} w_{kt}\,\tilde{y}_{t+k}$ define the log additive SAE obtained by application of the LAD. DP considers two alternative variance estimators depending on whether the SAE are used to estimate the seasonally adjusted values in the population or the trend levels. It what follows we consider for convenience the second case and denote $d_{2t} = [\hat{n}_t - \ell_t]$ where $\ell_t = \log(T_t^*)$. It follows from DP that under broad conditions $d_{2t} \equiv \sum_{K=-(t-1)}^{N-t} w_{Kt}\,\bar{e}_{t+K}$ where $\bar{e}_t = \log(e_t^*)$ is the compound error term in the additive decomposition model of $\tilde{y}_t$. See also Breidt (1992).

The method we propose for estimating the variances of the SAE for a multiplicative decomposition rests on the assumption that d_{2t} has a normal distribution with zero mean. This is a rather mild assumption considering that the time series analysed in practice are usually composed of means or aggregates computed from large surveys and that extreme values are ordinarily modified outside or by the X-11 procedure. (Each error term $\bar{e}_t$ is a linear combination of the observations $\tilde{y}_t$). The method consists of the following stages:

a) Apply the LAD mode of X-11 ARIMA

b) Apply the method of DP to estimate the variance and autocovariances of the compound errors $\{\bar{e}_t\}$ and hence the variances $Var^{(2)}(\hat{n}_t) = Var_C(d_{2t})$ where $Var_c(.)$ is the variance with respect to the joint distribution of the compound error terms

c) Estimate $Var^{(2)}(\hat{N}_t^*) = Var^{(2)}[(y_t/\hat{S}_t^*)/T_t^*]$ as

$$\hat{Var}^{(2)}(\hat{N}_t^*) = \{\exp[2\times\hat{Var}^{(2)}(\hat{n}_t)] - \exp[\hat{Var}^{(2)}(\hat{n}_t)]\} \qquad (2.1)$$

utilizing the relationship between the variance of the normal and lognormal distributions.

In order to illustrate the performance of the proposed method we carried out the following experiment. We applied the LAD to the series "Employed Women in Nova Scotia" and estimated the variance and autocorrelations of the corresponding error terms $\{\tilde{e}_t\}$ using the method of DP. The estimates obtained are $\hat{V}(\tilde{e}_t) = 1.9\times10^{-4}$; $\hat{\rho}_1 = .642, \hat{\rho}_2 = .421, \hat{\rho}_3 = .116, \hat{\rho}_k \cong 0$ for $k \geq 4$. Next we generated independently 300 random series $\{\tilde{e}_{t,j};\ t=1 \cdots 120\},\ j=1 \cdots 300$ from an MA(3) process defined by the estimated variance and autocorrelations. These random series had been added to the signals $\{\hat{m}_t = (\hat{\ell}_t + \hat{s}_t)\}$ estimated for the empirical series $\{\tilde{y}_t\}$. Let $y_{t,j} = \exp(\hat{m}_t + e_{t,j}),\ t=1 \cdots 120,\ j=1 \cdots 300$. We applied the multiplicative mode of X-11 to each of the series $\{y_{t,j}\}$ yielding the SAE $\hat{N}^*_{t,j} = (y_{t,j}/\hat{S}^*_{t,j})$ and hence the empirical standard deviations (SD), $\bar{SD}_t(\hat{N}^*) = [\sum_{j=1}^{300} (\hat{N}^*_{t,j} - \bar{N}^*_t)^2/300]^{½}$ where $\bar{N}^*_t = [\sum_{j=1}^{300} (\hat{N}^*_{t,j}/300]$. In the final stage we applied the method of variance estimation proposed in this section to each of the series $y_{t,j}$.

Figure 1 shows for each month $t=1 \cdots 120$ the empirical standard deviation $\bar{SD}_t(\hat{N}^*)$ and the mean $\bar{SD}_t(\hat{N}^*) = [\sum_{j=1}^{300} \hat{SD}_{t,j}/300]$ of the SD estimators $\hat{SD}_{t,j} = \exp(\hat{\ell}_t)[\hat{Var}^{(2)}(\hat{N}^*_{t,j})]^{½}$ where $\hat{Var}^{(2)}(\hat{N}^*_{t,j})$ is defined in (2.1). The picture revealed from the graph is that the estimators derived by application of the method are essentially unbiased. In fact, $[\sum_{t=1}^{120} \bar{SD}_t(\hat{N}^*)/120] = 1.80$ and $[\sum_{t=1}^{120} \bar{SD}_t(\hat{N}^*)/120] = 1.78$. Notice that both the empirical SD and the SD estimators increase over time as a result of the upgrowing trend levels $\{\hat{\ell}_t\}$.

3. ACCOUNTING FOR CHANGES IN THE DESIGN VARIANCES AND AUTOCOVARIANCES

As mentioned in the introduction, changes in the variances and autocovariances of the survey errors over time may be caused by changes in the sampling design or because the variance and autocovariances change with the level of the series. In what follows we assume the additive decomposition defined in (1.1) and the availability of monthly estimates for the design variances of the survey errors. Such estimates are routinely produced by statistical offices. Let $\hat{V}_0, \hat{V}_1, \ldots$ denote the estimated variance and autocovariances of the compound error terms as obtained by application of the method proposed in DP. Let $\hat{S}^2_{Dt}$ denote the design variance estimates and define $\hat{S}^2_D = [\sum_{t=1}^{N} \hat{S}^2_{Dt} / N]$ where N is the length of the series.

For the case where the changes in the design variances and autocovariances are caused by changes in the sampling design, we propose the following modifications to the estimates $\hat{V}_K$, $k=0,1,\ldots$ and hence to the estimates of the variances of the SAE. (As described in the introduction, the variances of the SAE can be approximated by functions of the variance and autocovariances of the compound error terms).

$$\tilde{V}_t = (\hat{V}_0 - \hat{S}^2_D) + \hat{S}^2_{Dt};\ \tilde{V}_{kt} = C\tilde{O}V(e_t, e_{t-k}) = (\tilde{V}_t \tilde{V}_{t-k})^{\frac{1}{2}} (\hat{V}_k / \hat{V}_0) \qquad (3.1)$$

The rationale for $\tilde{V}_t$ is that one can view $\hat{V}_0$ as an average variance in the sense that $\hat{V}_0 \cong V\hat{a}r_C(I_t + \epsilon_t) \cong \{V\hat{a}r_C(I_t) + [\sum_{t=1}^{N} V\hat{a}r_C(\epsilon_t)/N]\} \cong [V\hat{a}r_C(I_t) + \hat{S}^2_D]$ so that the proposed modification follows naturally. Notice also that $\sum_{t=1}^{N} (\tilde{V}_{0t}/N) = \hat{V}_0$. The modification to the covariances is somewhat more arbitrary since estimates for the design covariances are seldom available. It satisfies,
$c\tilde{o}rr(e_t, e_{t-k}) = (\hat{V}_k / \hat{V}_o) = c\hat{o}rr(e_t, e_{t-k})$, implying that the changes in the design do not affect the autocorrelations between the compound error terms which is a sensible assumption.

For the case where the design variance and autocovariances change with the level of the series, (evidently the more common case), we propose the following modifications

$$\tilde{V}_t = \hat{V}_0 \hat{S}^2_{Dt} / \hat{S}^2_D ; \tilde{V}_{kt} = (\tilde{V}_t \tilde{V}_{t-k})^{½} (\hat{V}_k / \hat{V}_0) = (\hat{S}^2_{Dt} \hat{S}^2_{D(t-k)})^{½} \hat{V}_k / \hat{S}^2_D \qquad (3.2)$$

The rationale for the modifications in this case is that if the variances of the survey errors change with the level of the series, then it is reasonable to postulate the same property for the variances of the error terms $\{ I_t \}$ in the decomposition model holding for the population values. Thus, both variances are modified by the ratio $\hat{S}^2_{Dt} / \hat{S}^2_D$. Notice that again $\sum_{t=1}^{N} (\tilde{V}_t / N) = \hat{V}_0$. The rationale for the modification of the covariances is as in the previous case, namely, that the autocorrelations between the compound error terms are constant over time despite the changes in the variances.

We applied the modifications (3.2) to the series "Unemployment Percentage Rates in Canada, 1982-1989" for which estimates of the design variances are routinely computed. (The original design variance estimates have been smoothed in order to diminish the effect of sampling variations. The differences between the original and the smoothed variances never exceed 6 percent and they are in most cases much smaller.) Figure 2 shows the monthly design variances (multiplied by 1000), along with the seasonal effects of the original series as estimated by application of the additive mode of X-11 ARIMA. As can be seen, the design variances exhibit a seasonal pattern which is close to the seasonal pattern of the original series, implying that the magnitude of the variances indeed depends on the level of the series. (The seasonal effects explain 84 percent of the month to month variation of the original series.)

In order to illustrate the performance of the modifications (3.2), we carried out a similar experiment to the experiment described in Section 2. Thus, we first estimated the variance and autocorrelations of the compound error terms of the original series using the method of DP. The estimates obtained are: $\hat{V}_0 = .0258$, $\hat{\rho}_1 = .33$, $\hat{\rho}_2 = .07$, $\hat{\rho}_3 = .28$, $\hat{\rho}_k \cong 0$ for $k \geq 4$. Next we generated independently 300 random series $\{e^u_{t,j};\ t=1 \cdots 96\}$, $j=1 \cdots 300$ from an MA(3) process defined by this variance and autocorrelations. These random series had been modified as $e_{t,j} = [(\hat{S}^2_{Dt} / \hat{S}^2_D)^{\frac{1}{2}}\ e^u_{t,j}]$ so that the variance and autocovariances of the modified series are as in (3.2). The modified series were added to the signals $\hat{M}_t = (\hat{T}_t + \hat{S}_t)$ estimated for the original series. In the final stage we applied the additive mode of X-11 to each of the series $\{y_{t,j} = \hat{M}_t + e_{t,j},\ t=1 \cdots 96\}$, $j=1 \cdots 300$ and estimated the variances of the SAE by first estimating the variance and autocovariances using the method proposed in DP and then applying the modifications (3.2).

Figure 3 shows for each month $t=1 \cdots 96$ the empirical SD of the SAE over the 300 series and the means of the SAE SD estimators as obtained when ignoring the changes in the design variances, (*i.e.* using the estimators $\hat{V}_k$, $k=0, 1, \ldots$ for the variance and autocovariances of the compound error terms), and when applying the modifications in 3.2. The picture revealed from the three graphs is that both sets of estimators are essentially unbiased in the sense that the averages of the SD estimators over the 96 months are practically the same as the corresponding average of the empirical SD. (The value obtained for the three averages is 0.14). However, the SD estimators obtained by application of the modifications in (3.2) follow much closer the empirical SD. Notice from Figure 2 that the design SD are around .04 implying that for this series the SD of the survey errors are much smaller than the SD of the SAE, indicating large variances of the irregular terms $\{I_t\}$.

4. THE EFFECTS OF IDENTIFICATION AND REPLACEMENT OF EXTREME VALUES

The procedure used by the additive mode of X-11 for the identification and replacement of extreme values is to compute the standard deviation σ of the estimated error terms in moving sections of 5 years and then compare the estimated errors in the central year of each section with the corresponding value of σ. For the first (last) 2 years, the σ value obtained for the first (last) 5 years are used for the comparisons. Error terms with absolute values larger than $2.5 \times \sigma$, referred to hereafter as *outliers*, are assigned a zero weight. Absolute errors smaller than $1.5 \times \sigma$ are assigned full weight whereas absolute errors between $1.5 \times \sigma$ and $2.5 \times \sigma$, referred to hereafter as *extreme* errors, are assigned a linearly graduating weight between zero and one. The weights are used to modify the estimates $\hat{Se}_{t^*} = (\hat{S}_{t^*} + \hat{e}_{t^*})$ corresponding to months t^* for which the error terms receive less than full weight. The modification consists of replacing $\hat{Se}_{t^*}$ by a weighted average of $\hat{Se}_{t^*}$ and the estimates $\hat{Se}_t$ obtained for the same calendar month in adjacent years. A similar procedure is applied under the multiplicative mode of X-11. See Dagum (1988) for details.

Outlier values are usually the outcome of unusual events like strikes, severe weather conditions, special government policies etc. As such, there is not much point in estimating the variances of the SAE corresponding to these values. The interesting question, however, is whether the existence of values with weights less than 1 and the replacement of these values, not accounted for by the linear filters of X-11, has a major effect on the variances of the SAE and the variance estimators in months not identified as having outlier errors.

In order to assess the impact of the outlier and extreme values on the variances of the SAE, we generated again 300 series of random error terms $\{e^{n}_{t,j}; t=1 \cdots 120\}$, $j=1 \cdots 300$ with variance and covariances as estimated for the series "Total Unemployment in Canada, 1980-1989". We chose this series because it contains a relatively large number of outlier and extreme values. Months with outlier values (9 months altogether, see Figure 5) have been excluded from the computation of the empirical variance and autocovariances of the estimated X-11 error terms which are used for the implementation of the method of DP. (See Stage 4 of the description of the method in the introduction). The variance and autocorrelations of the compound error terms were estimated as:

$\hat{V}_0 = 988.10$, $\hat{\rho}_1 = .82$, $\hat{\rho}_2 = .61$, $\hat{\rho}_3 = .34$, $\hat{\rho}_4 = .13$, $\hat{\rho}_k \cong 0$ for $k \geq 5$. The series $\{e^n_{t,j}]$ were added to the signals $\hat{M}_t = (\hat{T}_t + \hat{S}_t)$ estimated for the original series to form 300 new series $\{y^n_{t,j}, t=1 \cdots 120\}$, $j=1 \cdots 300$. These series are essentially free from outlier values.

In the next stage we modified the series $\{e^n_{t,j}\}$ by replacing errors generated for the months t^* with outlier values in the original series, by error terms generated from a normal distribution with mean zero and variance $\sigma^2_{t^*} = (R^2_{t^*} \cdot \hat{V}_0 / \hat{U}_0)$ where $R^2_{t^*}$ is the outlier X-11 estimated error term and $\hat{U}_0$ is the empirical variance of the estimated error terms. The 300 modified series of random errors obtained this way were again added to the signals $\hat{M}_t$ to form 300 additional series $\{y^o_{t,j}, t=1 \cdots 120\}$, $j=1 \cdots 300$.

We applied the additive mode of X-11 but without the option of identification and replacement of extreme and outlier values to each of the 300 series $\{y^n_{t,j}\}$. We then applied the additive mode with that option to each of the 300 series $\{y^o_{t,j}\}$ and estimated the variance and autocovariances of the error terms and hence the variances of the SAE using the method of DP. The average estimates of the variance and autocorrelations of the error terms over the 300 series were found to be, $\bar{V}_0 = 946.5$, $\bar{\rho}_1 = .81$, $\bar{\rho}_2 = .59$, $\bar{\rho}_3 = .30$, $\bar{\rho}_4 = .09$, $\bar{\rho}_K \cong 0$ for $k \geq 5$ which shows a very close fit to the variance and autocorrelations used to generate the series $\{e^n_{t,j}\}$, despite the existence of outlier values and the use of the option of identification and gradual replacement of extreme and outlier values. (Outlier values were again excluded from the computation of the empirical variance and autocovariances of the X-11 estimated error terms).

Figure 4 shows for each month $t=1 \cdots 120$ the empirical SD $[\tilde{SD}^O_t(\hat{N})]$ of the SAE as obtained for the series, $\{y^n_{t,j}\}$, the empirical SD $[\tilde{SD}^O_t(\hat{N})]$ of the SAE as obtained for the series $\{y^{o}_{t,j}\}$ and the mean of the SD estimators $[\overline{SD}^{o}_{t}(\hat{N})]$ as obtained for the latter group of series. The peaks marked with an asterisk correspond to the 9 months with outlier values.

The two notable results revealed from the graphs are:

a- The existence of outlier and extreme error terms has only a small effect on the variances of the SAE, except in the months with outlier values. Notice that the differences in the empirical variances of the SAE between the two groups of series observed for the months $t = 106 - 113$ are not the result of the outlier errors but rather the result of a random accumulation of extreme values in and around these months. The outcome that the variances of the SAE obtained for the series $\{y^{n}_{t,j}\}$ (with no identification and replacement of extreme values) are smaller in these months than the variances obtained for the series $\{y^{o}_{t,j}\}$ is explained by the fact that the way X-11 replaces the extreme values se_t, it generally increases the absolute differences $|y_t - \hat{S}_t|$ and hence the variances of the SAE.

b- The SAE variance estimators are naturally closer to the empirical variances obtained for the series $\{y^{n}_{t,j}\}$. In view of the first result, however, the performance of the estimators is nonetheless satisfactory. The average value of the means of the variance estimators over the 111 months with no outlier values is 29.83, as compared to 29.78 for the series $\{y^{n}_{t,j}\}$ and 30.04 for the series $\{y^{o}_{t,j}\}$.

5. THE EFFECTS OF ARIMA EXTRAPOLATIONS ON THE VARIANCES OF THE SAE

With the ARIMA option of X-11 ARIMA, an ARIMA model is identified to the observed series and used for the extrapolation of one or two years of data at either end of the series. The

X-11 procedure is then applied to the augmented series. The use of this option is known to improve over the original census X-11 estimators in terms of point estimation, and the magnitude of the revisions to existing estimates as new data become available. See e.g. Pierce (1980), Dagum (1983) and Dagum and Laniel (1987) for discussions. DP concludes that the SAE produced by the use of this option are approximately c-unbiased where the c-expectation is taken over the distribution of the compound error terms.

The filter used by X-11 ARIMA for estimating the SAE for a given month t in the beginning or the end section of the series is a convolution of the original census X-11 filter and the ARIMA extrapolation filter. Since both filters are linear and asymmetric, the same holds for the convoluted filter. Thus, the "nonlinearily" of the ARIMA option results from the model identification and estimation stage and not from the use of the extrapolated data per se. The method proposed in DP for estimating the variances of the SAE applies to any linear c-unbiased estimator and hence it can be adapted to the use of the convoluted filters straightforwardly. The method ignores however the identification and estimation aspects of the use of the ARIMA option.

In order to assess the effects of the use of ARIMA extrapolations on the variances of the SAE, we applied the ARIMA option to each of the series $\{y^{\circ}_{t,j}\}$, generated for assessing the effects of extreme and outlier observations (see Section 4). We used the "automatic option" which selects the model that fits best the data out of five optional models. See Dagum (1988) for the criteria used for selecting the model. The results of this exercise are summarized in table 1 which shows the number of times each model has been selected and the average values of the parameter estimates. All the models listed in the table have the general form $(1-\phi_1 B-\phi_2 B^2)(1-B)^d(1-B^{12})\,y_t = (1-\theta_1 B-\theta_2 B^2)(1-\theta_{12}B^{12})\,\epsilon_t$, where B is the backshift operator, $By_t = y_{t-1}$.

Table 1: Frequency of Selected ARIMA Models and Average Parameter Estimates

		Average Parameter Estimates				
Models	Frequency	ϕ_1	ϕ_2	θ_1	θ_2	θ_{12}
$(0,1,1)(0,1,1)_{12}$	146	-	-	.028	-	.733
$(0,1,2)(0,1,1)_{12}$	101	-	-	-.028	-.350	.659
$(2,1,0)(0,1,1)_{12}$	32	.203	.330	-	-	.605
$(0,2,2)(0,1,1)_{12}$	7	-	-	.910	-.180	.670
$(2,1,2)(0,1,1)_{2}$	2	.303	-.176	.227	-.583	.682
No model selected	12	-	-	-	-	-

The models listed in Table 1 are different and yield different extrapolated values which is somewhat disturbing considering that the 300 simulated series obey the same stochastic structure. The convoluted filters resulting from the use of these models are, however, quite similar implying that the SAE and the variances of the SAE are less sensitive to the choice of the model and the parameters' estimators.

Figure 5 displays the empirical SD of the SAE with $[\tilde{SD}_t^A(\hat{N})]$ and without $[\tilde{SD}_t^O(\hat{N})]$ the use of the ARIMA option and the mean $[\bar{SD}_t^A(\hat{N})]$ of the SD estimators as obtained when accounting for the ARIMA extrapolations. When applying the ARIMA option, we added 12 months of extrapolated data at only the end of the series which is the common practice. It allows us also to compare the behaviour of the SAE variance estimators with and without the use of the ARIMA option (see below). Notice that the graph of the empirical SD without the use of the ARIMA option is the same as the graph of the empirical SD, $[\tilde{SD}_t^O(\hat{N})]$ in Figure 4, displaying the SD of the SAE obtained for the series $\{y_{t,j}^O\}$ with outlier observations.

The most striking result revealed from the graph is the decrease in the variances of the SAE towards the end of the series when applying the ARIMA option. The effect of the use of ARIMA extrapolations on the variances of the SAE has not been studied directly before although it could be inferred from other related studies. As can be seen, the SD estimators account for the use of the convoluted filters and they are lower therefore at the end of the series than at the beginning of the series where no ARIMA extrapolations have been used. We noted in Section 4 that the empirical SD obtained for the months $t = 106\text{-}113$ are affected by a random accumulation of extreme values. This fact is illuminated in Figure 5 showing for example that the empirical SD obtained for the month $t = 108$ when using the ARIMA option is higher than in the months $t = 114$ and $t = 118$ which is not the case with respect to the empirical SD obtained without the use of the ARIMA option. The averages of the empirical SD over the 111 months without outlier observations are 30.04 for the case of no the ARIMA extrapolations and 29.84 with the ARIMA extrapolations. The corresponding average of the SD estimators is 29.65.

ACKNOWLEDGEMENT

Work on this study was carried out while the first author was staying at Statistics Canada under its Research Fellowship Program. The authors are grateful to Judy Clarke for her technical assistance.

REFERENCES

Breidt, F.J. (1992) Variance of seasonally adjusted time series. *Proceedings of the Business and Economic Statistics Section, American Statistical Association.*

Dagum, E.G. (1983) Spectral properties of the concurrent and forecasting seasonal linear fitters of the X-11 ARIMA method. *The Canadian Journal of Statistics, 11, 73-90.*

Dagum, E.B. (1988) The X11ARIMA/88 seasonal adjustment method-foundations and user's manual. Time Series Research and Analysis Division, Statistics Canada, Ottawa K1A-0T6.

Dagum, E.B. and Laniel, N. (1987), Revisions of trend-cycle estimators of moving average seasonal adjustment methods. *Journal of Business and Economic Statistics*, 5, 177-189.

Pfeffermann, D. (1993) A general method for estimating the variances of X-11 seasonally adjusted estimators. *Journal of Time Series Analysis* (Forthcoming).

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

Figure 1: Empirical Standard Deviations [$\tilde{SD}_t(\hat{N}^*)$] and Means of Standard Deviation Estimators [$\bar{SD}_t(\hat{N}^*)$] Under a Multiplicative Decomposition

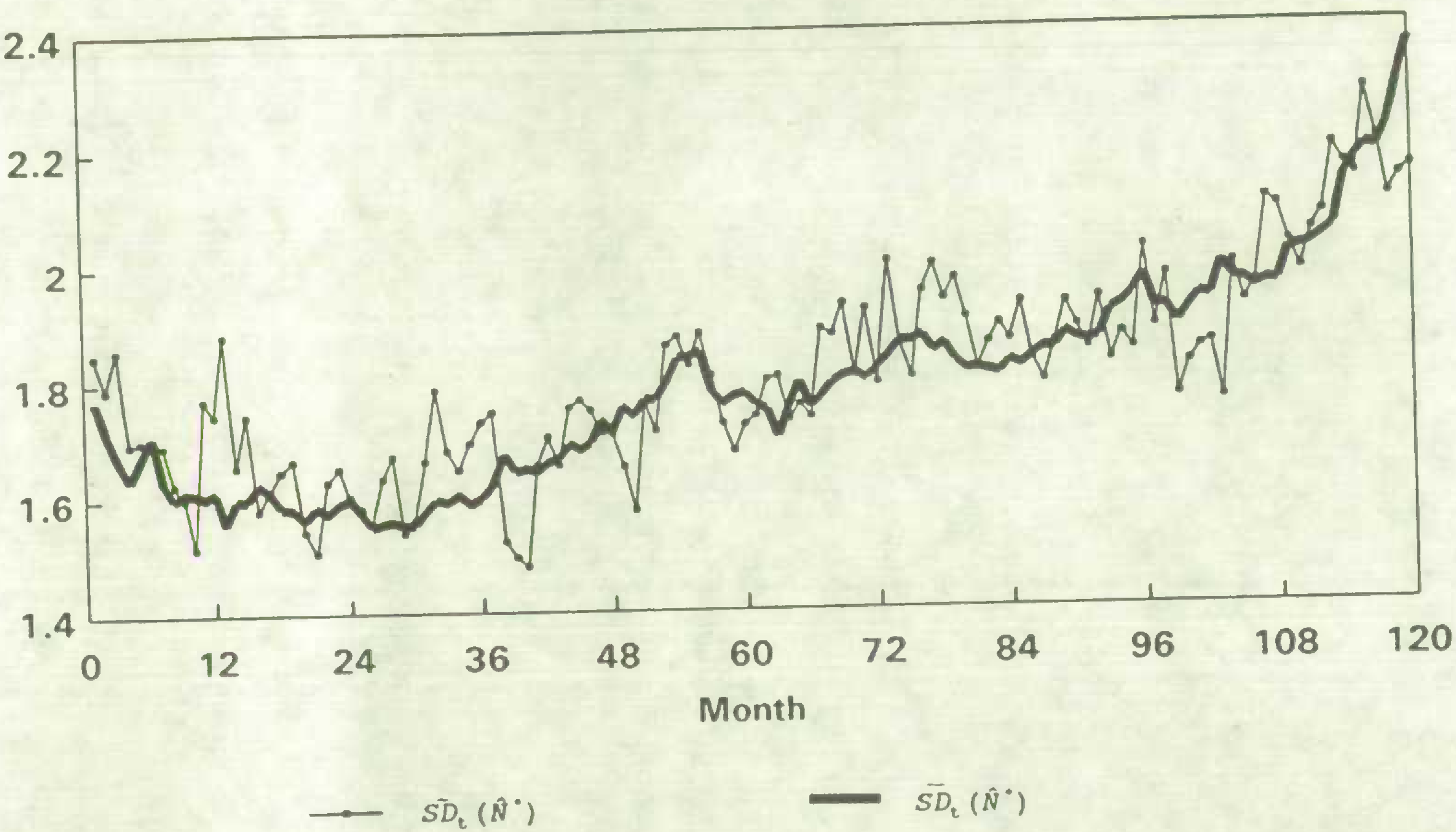


Figure 2: Estimates of the Seasonal Component and the Design Variances (× 1000).

Canada Unemployment Rates, 1982-1989

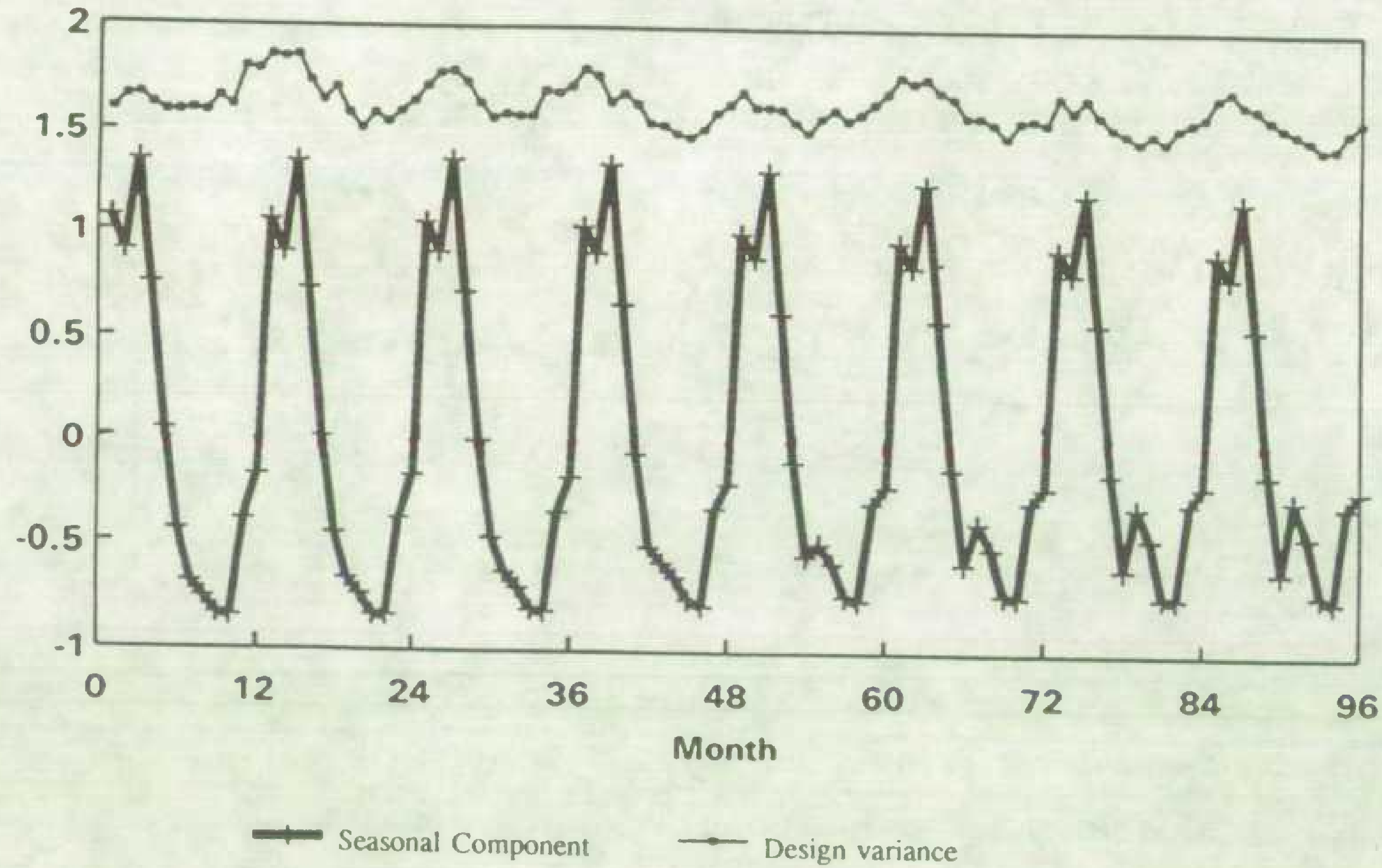


Figure 3: Empirical Standard Deviation [$\bar{SD}_t(\hat{N})$] and Means of Standard Deviation Estimators With [$\bar{SD}_t^w(\hat{N})$] and Without [$\bar{SD}_t^n(\hat{N})$] Accounting for Changes in the Design Variances

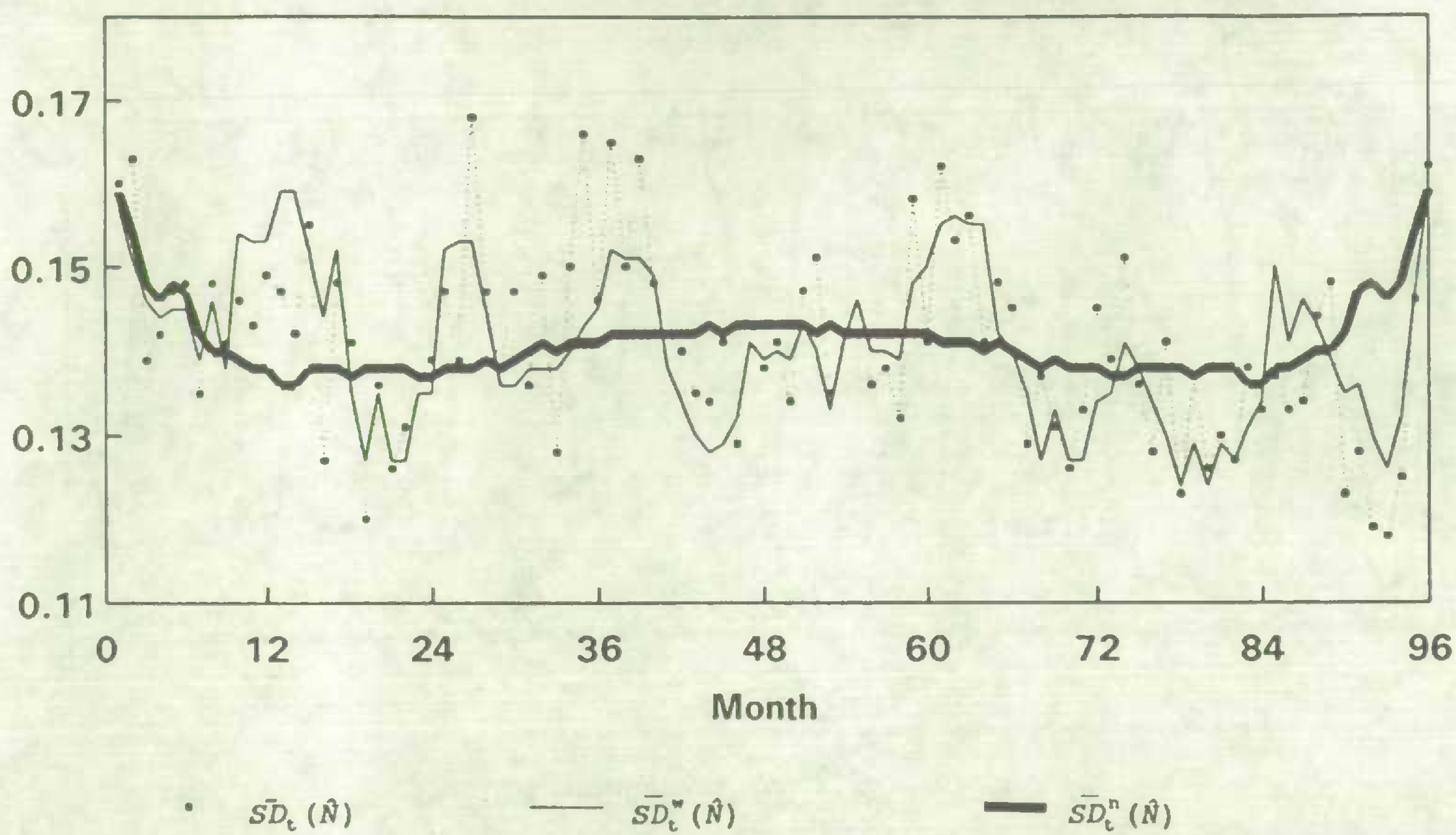


Figure 4: Empirical Standard Deviations With [$\bar{SD}_t^{o}(\hat{N})$] and Without [$\bar{SD}_t^{n}(\hat{N})$] the Identification and Replacement of Extreme and Outlier Values and Means of Standard Deviation Estimators $\bar{SD}_t^{o}(\hat{N})$

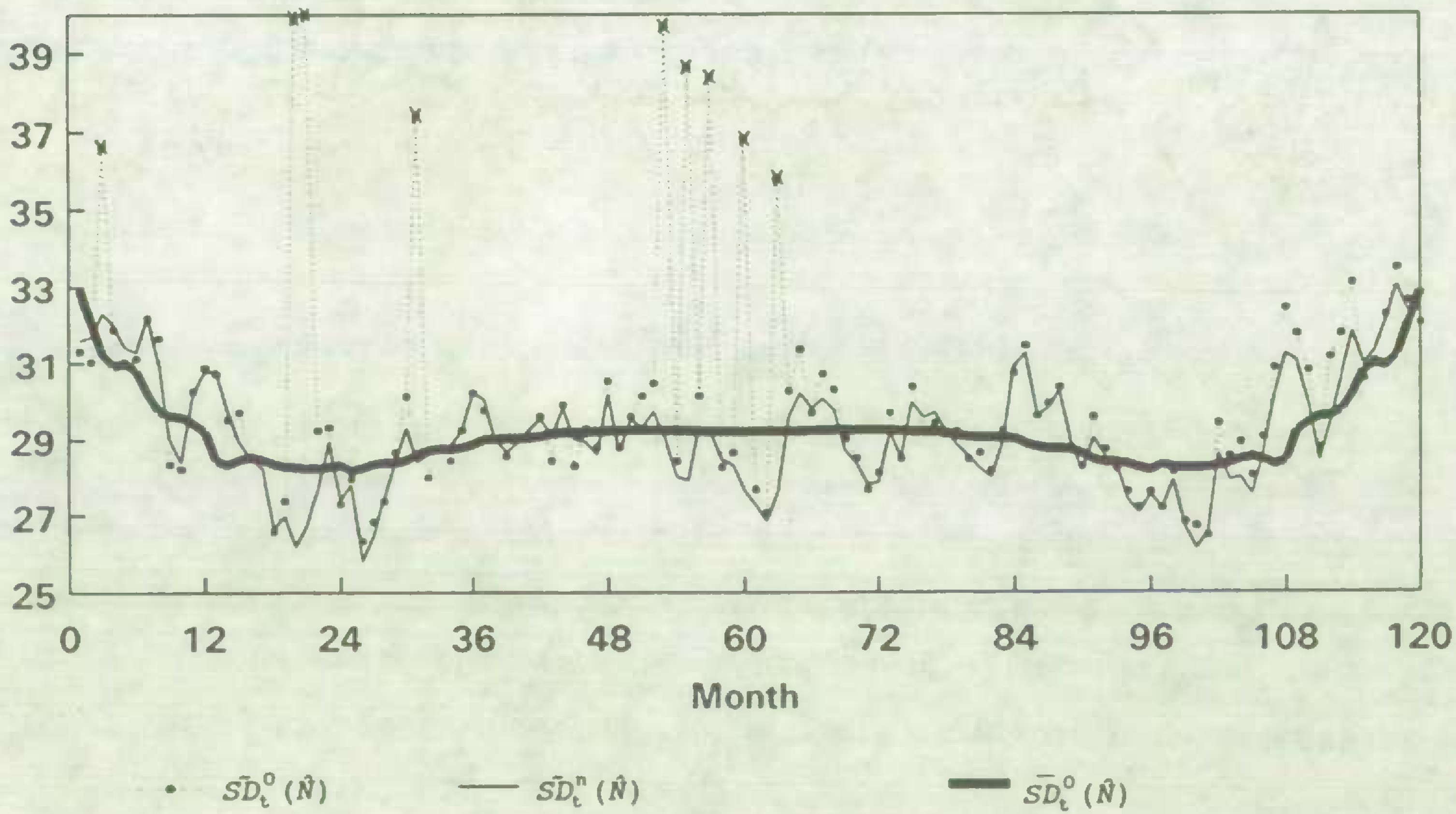
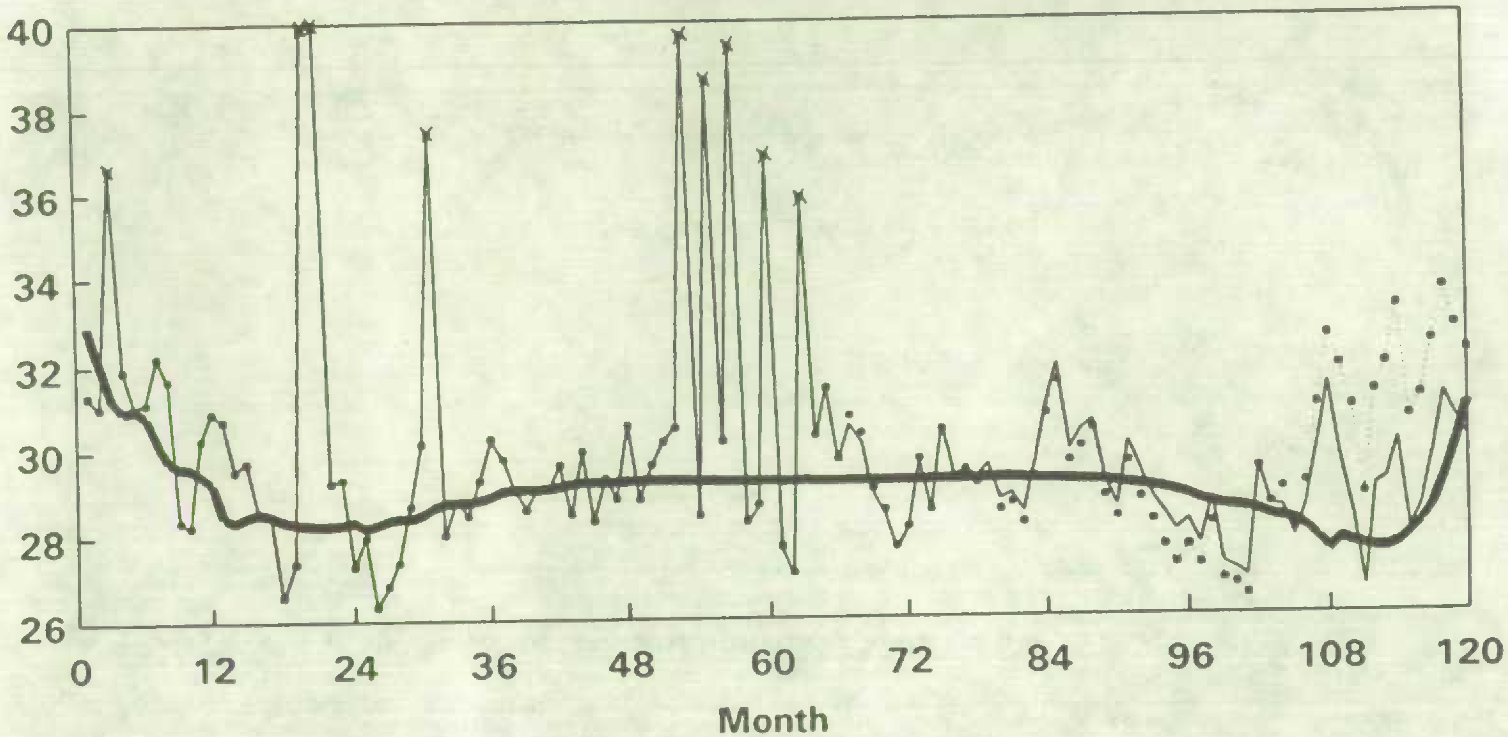


Figure 5: Empirical Standard Deviations With $[\bar{SD}_t^A(\hat{N})]$ and Without $[\bar{SD}_t^O(\hat{N})]$ the use of ARIMA Extrapolations and Means $[\bar{SD}_t^A(\hat{N})]$ of Standard Deviation Estimators When Accounting for ARIMA Extrapolations



$\bar{SD}_t^A(\hat{N})$ $\bar{SD}_t^O(\hat{N})$ $\bar{SD}_t^A(\hat{N})$