Working Paper No. 2003/1

ESTIMATION OF SEASONAL FACTORS FOR A SHORT TIME SPAN USING MULTI-LEVEL MODELLING WITH MIXED EFFECTS

Craig H. McLaren
Xichuan (Mark) Zhang

March 2003

ABS Catalogue No. 1351.0

EMBARGO: 11:30AM (CANBERRA TIME) FRI 16 May 2003

Working Paper No. 2003/1

Estimation of seasonal factors for a short time span using multi-level modelling with mixed effects

Craig H. McLaren and Xichuan (Mark) Zhang

March 2003

ABS Catalogue No. 1351.0

Contents

INQUIRIES

- *For further information about the contents of this Working Paper, please contact the author:* Craig McLaren — *telephone:* Canberra 02 6252 6540 *or facsimile:* Canberra 02 6252 8015 *or e-mail:* craig.mclaren@abs.gov.au

- *For information on the* Working Papers in Econometrics and Applied Statistics, *please contact the Managing Editor:* Valentin M. Valdez — *telephone:* Canberra 02 6252 7037 *or facsimile:* Canberra 02 6252 5251 *or e-mail:* valentin.valdez@abs.gov.au *or write* c/- Analytical Services Section, Australian Bureau of Statistics, PO Box 10, Belconnen, ACT 2616, Australia.

- *For information about other ABS statistics and services,* please refer to the back cover of this publication.

Abstract

Many time series are based on data obtained from sample surveys. Changes to the data source or methodology of a survey can have an impact on the original survey estimates and the seasonally adjusted and trend estimates derived from them. To assess the impacts of the changes to the survey, data can be collected from the old and the new survey for one or more overlapping time periods. This produces parallel series of original estimates. The number of overlapping time periods is typically small due to cost constraints. It is desirable to assist users by calculating seasonally adjusted and trend estimates for the new survey. Traditional seasonal adjustment methods cannot adequately calculate seasonally adjusted and trend estimates for short time series. Theory for estimating seasonal factors for short spans of time series data is given and illustrated with an example.

Keywords: short time series, multi-level modelling, seasonal factors, mixed effects

1 Introduction

The Australian Bureau of Statistics (ABS) is continually improving the sources and methods used in its surveys. This can occur by: changes in the sampling frames, scope and coverage of a survey, improved estimation techniques, new ways of collecting information or ultilising other sources of data such as administrative by product data. Methodological changes to the source or methods of a survey will impact on the original survey estimates and the associated seasonally adjusted and trend estimates. When such changes are substantial, for example the scope of the survey expands significantly, comparability of the data over time can be compromised.

Methodological changes to a survey can lead to the survey being replaced with a new survey that uses a vastly different survey design. It is desirable to assist users by calculating seasonally adjusted estimates for the new survey. Traditional seasonal adjustment methods cannot adequately calculate such estimates for short time series. This means there could be several years before reliable seasonally adjusted estimates are produced from the new survey.

This paper describes a method for estimating seasonal factors for short spans of time series data. The method uses a multi-level model with mixed effects. This method has been used to estimate and test for differences in the seasonal factors for the survey. We apply the method to estimate seasonal factors for the private sector gross earnings estimates obtained from the Quarterly Business Indicator Survey (ABS, 2001a), which is the replacement for the Survey of Employment and Earnings (ABS, 2001b), using four parallel quarter estimates in 2001. The results have been used in the compilation of the ABS National Accounts for March quarter 2002. The method described in this paper could also be applied to time series generated using an alternative approach to a survey, e.g. the use of administrative data sources.

2 Modelling framework

2.1 Old and new surveys

Consider the case where a new survey has started that will replace an existing survey, now refered to as the old survey. A realistic assumption may be that the new survey measures the same underlying activity as the old survey although the coverage and scope may be different. This means that the underlying movements are the same but the time series may shift to a different level. The ABS use the trend component (ABS, 1987) to represent the long term movement in a time series. When methodological changes occur, the trend level for the new survey may be different from the old survey, while the movement of the trend could be the same between the old and new survey. The seasonal factors could be different for the old and new surveys because of the different scope and coverage and other methodological changes that have occurred. For example, the new survey might have better coverage of small businesses and improved classification or identification of large businesses. This will mean the seasonality of the two surveys may be different, even though the same underlying activity is still being measured.

To help assess the impact of the changes, original estimates can be calculated using data collected from the old and the new survey for one or more overlapping time periods. The number of overlapping time periods is typically short due to cost constraints of conducting two surveys simultaneously.

Our objective is to test whether the seasonal patterns are different between the old and new surveys, and derive seasonal factors for the new survey if the differences are found. We will show how the information from these two parallel surveys, disaggregated over a number of lower level series, can be used in multi-level modelling to test the differences in the seasonal factors between the old and new survey data at an aggregate level for the overlapping time periods. Seasonal factors for the time series obtained from the new survey can also be obtained.

2.2 Multi-level and mixed effects models

Multi-level models can be used to reflect variation in parameters between groups in a hierachical structure in a given system (Goldstein, 1995).

In the case where a new survey and an old survey are being conducted and both surveys are measuring the same underlying activity, it is appropriate to use a mixed effects incorporating nested random effects. Laird and Ware (1982) describe a general approach to this type of model. The mixed effects model allows for fixed (common) effects between the new and old surveys and between different groups, and random effects across the new and old surveys and also across the different groups.

Denote an original series at time t by y_{ijt} where for the given survey i, the new survey is denoted by $i = 1$ and the old survey by $i = 0$; data is available at a number of group levels $j = 1, ..., n$. The original series can be decomposed using a multiplicative relationship

$$y_{ijt} = T_{ijt} \times S_{ijt} \times I_{ijt} \qquad (1)$$

where T_{ijt} is the trend component, S_{ijt} is the seasonal component, and I_{ijt} is the irregular component with a mean of 1.0. For a multiplicative related time series, a seasonal component of $S_{ijt} = 1.0$ is interpreted as that time point being seasonally neutral. $S_{ijt} > 1.0$ is interpreted as that time point being seasonally high, and $S_{ijt} < 1.0$ is interpreted as that time point being seasonally low.

Firstly, assume a log additive model for the time series decomposition. The Appendix gives an explicit correction factor for the bias introduced by using the log transformation. The results in this paper did not correct for the bias of the log transform. This is not expected to impact significantly on the conclusions.

Secondly, assume the logged trend component can be modelled as a locally linear process within the overlapping period. For example

$$\log(T_{ijt}) = (a + a_{ij}) + (b + b_j)t$$

where a represents the intercept term for the trend. a_{ij} denotes the intercept random effect at the survey and group level, b represents the fixed slope term for the trend, b_j represents the random effect of the slope at group level, $a_{ij} \sim N(0, \Sigma_{a_{ij}})$ and $b_j \sim N(0, \Sigma_{b_j})$. Note that the growth rate b_j does not depend on the survey. That is, both surveys have the same growth but will have different intercepts, a_{ij}. This means that the level of the trend are dependent on the survey and group, and the growth of the each group are the same for the both surveys.

For the seasonal component

$$\log(S_{ijt}) = (\mathbf{c}+\mathbf{c}_{ij})'\mathbf{s}_t$$

where $\mathbf{s}_t$ is a seasonal dummy variable, $\mathbf{c}$ represents the common fixed seasonal factor vector, $\mathbf{c}_{ij}$ denotes the seasonal random effect at the survey and group level, and $\mathbf{c}_{ij} \sim N(0, \Sigma_{\mathbf{c}_{ij}})$. Here, $\mathbf{c} = (c_{(1)}, ..., c_{(p)})$ and $\sum_{k=1}^{p} c_{(k)} = 0$ where p is the number of periods ($p = 4$ for quarterly and $p = 12$ for monthly). Similarly for $\mathbf{c}_{ij}$.

Combining these two steps and representing the trend, seasonal and irregular components in terms of a model based approach gives

$$\begin{aligned}\log(y_{ijt}) &= \log(T_{ijt}) + \log(S_{ijt}) + \log(I_{ijt}) \\ &= \underbrace{(a + a_{ij}) + (b + b_j)t}_{\text{trend}} + \underbrace{(\mathbf{c} + \mathbf{c}_{ij})'\mathbf{s}_t}_{\text{seasonal}} + \underbrace{\varepsilon_{ijt}}_{\text{irregular}}\end{aligned} \tag{2}$$

Due to the multiple levels hierarchical structure, we consider that the data is grouped with nested classifications factors ‘group’ and ‘survey’. The two nested grouping factors could have a different effect on the observed responses. With this setting, the log additive decomposition model (2) can be rewritten as a multiple level mixed random model

$$\begin{aligned}\log(y_{ijt}) &= \log(T_{ijt}) + \log(S_{ijt}) + \log(I_{ijt}) \\ &= \underbrace{(a + bt + \mathbf{c}'\mathbf{s}_t)}_{\text{fixed effect}} + \underbrace{(a_{ij} + b_j t + \mathbf{c}_{ij}'\mathbf{s}_t)}_{\text{random effect}} + \varepsilon_{ijt}\end{aligned} \tag{3}$$

where i= survey indicator (0,1), j= group indicator, a_{ij} and $\mathbf{c}_{ij}$ represent the intercept, and seasonal random effects at the survey within group level, b_j denotes the random effect of the slope at group level, ε_{ijt}= error term and t=time.

Under certain normality conditions, the fixed and random coefficients can be estimated by fitting the model (3) in the mixed model framework (Pinheiro and Bates, 1999).

The coefficients of the seasonal dummy variables $\mathbf{s}_t$ for each level of a factor cannot usually be estimated because of dependencies among the coefficients of the overall model. For example, the sum of all the seasonal dummy factors, which is a vector of all ones. This corresponds to the term used for fitting an intercept. Overparametrization induced by dummy variables needs to be removed prior to fitting the model. This is done by replacing the dummy variables with a set of linear combinations of the dummy variables. A particular choice of linear combinations of the dummy variables for the seasonal factors alters the specific individual coefficients in the model, but does not change the overall contribution of the term to the fit. Therefore, the coefficients of the seasonal dummy variables of model (3) may not be unique. This means that the estimated seasonal coefficients from model (3) may not be useful. However, the estimated intercepts and slopes can be used to produced a detrended series

$$\log(y^*_{ijt}) = \log(y_{ijt}) - (\hat{a} + \hat{a}_{ij} + \hat{b}t + \hat{b}_j t) \qquad (4)$$

The detrended series can be considered as containing a seasonal pattern which is affected by the group, survey within group, and irregulars.

Our objective is to test whether the seasonal factors have been changed from the old survey to new survey at the aggregated level. The fixed seasonal factors of the old and new surveys with the random seasonal effects at group level and at survey within group in model (5) can be used to provide statistical inferences to judge whether the seasonal factors have been changed. That is

$$\log(y^*_{ijt}) = \underbrace{\mathbf{c}_i'\mathbf{s}_t}_{\text{fixed effect}} + \underbrace{(\mathbf{c}_j + \mathbf{c}_{ij})'\mathbf{s}_t}_{\text{random effect}} + \varepsilon_{ijt} \qquad (5)$$

where $\mathbf{c}_i$ is the fixed seasonal effect for the old and new survey ($i = 0, 1$) respectively, and $\mathbf{c}_j \sim N(0, \Sigma\mathbf{c}_j)$ and $\mathbf{c}_{ij} \sim N(0, \Sigma\mathbf{c}_{ij})$ are random seasonal effects at group level and at survey within group level respectively.

Section 2.5 describes a practical approach to estimate seasonal factors at an aggregated level. We consider the case where the groups j are 15 industries to give state and Australia total estimates of the seasonal factors. Alternatively, the groups could have been taken as 8 states to give an Australian total estimate of the seasonal factors.

2.3 Bridging the model and filter based seasonal factors

The ABS does not currently use model based methods to estimate seasonal factors. The ABS uses a filter based approach based on the commonly used seasonal adjustment packages X11 (Shiskin et. al, 1967) and X12ARIMA (Findley et. al, 1998). For a description of ABS's approach see ABS (2001c).

Estimation of the parameters in the model based framework described in Section 2.2 can enable us to compare the difference between the old and new seasonal factors. We will assume for practical purposes that the ratio between the new and old seasonal faactors based on the model will be a good approximation to the filters as estimated under X11.

If the difference is statistically significant, the model based estimates of the seasonal factors need to be converted into a filter based framework. We assume that the proportional relationship between the seasonal factors for the new and old surveys is consistent between the filter and model based frameworks. This can be expressed as

$$\frac{Filter^{new}}{Filter^{old}} = \frac{\exp(Model^{new})}{\exp(Model^{old})} \qquad (6)$$

where $Filter^{new}$ are the unknown filter based seasonal factors for the new survey, $Filter^{old}$ are the known filter based seasonal factors for the old and can be estimated, $\exp(Model^{new})$ and $\exp(Model^{old})$ are the estimated model based seasonal factors for the new and old surveys respectively. Equation (6) is used to estimate $Filter^{new}$ which can then be applied within the ABS framework.

2.4 Steps involved in modelling seasonal factors

For a given time series these steps should be followed:

a. Estimate all the parameters in the full model as described in equation (3). For example, the estimation procedure nlme in Splus could be used (Pinheiro and Bates, 1999). Alternative estimation procedures could be used.

b. For practical convenience, we remove the estimate of the intercept and the slope (trend) from the original data

$$\log(y^*_{ijt}) = \log(y_{ijt}) - (\hat{a} + \hat{a}_{ij} + \hat{b}t + \hat{b}_j t)$$

Here, $\log(y^*_{ijt})$ contains seasonal patterns and irregulars from which the seasonal factors can be estimated by the seasonal dummy coefficients.

c. Estimate the common seasonal factors $\mathbf{c}$ with the seasonal random effects at group level at survey within group by using the model (without intercept)

$$\log(y^*_{ijt}) = \underbrace{\mathbf{c}_i^{\prime}\mathbf{s}_t}_{\text{fixed effect}} + \underbrace{(\mathbf{c}_j + \mathbf{c}_{ij})^{\prime}\mathbf{s}_t}_{\text{random effect}} + \varepsilon_{ijt}$$

The constrast between seasonal factors $\mathbf{c}_i$ $(i = 0, 1)$ with random seasonal effects at group and survey within group in model (5), can be used to provide statistical inferences to judge whether the seasonal factors have been changed.

If differences are found to be statistically significant, the estimates for the fixed common seasonal factors can be thought of as the seasonal factors for the new survey.

Equation (5) implies $\sum_{k=1}^{p} c_{(k)} = 0$ where $\mathbf{c} = (c_{(1)}, ..., c_{(p)})$ and p is the number of periods ($p = 4$ for quarterly and $p = 12$ for monthly). However, the model is fitted (estimation of $\mathbf{c}$) by a method of maximizing the restricted log-likelihood so this criteria may not be automatically met numerically. This means that the estimates for $\mathbf{c}$ will need to be scaled by an appropriate factor, f. For a multiplicative model this constraint is used to ensure that the estimated seasonal factor $\tilde{\mathbf{c}}_k$ satisfy $\sum_{k=1}^{p} \tilde{c}_{(k)} = p$ where $\tilde{c}_{(k)} = f\exp(c_{(k)})$.

d. Apply the model based seasonal factors to estimate X11 based seasonal factors using equation (6).

3 Real world example

The Quarterly Business Indicator Survey (QBIS) (ABS, 2001a) produces estimates of private wages and salaries on an accruals (accounts) basis. This survey replaced the Survey of Employment and Earnings (SEE) (ABS, 2001b), which produced estimates for private and public gross earnings on a cash (payroll) basis. SEE has now been replaced by a third survey which only collects public gross earnings. The private wages information collected from QBIS is used in the estimation of compensation of employees within the Australian system of National Accounts.

To estimate the impact of the change over to the new survey, four parallel quarter estimates were produced from March quarter 2001, to December quarter 2001 inclusive. This gave original estimates at these time points for both the QBIS and pay period adjusted SEE survey. An additional 4 data points for only the SEE data were used from March quarter 2000 to December quarter 2000 inclusive to ensure that the model described in (2) is identifiable and therefore numerically feasible. The original estimates for each survey were at the state by industry level. For example, NSW at Mining, Manufacturing, Electricity Gas and Water supply, Construction, Wholesale Trade, Retail Trade, Accommodation Cafes and Restaurants, Transport and Storage, Communication Services, Finance and Insurance, Property and Business Services, Education, Health and Community Services, Cultural and Recreational Services, Personal and Other Services. This industry level information was used to estimate seasonal factors at the NSW level. A separate analysis was done for all eight states and at the Australian total using industry as the groups. In this example, the number of data points available is small, 15 industries with 4 overlapping data points. Maas and Cox (2002) show that the small sample sizes have a negligible impact on the estimates of multi-level model parameter estimates.

The Australian total can be derived two different ways. To ensure consistency with the state total estimates the Australian total was calculated by aggregating over the industry totals. Alternatively, the total could have been derived by aggregating across the state totals.

The results of this analysis were used in the compilation of the ABS National Accounts for March quarter 2002. Seasonal factors were generated for March quarter 2002 to December quarter 2002. These seasonal factors are not presented here.

3.1 Results for Quarterly Business Indicator Survey and Survey of Employment and Earnings

Figures 1a and 1b show the estimated model based seasonal factors on the exponential scale for the new survey (QBIS) and the differences between the estimated model based seasonal factors on the exponential scale for the new and old survey (QBIS - SEE) for each state and total Australia level. An 80% confidence interval on the differences has also been calculated using estimates of the standard errors of the differences obtained from the estimation procedure. This confidence level was chosen as it gives a margin of error around 1%. For example, a movement in the seasonal factor from 1.00 to 0.99 equates to 1%. In practical terms this movement can be quite important.

All states and the Australian total, apart from Tasmania, had a low seasonal factor for the new survey for March quarter 2001. The June and September 2001 quarters seasonal factors for the new survey were a mixture of high and low seasonal factors across the different states and total. All states and the Australian total, apart from ACT, had a high seasonal factor for the new survey for December quarter 2001.

Queensland and the Northern Territory had the most differences between the seasonal factors for the old and new survey. In some cases, the differences were up to 6%. States which did not have significant difference between the model based seasonal factors for the old and new survey were: Victoria, Tasmania, ACT and the Australian total.

Figure 1a: Model based seasonal factors on the exponential scale and difference between the model based new and old seasonal factor point estimates for states with the bars denoting an 80% confidence interval for March quarter 2001 to December quarter 2001: New South Wales (NSW), Victoria (Vic), Queensland (Qld), South Australian (SA) and Western Australia (WA).

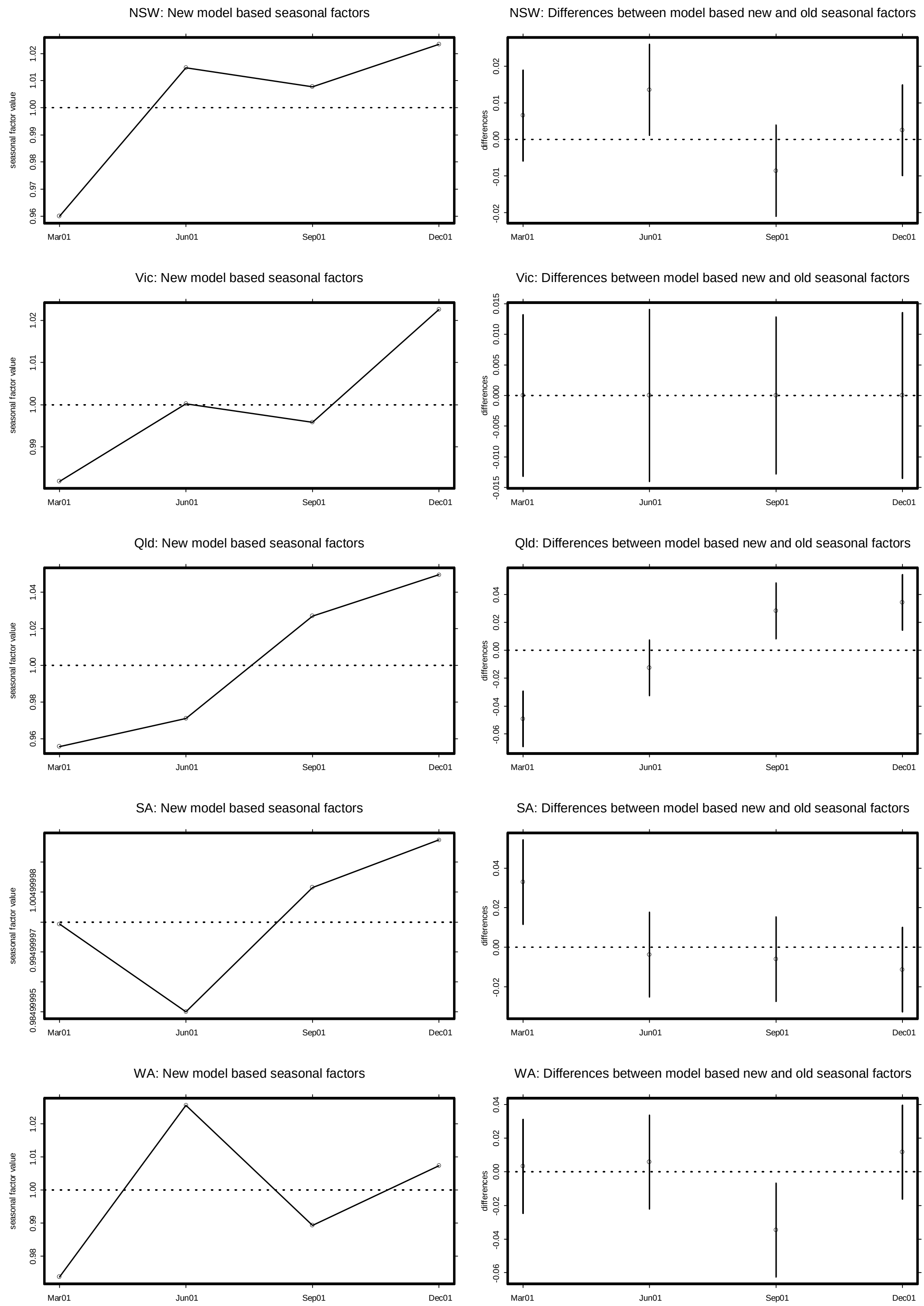

Figure 1b: Model based seasonal factors on the exponential scale and difference between the model based new and old seasonal factor point estimates for states with the bars denoting an 80% confidence interval for March quarter 2001 to December quarter 2001: Tasmania (Tas), Northern Territory (NT), Australian Capital Territory (ACT), and Total by Industry (Australian Total).

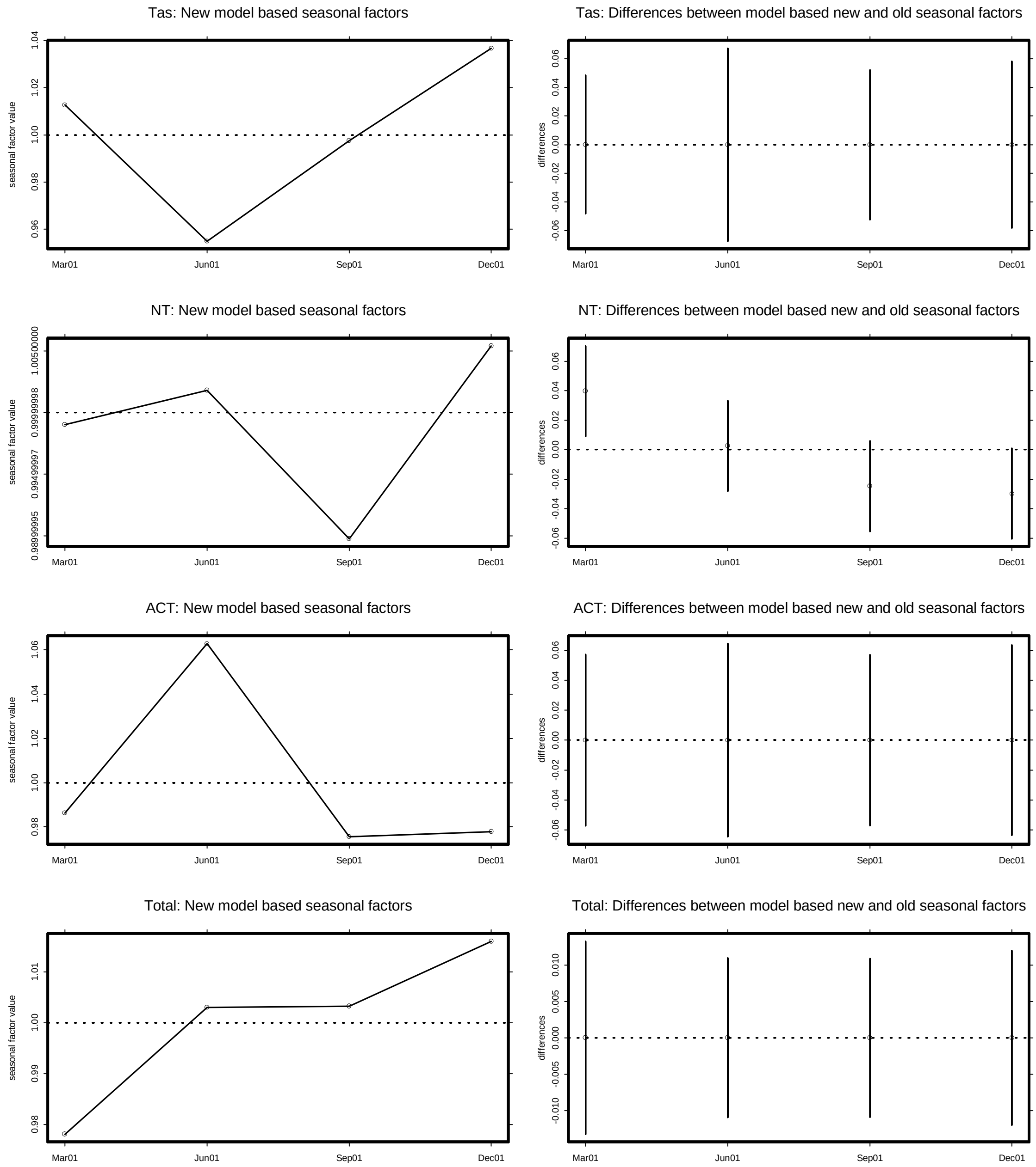

Table 1 gives comparisons between seasonal factors for the old and new survey at each state and total level after the model based seasonal factors have been converted to the filter based seasonal factors. This is essentially the same information as presented in Figures 1a and 1b and similar comments apply.

Table 2 gives 80% confidence intervals for the final filter based seasonal factors. As the seasonal factors for Victoria, Tasmania, ACT and Australian Total are not statistically significant between the old and new survey, the seasonal factors known for the old survey were used. For NSW, Queensland, South Australia, Western Australia, and Northern Territory the filter based seasonal factors estimated for the new survey were used. If the confidence intervals contain 1.0, then there is a chance that quarter may be seasonally neutral.

These results were used in the final decision on which were the appropriate seasonal factors to use. As the seasonal factors between the old and new survey were not significantly different for the Australian Total, the decision was made to continue to use the old (SEE) seasonal factors for the March 2002 quarter.

Table 1: Comparison of final filter based seasonal factors for the old and new survey for different states. * denotes that the estimated seasonal factors were not significantly different between the old and new survey (at the 20% confidence level).

	Series		Mar 2001	Jun 2001	Sep 2001	Dec 2001
	NSW	New	0.984	1.004	0.983	1.029
		Old	0.981	0.994	0.995	1.030
*	Victoria	New	0.983	0.999	0.984	1.034
		Old	0.985	0.993	0.992	1.031
	Queensland	New	0.924	0.979	1.019	1.081
		Old	0.972	0.992	0.991	1.045
	SA	New	0.998	0.985	0.999	1.015
		Old	0.969	0.992	1.008	1.030
	WA	New	0.986	1.024	0.962	1.029
		Old	0.979	1.015	0.993	1.013
*	Tasmania	New	0.988	1.023	0.967	1.022
		Old	0.974	1.019	0.965	1.042
	NT	New	1.047	1.001	0.966	0.987
		Old	1.004	0.995	0.987	1.014
*	ACT	New	0.986	1.021	0.967	1.024
		Old	0.962	1.043	0.990	1.005
*	Total x Ind	New	0.977	0.999	0.988	1.036
		Old	0.980	0.996	0.993	1.031

Table 2: Upper and lower confidence intervals (80%) estimated for the final filter based seasonal factors.

series		Mar 2001	Jun 2001	Sep 2001	Dec 2001
	lower	0.972	0.992	0.971	1.016
NSW		0.984	1.004	0.983	1.029
	upper	0.996	1.017	0.996	1.042
	lower	0.972	0.979	0.979	1.017
Victoria		0.985	0.993	0.992	1.031
	upper	0.998	1.007	1.005	1.045
	lower	0.906	0.959	0.999	1.060
Queensland		0.924	0.979	1.019	1.081
	upper	0.943	0.998	1.039	1.103
	lower	0.977	0.964	0.978	0.994
SA		0.998	0.985	0.999	1.015
	upper	1.020	1.007	1.021	1.037
	lower	0.958	0.996	0.935	1.000
WA		0.986	1.024	0.962	1.029
	upper	1.014	1.053	0.989	1.058
	lower	0.928	0.952	0.916	0.983
Tasmania		0.974	1.019	0.965	1.042
	upper	1.023	1.090	1.017	1.105
	lower	1.015	0.970	0.936	0.957
NT		1.047	1.001	0.966	0.987
	upper	1.080	1.032	0.996	1.017
	lower	0.909	0.977	0.935	0.943
ACT		0.962	1.043	0.990	1.005
	upper	1.019	1.112	1.048	1.071
	lower	0.967	0.985	0.982	1.019
Total x Ind		0.980	0.996	0.993	1.031
	upper	0.993	1.007	1.004	1.044

4 Conclusions

Methodology changes will continue to occur to surveys conducted by government statistical agencies. When this happens, new surveys which include the methodological changes may replace the previously used surveys. In the case where lower level group data are available, the approach of using a multi-level model with mixed effects as described in this paper provides an approach to estimating seasonal factors for short span time series at an aggregate level. Once seasonal factors have been estimated, seasonally adjusted and trend estimates can be derived for short span time series. This methodology can also be used when survey data is replaced by taxation or other administrative data.

Topics for future work will include: 1. Correction of the log bias transformation using the approach outlined in the Appendix. 2. The splicing of the new and old survey into one continuous time series. This can be achieved by obtaining an estimate from the multi-level model with mixed effects of the magnitude of the level shift between the two surveys. A seasonal break correction factor can be estimated once the seasonal factors for the new survey are known. This information can be used to give a trend and seasonal break into the spliced series to produce quality seasonally adjusted and trend estimates. 3. Improved estimates of the seasonal factors may be obtained by the use of appropriate covariance structures to handle survey sampling error in the mixed model estimation.

5 Acknowledgements

We would like to thank David Steel and Carole Birrell (University of Wollongong), Robert Tanton and Frank Yu (Australian Bureau of Statistics) for their insightful comments and advice.

5 References

Australian Bureau of Statistics (1987). A Guide to Smoothing Time Series - Estimates of "Trend". Australian Bureau of Statistics, 1316.0, Canberra, Australia.

Australian Bureau of Statistics (2001a). Business Indicators, Catalogue number 5676.0, Canberra, Australia.

Australian Bureau of Statistics (2001b). Wage and Salary Earners, Australia, Catalogue: 6248.0., Canberra, Australia.

Australian Bureau of Statistics (2001c). Introductory course on Time Series Analysis, Http://www.abs.gov.au/ausstats/abs@.nsf/0/8464A549BD14F38CCA256A5A0002E5C6?Open (free download)

Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C. and Chen B. (1998). New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program. *Journal of Business and Economic Statistics*, Vol. 16, No. 2., 127-177.

Franses, P.H., Arino M.A. and Hobijn B. (1997). Are many current seasonally adjusted data downward biased?, Erasmus University Rotterdam, The Netherlands, Report 9717/A. http://www.eur.nl/few/ei/papers

Goldstein, H. (1995). Multilevel statistical models, Halstead Press, New York.

Laird, N.M. and Ware, J.H. (1982). Random effects models for longitudinal data, Biometrics, 38: p963-974.

Maas, C.J.M. and Hox, J.J. (2002). Robustness of multilevel parameter estimates against small sample sizes. International Sociological Association, XVth ISA World Congress of Sociology, July 7-13, 2002.

Pinheiro, J.C. and Bates, D.M. (1999). Linear and Nonlinear Mixed-effects Methods and Classes for S and S-Plus. http://nlme.stat.wisc.edu//

Shiskin, J., Young, A. H. and Musgrave, J.C. (1967). The X11 Variant of the Census Method II Seasonal Adjustment Program. Technical Paper 15, Bureau of the Census, U.S. Department of Commerce, Washington, D.C.

A Appendix

See Franses et. al (1997) for an alternative approach to deriving a correction factor for the log transform of an ARIMA time series forecast.

Assume a multiplicative decomposition model for an original value Y_t at time t.

$$Y_t = T_t \times S_t \times I_t$$

This model is often converted to a log additive model

$$y_t = \tau_t + s_t + i_t$$

where $y_t = \log(Y_t)$, $\tau_t = \log(T_t)$, $s_t = \log(S_t)$ and $i_t = \log(I_t)$. Various statistical techniques are easier to apply to the additive model rather than the multiplicative model.

The conversion method for estimation of T_t is to use $T_t = \exp(\tau_t)$. This conversion gives biased estimators for T_t. The more generic problem can be described as follows.

Suppose X is a random variable with a lognormal distribution Lognormal$(E(X), Var(X))$. Therefore $Y = \log(X)$ is a random variable with normal distribution $\mathrm{N}(E(Y), Var(Y))$.

This exercise proves that

$$E(X) = \exp(E(Y) + 0.5Var(Y))$$

$$Var(X) = \{\exp(Var(Y)) - 1\}\exp(2E(Y) + Var(Y))$$

This result is important to correct bias when transforming from a logged additive model to a multiplicative model. An alternative general derivation can be done using Taylor series expansion.

Let $a = E(Y)$ and $\sigma^2 = Var(Y)$. Then $E(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^y e^{\frac{-(y-a)^2}{2\sigma^2}} dy$ and let $z = \frac{y-a}{\sigma}$ so

$$\begin{aligned}
E(X) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{z\sigma+a} e^{\frac{-z^2}{2}} dy \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(z^2-2z\sigma+\sigma^2-\sigma^2-2a)} dz \\
&= \frac{1}{\sqrt{2\pi}} e^{a+\frac{\sigma^2}{2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(z-\sigma)^2} dz \\
&= e^{a+\frac{\sigma^2}{2}} \\
&= \exp(E(Y) + \frac{1}{2}Var(Y))
\end{aligned}$$

Consider

$$
\begin{aligned}
Var(X) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (e^y - E(X))^2 e^{-\frac{(y-a)^2}{2\sigma^2}}\, dy \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{2\sigma z + 2a} e^{-\frac{z^2}{2}}\, dz - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} 2e^{\sigma z + a} E(X) e^{-\frac{z^2}{2}}\, dz + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} E(X)^2 e^{-\frac{z^2}{2}}\, dz
\end{aligned}
$$

Expanding each term gives

$$
\begin{aligned}
Var(X) &= e^{2(a+\sigma^2)} - 2E(X)^2 + E(X)^2 \\
&= e^{2(a+\sigma^2)} - e^{2a+\sigma^2} \\
&= (e^{\sigma^2} - 1)e^{2a+\sigma^2} \\
&= \{\exp(Var(Y)) - 1\} \exp(2E(Y) + Var(Y))
\end{aligned}
$$

FOR MORE INFORMATION...

INTERNET **www.abs.gov.au** the ABS web site is the best place to start for access to summary data from our latest publications, information about the ABS, advice about upcoming releases, our catalogue, and Australia Now—a statistical profile.

LIBRARY A range of ABS publications is available from public and tertiary libraries Australia-wide. Contact your nearest library to determine whether it has the ABS statistics you require, or visit our web site for a list of libraries.

CPI INFOLINE For current and historical Consumer Price Index data, call 1902 981 074 (call cost 77c per minute).

DIAL-A-STATISTIC For the latest figures for National Accounts, Balance of Payments, Labour Force, Average Weekly Earnings, Estimated Resident Population and the Consumer Price Index call 1900 986 400 (call cost 77c per minute).

INFORMATION SERVICE

Data which have been published and can be provided within five minutes are free of charge. Our information consultants can also help you to access the full range of ABS information—ABS user-pays services can be tailored to your needs, time frame and budget. Publications may be purchased. Specialists are on hand to help you with analytical or methodological advice.

PHONE **1300 135 070**

EMAIL **client.services@abs.gov.au**

FAX 1300 135 211

POST Client Services, ABS, GPO Box 796, Sydney 2001

WHY NOT SUBSCRIBE?

ABS subscription services provide regular, convenient and prompt deliveries of ABS publications and products as they are released. Email delivery of monthly and quarterly publications is available.

PHONE 1300 366 323

EMAIL subscriptions@abs.gov.au

FAX 03 9615 7848

POST Subscription Services, ABS, GPO Box 2796Y, Melbourne 3001

Produced by the Australian Bureau of Statistics