THE EFFECT OF DIFFERENT ROTATION PATTERNS ON THE SAMPLING VARIANCE OF SEASONAL AND TREND FILTERS

Craig H. McLaren and David G. Steel, University of Wollongong, NSW Australia
Craig H. McLaren, School of Mathematics and Applied Statistics,
University of Wollongong, NSW 2522, Australia (craig_mclaren@uow.edu.au)

Key Words: Henderson Moving Average, X-11, Seasonal Adjustment, Trend Estimation.

1. Introduction

Many important time series are based on repeated sample surveys which have complex designs and patterns of sample overlap from period to period. The use of sampling means that the estimated time series have a component of variability due to sampling errors and for many series this will be a major source of variability. The sample design, in particular the overlap pattern, affects the variability of the time series of the survey estimates.

Increasingly, analysis of time series is concentrating on assessing trends based on analysis of the seasonally adjusted series. The Australian Bureau of Statistics (ABS) has calculated seasonally adjusted series for many years. More recently, it has published trend series obtained by applying Henderson moving-averages to the seasonally adjusted series to improve the interpretion of variation in the original series (Linacre and Zarb, 1991). Since seasonally adjusted and trend estimates are obtained by various processes applied to the original data, they are also influenced by the sampling error. Some series are based on independent samples over time, but usually the samples used have a degree of overlap from period to period to reduce costs and the standard errors associated with the estimates of change between two adjacent time periods; for example monthly or quarterly changes.

A key issue in the development of the design of a repeated survey is the rotation pattern, that is, the pattern of a selected unit's inclusion in the survey over time, which will determine the sample overlap. The aim of this paper is to determine the effects of the rotation scheme used on the sampling error of the estimated seasonally adjusted and trend series.

This research was supported by the Australian Research Council and the Australian Bureau of Statistics. We would like to thank Andrew Sutcliffe and Phillip Bell for their advice during this project.

2. Rotation Patterns

Consider a univariate time series with values $y_t, t = 1, \dots, T$, obtained from a repeated sample survey. The observed value at time t is related to the true value of the series in the finite population, Y_t, by

$$y_t = Y_t + \epsilon_t + e_t$$

where ϵ_t is the effect due to non-sampling errors and e_t is the sampling error. The series Y_t is thought to consist of trend-cycle, seasonal and irregular components T_t, S_t and I_t, so that

$$y_t = T_t + S_t + I_t + \epsilon_t + e_t$$

In some cases multiplicative models may be more appropriate. Many statistical agencies produce seasonally adjusted series by attempting to estimate S_t and remove it from the series, usually using some combination of linear filters, as in the US Census Bureau's X-11 method (Shiskin *et al*, 1967). The ABS is also producing trend estimates using Henderson Moving Averages applied to the seasonally adjusted series and encouraging their use when interpreting time series (see ABS, 1987 and 1993).

The autocorrelation structure of the observed series is determined by the autocorrelation of the series Y_t, ϵ_t and e_t, which will then affect the estimation of the trend, seasonal and irregular components. The covariance structure of the sampling error series e_t can be estimated from the unit level survey data. By obtaining such estimates, it is possible to obtain the sampling variance of the estimated trend, seasonally adjusted and irregular series. Various methods for doing this have been proposed; for example Steel and DeMel (1988) consider the effect of linear filters on the spectrum of the sampling error series and Wolter and Monsour (1981) used an approach based on the effect of linear filters on the autocovariance function. Sutcliffe (1993) adopts a similar approach using an approximation to the X-11 procedure. Pfeffermann (1994) proposed a method which develops an estimate of sampling error directly from the estimated time series using various simplifying assumptions. These approaches do not explicitly model the time series. Other authors, for example Bell and Wilcox (1993), Burridge and Wallis (1985) and Hausman and Watson (1985), consider explicit ARIMA models for both the true series and the sampling error series, and concentrate on the estimation of the parameters of the models.

Typically, the repeated survey generating the time series does not use an independent sample at each time period but involves a significant degree of sample overlap. A rotation scheme is usually employed in which selected units are retained in the sample for several periods. The rotation pattern used will affect the autocorrelation structure of the sampling error series and hence the sampling variance of the original, seasonally adjusted and trend estimates.

Several considerations are taken into account in deciding upon a rotation scheme. High sample overlap between consecutive periods reduces the sampling variance of estimates of change between the periods and high sampling overlap between periods 12 months apart reduces the sampling variance of estimates of annual change. Also, the first interview is usually the most expensive and so by keeping selected dwellings in the survey for longer the cost of the survey is reduced. These factors by themselves would lead to having no rotation and using the same sample at each period and hence having complete sample overlap. There would have to be some updating of the sample to represent births in the population. This leads to rotation schemes in which a selected unit is included every period for as long as possible. However, a selected unit must eventually be rotated out of the survey. Besides the ethical consideration of spreading respondent load, there is the possible deterioration in response rate and quality of data reported if the same unit is included on several occasions.

Rotation patterns vary in terms of the number of times a dwelling is included in the survey and the time interval between inclusions. We will concentrate on monthly labour force surveys (MLFSs). Different countries have developed different rotation schemes. Balancing the considerations discussed above led Australia to include selected dwellings for 8 consecutive months and Canada to include them for 6 consecutive months. The United States' Current Population Survey (CPS) includes households for 4 months, then leaves them out for 8 months and re-includes them for a further 4 months. Japan has a MLFS which includes households for 2 months, then leaves them out for 10 months and re-includes them for a further 2 months. These last two designs give some sample overlap between the same months a year apart, which improves the standard error of the estimates of change between these months.

We will consider the following rotation schemes:

(a) Selected dwellings are included for m months after which they are removed from the sample. This pattern leads to a $1 - s/m$ overlap between samples s months apart, for $s = 1, \ldots, m-1$ and no overlap for months m or more months apart. Unless m exceeds 12 there will be no overlap for months a year apart. The case $m = 6$ corresponds to the Canadian rotation pattern and $m = 8$ corresponds to the Australian pattern. We will denote this pattern as *in-for-m*.

(b) Selected dwellings are in the survey for one month then removed for two months, then included again. This pattern is repeated until dwellings are included for a total of m times. This leads to no sample overlap between months one or two months apart, but an overlap of $1 - s/3m$ for $s = 3, 6, \ldots, 3m$. The overlap between months a year apart is $1 - 4/m$ provided m is 5 or more. The case of $m = 5$ roughly corresponds to the current British quarterly LFS regarded as a monthly survey. We will denote this pattern 1-2-1(m).

(c) Selected dwellings are included for 4 consecutive months, removed from the survey for 8 months and then included again for a further 4. This leads to an overlap of $1 - s/4$ for months s months apart, $s = 1, 2, 3$. For $s = 12$ the overlap factor is $4/8$. In fact the overlap is $4/8 - |s - 12|/8$ for $s = 9, \ldots, 15$.

There is no overlap for $s = 4, \ldots, 8$. This is the rotation scheme used in the US for the CPS. We will denote this pattern as 4-8-4(8).

(d) Selected dwellings are included for 2 consecutive months, removed from the survey for 10 months and then included again for a further 2. This leads to an overlap of 1/2 for adjacent months. For $s = 12$ the overlap factor is also 1/2. This is the rotation scheme used in Japan. We will denote this pattern as 2-10-2(4).

The United States and Japan rotation schemes are special cases in which selected dwellings are included for a consecutive months, removed from the survey for b months and then included again for a further a months. The pattern is repeated so that dwellings are included for a total of m occassions. These rotation schemes will be denoted a-b-$a(m)$. Some further examples of this class are considered below, which, as far as we know, have not be used in practice.

(e) 6-6-6(12). This leads to an overlap of $1 - s/6$ for months s months apart, $s = 1, \ldots, 5$. For $s = 12$ the overlap factor is 6/12. In fact the overlap is $6/12 - |s - 12|/12$ for $s = 7, \ldots, 17$.

(f) 1-1-1(6). This leads to no sample overlap between months one month apart. An overlap of $1 - s/12$ occurs for $s = 2, 4, \ldots, 10$.

3. Sampling Variance of Seasonally Adjusted and Trend Estimates

Let $\boldsymbol{y}_T$ be the vector containing the values of the time series of survey estimates up to time T. Consider a linear filter which is used to obtain values from $\boldsymbol{y}_T$ by applying a vector of weights $\boldsymbol{w}_t$, giving the filtered value at time t

$$\tilde{y}_t = \boldsymbol{w}_t' \boldsymbol{y}_T$$

Then

$$V(\tilde{y}_t) = \boldsymbol{w}_t' V(\boldsymbol{y}_T) \boldsymbol{w}_t$$

The seasonally adjusted and the final trend estimates produced by X-11 procedure can be approximated by linear filters. Sutcliffe (1993) showed that the standard X-11 package can be modelled by a series of matrix operations. Using this method, a set of linear filters which realistically approximate the X-11 process in the middle and at the ends of the series can be generated. We use these linear approximations corresponding to a 13 term Henderson Moving Average for estimation of trend, 3×5 moving average for estimation of seasonal factors and no modification for outliers.

To determine the sampling variance of a particular filtered series, we need an estimate of $V(\boldsymbol{y}_T)$ for different rotation patterns. Previous work on estimating variances of seasonally adjusted series has either ignored the rotation pattern or taken it as fixed and used an estimation method that takes it into account. We need a model for $V(\boldsymbol{y}_T)$ that reflects the effect of the different rotation schemes that could be used.

The analysis of the effect of sampling error is simplified if the series of sampling errors has a stable autocorrelation structure. The precise form of the autocorrelation function will depend on the series and the sample design. The autocorrelation function must reflect the complexities of the design. For example Steel and DeMel (1988) suggest a model for the Australian Monthly Labour Force data and Bell and Wilcox (1993) suggest a model for the United States Retail Trade series. Development and estimation of realistic autocorrelation functions for the sampling error series associated with important series taking into account the sample design used is an important and complicated issue.

Our approach is to assume that the series of sampling errors e_t has constant variance and that the correlation between estimates s periods apart for a two stage sampling design model is of the form given by Steel and DeMel (1988)

$$r(s) = \frac{\bar{n}\delta(0) + k(s)(\rho(s) - \delta(s))}{1 + (\bar{n} - 1)\delta(0)} \tag{1}$$

where $\bar{n}$ is the average number of dwellings selected per selected first stage unit, $k(s)$ is the overlap factor appropriate for the rotation scheme, and $\rho(s)$ is the correlation between the same dwellings at lag s. For some sample designs, there would be some correlation even when there is no houshold overlap. For example when dwellings are rotated out of the survey they may be replaced by dwellings in the same first stage sampling unit. The term in (1), $\delta(s)$ which is the correlation between different dwellings in the same first stage unit at lag s, reflects this.

The assumption that the variance of the sampling error series is constant implies that no major changes to the sample design or the population structure occur, at least over the effective length of the filters being considered. The assumption of constant autocorrelation, $\rho(s)$ and $\delta(s)$, for the population correlation also implies no major changes to the population.

Sutcliffe and Lee (1995) studied the standard errors of seasonally adjusted and trend estimates of level and movement under a small number of different rotation schemes. They assumed a simple pure geometric decay model for the correlations between survey estimates with a population correlation of $\rho = 0.8$, i.e. $r(s) = \rho^s$. The geometric decay model corresponds to an AR(1) process and decreases more rapidly than might be expected in practice.

The values for $\rho(s)$ and $\delta(s)$ in (1) were obtained from the Australian Labour Force Survey for the proportion of persons employed and also the proportion of persons unemployed.

4. Results

Table 1 summarises the effect of different rotation patterns by giving, for each of the two variables and a selection of rotation patterns, the ratio of the sampling variance of the estimates under consideration divided by the sampling variance that would be obtained when there is complete rotation each month. This is done for the level and one month difference for the seasonally adjusted and trend estimates at the very end of the series. Similar results apply to the middle of the series.

Figures 1 to 4 show the variance of the level and one month difference for the seasonally adjusted and trend estimates at the end of the series divided by the variance of the original estimate of level. These plots give results for the *in-for-m* rotation schemes for m going from 1 to 30 as well as the rotation patterns considered in the Table 1.

Figure 1 shows that increasing m in the *in-for-m* rotation patterns increases the variance of the seasonally adjusted level estimates, although the increase is small. Columns 1 and 2 in Table 1 shows that rotation patterns with low monthly overlap but good annual overlap such as the 2-10-2(4) and the 1-2-1(8) perform well for both variables.

However, when we consider the one month change in seasonally adjusted estimates, the benefit of having high monthly overlap becomes evident (see Figure 2 and columns 3 and 4). Those rotation patterns used in Canada and Australia perform well. The best option is no rotation but, as discussed in Section 2, this is not a practical option.

For the level of trend estimates at the end of the series the rotation schemes of 1-2-1(5) and 1-2-1(8) perform almost as well as an independent sample and considerably better than rotation schemes that involve monthly overlap (see Figure 3 and columns 5 and 6). This is primarily due to the fact that for a moving average, it is better to average over independent observations than positively correlated ones.

Figure 4 and columns 7 and 8 show that for one month changes in trend estimates the *in-for-4* rotation pattern seems to be the worst among those considered, and that the currently used rotation patterns can be significantly improved upon. For changes in the trend estimates the best performing rotation patterns are 1-2-1(5) and 1-2-1(8) which perform even better than the independent sample produced by complete rotation each month. This may seem a suprising result, but it arises because changes in trend estimates effectively look at differences in the seasonally adjusted series a few months apart and the 1-2-1(m) rotation patterns lead to postive correlations between estimates 3 months apart.

The rotation patterns currently used are sensible if the one month change in seasonally adjusted estimates are the key statistics to be analysed. However, we believe that the main purpose of looking at seasonally adjusted estimates is to assess the current trend in the series. If this is done using 13 point Henderson Moving Averages then the results here suggest that quite different rotation schemes should be used. Even if analysts do not formally use 13 point Henderson Moving Averages, the assessment of trend will often involve looking at changes in seasonally adjusted estimates a few months apart. The 1-2-1(m) rotation schemes will be suitable if the assessment of trends involve looking at changes in seasonally adjusted estimates over 3 months.

The evaluation criterion used in this paper has been the sampling variances of the trend and seasonally adjusted estimates, which are conditional on the values of true series, Y_t. Thus, only the sampling variance has been considered and the variance associated with Y_t has not be taken into account. However, it is the sampling variance that can be altered by sample design considerations. Wolter and Monsour (1981) discuss the issue of total variance versus sampling error variance. Another criterion to use to assess different rotation schemes is the degree of revisions at the end points, see for example, Dagum (1996) and Gray and Thomson (1997). Further work will involve developing more realistic correlation models, which reflect breaks in the sampling error series and removal of the non-stationary variance assumption. The total variance can be taken into account in our approach by developing correlation models that reflect the autocorrelation structure of Y_t and e_t. Given the complexity of the X-11 package, this work has used a set of linear weights which realistically approximate the widely used X-11 seasonal adjustment package. Other linear weights can be easily considered within this current framework. An example of such weights are weights derived to allow for correlated errors within a series.

5. References

Australian Bureau of Statistics (1987): *A guide to smoothing time series - estimates of "trend".* Australian Bureau of Statistics, Cat No 1316.0, Canberra.

Australian Bureau of Statistics (1993): *A guide to interpreting time series - monitoring "trends", an overview.* Australian Bureau of Statistics, Cat No 1348.0, Canberra.

Bell, W.R. and Wilcox, D.W. (1993): The effect of sampling error on the time series behavior of consumption data. *Journal of Econometrics*, 555, 235-265.

Bell, W.R. and Hillmer S. (1987): *Time series methods for survey estimation*, Bureau of the Census, RR-87/20

Burridge, P. and Wallis, K.F. (1985): Calculation of seasonally adjusted series. *Journal of the American Statistical Association*, 80, No 391, 541-552.

Dagum, E.B. (1996): A new method to reduce unwanted ripples and revisions in trend-cycle estimates from X11-ARIMA. *Survey Methodology*, 20, No. 1, 77-83

Gray, A. and Thomson, P. (1997): *Design of moving-average trend filters using fidelity, smoothness and minimum revisions criteria*, Bureau of the Census, RR96/01

Hausman, J.A. and Watson, M.W. (1985): Error in variables and seasonal adjustment procedures. *Journal of the American Statistical Association*, 80, No 391, 531-540.

Linacre, S. and Zarb, J. (1991): Picking turning points in the economy. *Australian Economic Indicators*, April 1991. Australian Bureau of Statistics, Cat No 1350.0.

Pfeffermann, D. (1994): A general method for estimating the variances of X-11 seasonally adjusted estimators. *Journal of Time Series Analysis*, 15, No 1, 85-116.

Shiskin, J., Young, A.H., and Musgrave, J.C. (1967). *The X-11 variant of the Census method II seasonal adjustment program.* Technical Paper 15, Bureau of the Census, U.S. Department of Commerce, Washington, D.C.

Steel. D.G. and DeMel, R. (1988): *The contribution of sampling error to the variability of statistical series.* Paper Presented at the National Mathematical Sciences Congress, Canberra, May 1988.

Sutcliffe, A. (1993): *X-11 Time series decomposition and sampling errors.* Working Papers in Econometrics and Applied Statistics, No 93/2. Australian Bureau of Statistics, Cat No 1351.

Sutcliffe, A. and Lee, G (1995). *Seasonal analysis and sample design.* Paper presented at the Conference of Survey Measurement and Process Quality. Bristol 1995.

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

	SA_t		$SA_{t+1} - SA_t$		$\hat{T}_t$		$\hat{T}_{t+1} - \hat{T}_t$	
	emp	unemp	emp	unemp	emp	unemp	emp	unemp
complete	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1-1-1(6)	1.01	1.01	1.00	1.00	1.07	1.06	0.78	0.86
in-for-8	1.04	1.03	0.33	0.50	1.26	1.25	1.03	1.09
in-for-6	1.03	1.03	0.36	0.52	1.25	1.25	1.14	1.15
6-6-6(12)	1.03	1.03	0.36	0.52	1.25	1.25	1.14	1.16
4-8-4(8)	1.03	1.03	0.42	0.57	1.22	1.22	1.28	1.24
2-10-2(4)	1.01	1.01	0.61	0.71	1.11	1.11	1.22	1.19
1-2-1(5)	1.00	1.00	0.99	1.00	1.01	1.01	0.77	0.84
1-2-1(8)	1.00	1.00	0.98	0.99	1.01	1.01	0.74	0.82
no rotation	1.05	1.04	0.25	0.44	1.29	1.29	0.81	0.94

Table 1: Ratio of the Variance at the ends of the series

Ai = 4-8-4(8), Bi = 2-10-2(4), Ci = 1-2-1(5), Di = 1-2-1(8), Hi = 1-1-1(6), Ji = 6-6-6(12),
—— (1) proportion of persons employed, - - - - (2) proportion of persons unemployed

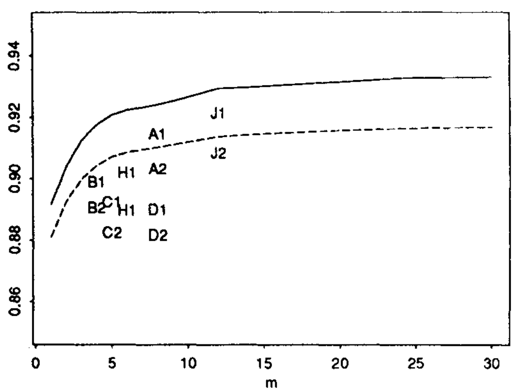

Figure 1: $V(SA_t)/V(y_t)$

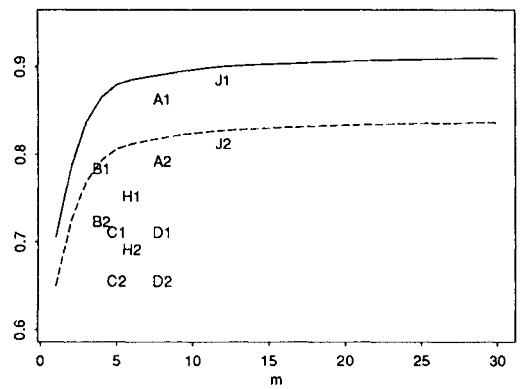

Figure 3: $V(\hat{T}_t)/V(y_t)$

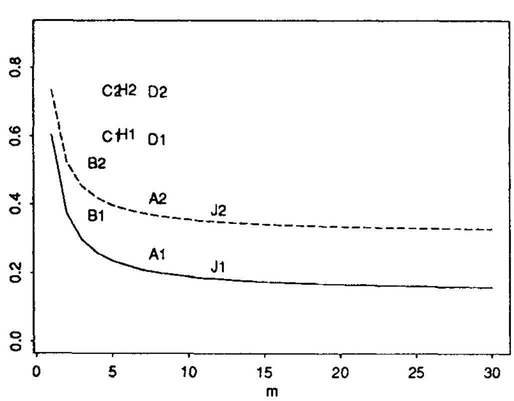

Figure 2: $V(SA_{t+1} - SA_t)/V(y_t)$

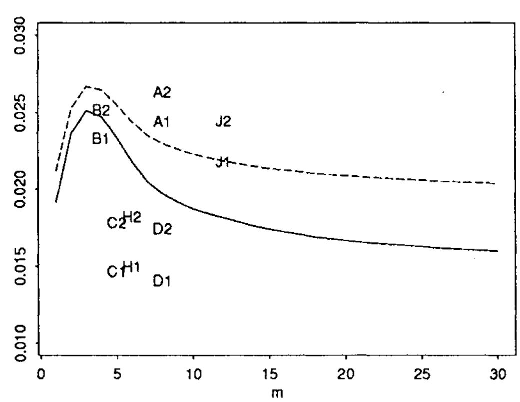

Figure 4: $V(\hat{T}_{t+1} - \hat{T}_t)/V(y_t)$