University of Wollongong
Research Online

Centre for Statistical & Survey Methodology
Working Paper Series

Faculty of Engineering and Information Sciences

2010

Parameter estimation and naïve bias for a seasonally adjusted aggregate series of different lengths using univariate and multivariate approaches

Carole Birrell
University of Wollongong, cbirrell@uow.edu.au

Yan Xia Lin
University of Wollongong

DAvid G. Steel
University of Wollongong

Recommended Citation

Birrell, Carole; Lin, Yan Xia; and Steel, DAvid G., Parameter estimation and naïve bias for a seasonally adjusted aggregate series of different lengths using univariate and multivariate approaches, Centre for Statistical and Survey Methodology, University of Wollongong, Working Paper 12-10, 2010, 32p.
http://ro.uow.edu.au/cssmwp/62

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Centre for Statistical and Survey Methodology

The University of Wollongong

Working Paper

12-10

Parameter estimation and naïve bias for a seasonally adjusted aggregate series of different lengths using univariate and multivariate approaches

(Running Title: *Seasonal adjustment of an aggregate series*)

Carole Birrell, Yan-Xia Lin and David G. Steel
University of Wollongong

Copyright © 2008 by the Centre for Statistical & Survey Methodology, UOW. Work in progress, no part of this paper may be reproduced without permission from the Centre.

Centre for Statistical & Survey Methodology, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

Parameter estimation and naïve bias for a seasonally adjusted aggregate series of different lengths using univariate and multivariate approaches

(Running Title: *Seasonal adjustment of an aggregate series*)

CAROLE L. BIRRELL [1] [2], YAN-XIA LIN[2] and DAVID G. STEEL[2]

University of Wollongong

Summary

An aggregate series is a time series resulting from the aggregation of two or more sub-series. This paper compares a model-based univariate and multivariate approach to seasonal adjustment of the aggregate series for different series lengths. A simulation study compares two outcomes: the accuracy of the estimated parameters of the aggregate series, and the naïve bias in the prediction error variance.

The results show that for the two examples studied, the use of the multivariate approach in the estimation of parameters improves the accuracy of the parameter estimates of the aggregated series. This was especially the case for short to medium length time series. The relative efficiencies of the seasonally adjusted aggregated series also showed good gains for the multivariate model. For one of the examples, there was a substantial decrease in the naïve bias with the use of the multivariate model.

Keywords: Aggregate series; basic structural model; Kalman filter; multivariate time series; naïve bias; seasonal adjustment; state space model.

[1]Author to whom correspondence should be addressed

[2]Centre for Statistical and Survey Methodology, School of Mathematics and Applied Statistics, University of Wollongong, 2522 NSW, Australia.

e-mail: cbirrell@uow.edu.au

Acknowledgements: This research was supported by a grant from the Australian Research Council and also by the Australian Bureau of Statistics. Special thanks to Mark Zhang for his valuable input into this project.

1. Introduction

Seasonal adjustment may be performed by filter-based methods or model-based methods. For a time series resulting from the aggregation of two or more sub-series, the adjustment can be made on the aggregate series directly, or it can be done indirectly using the sub-series. Although each of the sub-series is utilised in the indirect approach, each series is seasonally adjusted individually rather than jointly. A multivariate approach to seasonal adjustment is an alternative to both the direct and indirect methods. Limited research has been done in this area (see Geweke, 1978; Ghysels, 1997; Planas & Campolongo, 2001). Other studies including Harvey & Durbin (1986), and Harvey & Chung (2000), have shown that the multivariate basic structural model has been effectively used to achieve gains in the estimation of the components of a target series. The focus of this paper is the joint modelling of sub-series for the purpose of seasonal adjustment of their aggregated series.

A model-based univariate approach is compared with a multivariate approach to seasonal adjustment of an aggregate series for different series lengths. A univariate basic structural model (BSM) is applied directly to the aggregate series and a multivariate BSM is applied to a transformed system of sub-series. When the model parameters are unknown and are estimated from the data, the procedure of calculating the accuracy of the seasonally adjusted series becomes two-staged. The results of these two stages are studied in this paper.

In the first stage, the parameters of the aggregate series are estimated using maximum likelihood estimation. When considering the variance of the seasonally adjusted series the prediction mean squared error (PMSE) of the seasonal component is required. In the second stage, the estimated parameters are substituted into the state space model and using the Kalman filter, an estimate of the PMSE is obtained. This substitution is known to produce an underestimate of the true PMSE in univariate models. The bias is often

referred to as the 'naïve bias' and results from not accounting for the variability resulting from the estimation of the model parameters (Durbin & Koopman, 2001; Quenneville & Singh, 2000; and Pfeffermann & Tiller, 2005). A simulation study compares the outcomes of both stages of the univariate approach with the multivariate approach for series of lengths $T = 20,\ 24,\ 28,\ 40,\ 80,\ 120,\ 240$.

The BSM for the univariate and multivariate approaches are given in Section 2. The simulation study and the estimation of parameters is described in Section 3. The results are presented in Sections 4 and 5 with conclusions in Section 6.

2. Basic structural model (BSM)

2.1 Univariate BSM

For a single additive time series, the observations at time t denoted by Y_t, may be written as the sum of a local linear trend, L_t, a dummy seasonal component, S_t, and an irregular or disturbance term, $\varepsilon_{U,\,t}$. The univariate model adopted for the aggregate series is a local level seasonal model. For $t = 1, \ldots, T$, it is written as

$$Y_t = L_t + S_t + \varepsilon_{U,\,t}, \qquad \varepsilon_{U,\,t} \sim N(0, \sigma^2_{U,\,\varepsilon}) \tag{1}$$

$$L_{t+1} = L_t + \eta_{U,\,t}, \qquad \eta_{U,\,t} \sim N(0, \sigma^2_{U,\,\eta}) \tag{2}$$

$$S_{t+1} = -\sum_{j=1}^{s-1} S_{t+1-j} + \omega_{U,\,t} \qquad \omega_{U,\,t} \sim N(0, \sigma^2_{U,\,\omega}). \tag{3}$$

The disturbance terms $\eta_{U,\,t}$, $\omega_{U,\,t}$ and $\varepsilon_{U,\,t}$, are assumed to be serially and mutually independent, and their respective variances, $\{\sigma^2_{U,\,\eta},\ \sigma^2_{U,\,\omega},\ \sigma^2_{U,\,\varepsilon}\}$ are the parameters of the univariate model.

2.2 Multivariate BSM

For a multivariate BSM, the disturbance terms are decomposed into common effects, which are time specific, and time-unit specific effects (Marshall, 1992). The local level seasonal model for the observation for series k at time t, denoted by Y_{kt}, is given below with $k = 1, 2, \cdots, K$ representing the K sub-series with dummy seasonal components.

$$Y_{kt} = L_{kt} + S_{kt} + \varepsilon_t + \varepsilon^*_{kt}, \qquad \varepsilon_t \sim N(0, \sigma^2_\varepsilon), \quad \varepsilon^*_{kt} \sim N(0, \sigma^2_{k\varepsilon^*}),$$

$$\begin{aligned} L_{k,\,t+1} &= L_{kt} + \eta_t + \eta^*_{kt}, & \eta_t &\sim N(0, \sigma^2_\eta), \quad \eta^*_{kt} \sim N(0, \sigma^2_{k\eta^*}), \\ S_{k,\,t+1} &= -\sum_{j=1}^{s-1} S_{k,\,t+1-j} + \omega_t + \omega^*_{kt}, & \omega_t &\sim N(0, \sigma^2_\omega), \quad \omega^*_{kt} \sim N(0, \sigma^2_{k\omega^*}). \end{aligned} \tag{4}$$

The disturbance terms, ε_t, ε^*_{kt}, η_t, η^*_{kt}, ω_t, ω^*_{kt} are assumed to be mutually independent Normal random variables. The common effects are η_t, ω_t, ε_t and the time-unit specific effects are ε^*_{kt}, η^*_{kt}, ω^*_{kt}.

The aggregate series is given by:

$$\begin{aligned} Y_{tot,\,t} = \sum_{k=1}^{K} Y_{kt} &= L_{tot,\,t} + S_{tot,\,t} + \varepsilon_{tot,\,t}, & \varepsilon_{tot,t} &\sim N(0, \sigma^2_{tot,\varepsilon}), \\ L_{tot,\,t+1} &= L_{tot,\,t} + \eta_{tot,\,t}, & \eta_{tot,t} &\sim N(0, \sigma^2_{tot,\eta}), \\ S_{tot,\,t+1} &= -\sum_{j=1}^{s-1} S_{tot,\,t+1-j} + \omega_{tot,\,t}, & \omega_{tot,t} &\sim N(0, \sigma^2_{tot,\omega}). \end{aligned}$$

3. Simulation Experiment

To investigate the behaviour of the univariate and multivariate models for varying series lengths, a simulation experiment is carried out. Data is simulated for two sub-series and aggregated to obtain the total series for lengths, T = 20, 24, 28, 40, 80, 120,

240. For quarterly data, this corresponds to series of length 5, 6, 7, 10, 20, 30 and 60 years. Time series of length 5 to 7 years are often considered as short series, 10 to 20 years are considered moderate length series and 30 or more years are long length.

3.1 Generation of Data Series

The parameters for the aggregate series are set to

$$\sigma^2_{U,\,\eta} = 0.5, \quad \sigma^2_{U,\,\omega} = 1.0, \quad \sigma^2_{U,\,\varepsilon} = 1.0.$$

In a previous study by Birrell et al. (2010) it was shown that the parameter values of the aggregate series did not have a great effect on the results that were produced. The seasonal to non-seasonal ratio for this series is $\sigma^2_{tot,\omega}/(\sigma^2_{tot,\eta} + \sigma^2_{tot,\varepsilon}) = 1/1.5 = 0.67$.

The parameters of the aggregate series may be written in terms of the sub-series parameters. Given $K = 2$, the following relationships hold:

$$\sigma^2_{tot,\eta} = 4\sigma^2_{\eta} + \sigma^2_{1\eta^*} + \sigma^2_{2\eta^*}, \quad \sigma^2_{tot,\omega} = 4\sigma^2_{\omega} + \sigma^2_{1\omega^*} + \sigma^2_{2\omega^*}, \quad \sigma^2_{tot,\varepsilon} = 4\sigma^2_{\varepsilon} + \sigma^2_{1\varepsilon^*} + \sigma^2_{2\varepsilon^*} \quad (5)$$

Note that given the true parameters, the following equalities will hold: $\sigma^2_{U,\,\eta} = \sigma^2_{tot,\eta}$, $\sigma_{U,\,\omega} = \sigma^2_{tot,\omega}$ and $\sigma^2_{U,\,\varepsilon} = \sigma^2_{tot,\varepsilon}$. When parameters are estimated however, these equalities may not necessarily hold due to estimation using the two different approaches.

To study the effect of the relationship of parameters between series, a measure of the between series similarity of the stochastic properties for each component is defined. Let the c-ratio for the level component (c_η) be defined as follows and similarly for the seasonal and error components:

$$c_\eta = \frac{\mathrm{Var}(L_{1,t+1} - L_{1t})}{\mathrm{Var}(L_{2,t+1} - L_{2t})} = \frac{\sigma^2_{\eta} + \sigma^2_{1\eta^*}}{\sigma^2_{\eta} + \sigma^2_{2\eta^*}}, \quad c_\omega = \frac{\sigma^2_{\omega} + \sigma^2_{1\omega^*}}{\sigma^2_{\omega} + \sigma^2_{2\omega^*}}, \quad c_\varepsilon = \frac{\sigma^2_{\varepsilon} + \sigma^2_{1\varepsilon^*}}{\sigma^2_{\varepsilon} + \sigma^2_{2\varepsilon^*}}.$$

In a previous study by Birrell et al. (2010), the accuracy of the seasonally adjusted aggregate series given the true parameters were studied for different combinations of c_η, c_ω and c_ε. In this study, two of these combinations which were shown to produce different results have been chosen in order to investigate the effect of the estimated parameters. They will be named *Set 1* and *Set 2*. These two sets have different c-ratios but the same correlations:

$$\begin{aligned}
&\textit{Set 1}: \ c_\omega = 1, \quad c_\eta = 20, \quad c_\varepsilon = 20, \quad \rho_\omega = 0.1, \quad \rho_\eta = 0.2, \quad \rho_\varepsilon = 0.2;\\
&\textit{Set 2}: \ c_\omega = 10, \quad c_\eta = 0.1, \quad c_\varepsilon = 0.1, \quad \rho_\omega = 0.1, \quad \rho_\eta = 0.2, \quad \rho_\varepsilon = 0.2.
\end{aligned}$$

Given the values of the c-ratios, the correlation coefficient for each component and the series parameters for the aggregate series, the multivariate seasonal parameters may be calculated. Solved in terms of $\sigma^2_{tot,\omega}$, c_ω and ρ_ω, the seasonal parameters are

$$\sigma^2_\omega = \frac{\rho_\omega\sqrt{c_\omega}\sigma^2_{tot,\omega}}{1 + c_\omega + 2\rho_\omega\sqrt{c_\omega}}, \quad \sigma^2_{1\omega^*} = \frac{\sigma^2_{tot,\omega}(c_\omega - \rho_\omega\sqrt{c_\omega})}{1 + c_\omega + 2\rho_\omega\sqrt{c_\omega}}, \quad \sigma^2_{2\omega^*} = \frac{\sigma^2_{tot,\omega}(1 - \rho_\omega\sqrt{c_\omega})}{1 + c_\omega + 2\rho_\omega\sqrt{c_\omega}}.$$

Since $\sigma^2_{1\omega^*} \geq 0$, $\sigma^2_{2\omega^*} \geq 0$ and $\sigma^2_\omega > 0$, the restrictions on the correlations are such that if $c_\omega \geq 1$, then $0 < \rho_\omega \leq 1/\sqrt{c_\omega}$, and if $c_\omega < 1$, then $0 < \rho_\omega \leq \sqrt{c_\omega}$. Similar constraints apply to the level and error components. The multivariate parameter values are given in Table 1.

TABLE 1 ABOUT HERE

The data for each sub-series are generated from the model equations given by (4) using the model parameters in Table 1. The total series is then obtained by contemporaneously aggregating the two sub-series. This procedure is repeated to produce the 1000 realisations of each sub-series and their aggregate.

3.2 Estimation of Parameters

The univariate and multivariate BSMs may be written more concisely in state space form (SSF) and then analysis is carried out with the Kalman filter. Details are given in the Appendix. For more general discussion of state space models and the Kalman filter refer to Kalman (1960), Harvey(1989), and Durbin & Koopman (2001).

3.2.1 Univariate Model

For the univariate local level seasonal model given in (1) - (3), let the vector of the three parameters be $\psi^{(U)} = (\sigma^2_{U,\,\eta},\ \sigma^2_{U,\,\omega},\ \sigma^2_{U,\,\varepsilon})^\top$.

It is possible to reduce the dimension of the parameter vector and thereby reduce the dimensionality of the numerical search by concentrating out one parameter (denote by σ^2_c) from the log-likelihood function. This can also improve the numerical stability of the optimisation (see Zivot & Wang, 2006, p558 and Harvey, 1989, p183). The measurement error variance parameter, $\sigma^2_{U,\,\varepsilon}$ is chosen as the concentrated parameter, and a square root transformation is applied to ensure that the estimated model parameters are positive. The maximisation is therefore performed with respect to $\psi_c^{(U)} = \left(\sqrt{q_{U,\,\eta}},\ \sqrt{q_{U,\,\omega}}\right)'$. Thus, $q_{U,\,\eta}$

and $q_{U,\,\omega}$ are the signal-to-noise ratios for the level and seasonal components respectively and are defined as $q_{U,\,\eta} = \sigma^2_{U,\,\eta}/\sigma^2_{U,\,\varepsilon}$ and $q_{U,\,\omega} = \sigma^2_{U,\,\omega}/\sigma^2_{U,\,\varepsilon}$. The details of the state space form for the univariate model is given in the Appendix. Referring to (15) in the Appendix, $\mathbf{Q} = \mathrm{Var}\,(\eta_{U,\,t},\ \omega_{U,\,t})^\top$ and if $\sigma^2_{U,\,\varepsilon}$ is the concentrated parameter, then $\mathbf{Q}$ can be rewritten as $\sigma^2_{U,\,\varepsilon}\mathbf{Q}^c$ where $\mathbf{Q}^c = \mathbf{Q}/\sigma^2_{U,\,\varepsilon}$. Similarly, we can write $\mathrm{Var}(\varepsilon_{U,\,t}) = \sigma^2_{U,\,\varepsilon}H^c$ where $H^c = 1$.

The standard Kalman filter equations (given in the Appendix) are amended such that

$$\begin{aligned} F_t &= \sigma^2_{U,\,\varepsilon}F^c_t \\ F^c_t &= \mathbf{Z}\mathbf{P}^c_{t|t-1}\mathbf{Z}^\top + 1 \\ \mathbf{P}_{t|t-1} &= \sigma^2_{U,\,\varepsilon}\mathbf{P}^c_{t|t-1} \end{aligned}$$

The concentrated diffuse log-likelihood function for the univariate model is given as

$$\ln\,\mathrm{L}_{dc} = -\frac{T}{2}\ln(2\pi) - 2\ln 2 - \frac{(T-4)}{2}\ln(\sigma^2_{U,\,\varepsilon}) - \frac{1}{2}\sum_{t=5}^{T}\ln F^c_t - \frac{1}{2\sigma^2_{U,\,\varepsilon}}\sum_{t=5}^{T}\frac{\nu_t^2}{F^c_t} \tag{6}$$

where ν_t and F^c_t do not depend on $\sigma^2_{U,\,\varepsilon}$. To determine the value of $\sigma^2_{U,\,\varepsilon}$ conditional on a given value of $\psi^{(U)}_c$, the concentrated diffuse log-likelihood (6) is differentiated with respect to $\sigma^2_{U,\,\varepsilon}$ and solved to give

$$\hat{\sigma}^2_{U,\,\varepsilon}(\psi^{(U)}_c) = \frac{1}{(T-4)}\sum_{t=5}^{T}\frac{\nu_t^2}{F^c_t}.$$

This function is maximised with respect to the elements of $\psi^{(U)}_c$ (Harvey, 1989, p127) to obtain the maximum likelihood estimates of $\sqrt{q_{U,\,\eta}}$ and $\sqrt{q_{U,\,\omega}}$.

3.2.2 Multivariate Model

The multivariate local level seasonal model described in (4) has two existing sub-series ($K = 2$) and hence nine parameters in the model. Let the original nine parameters in this multivariate model be denoted by $\psi^{(m)}$ such that:

$$\psi^{(m)} = \left(\sigma^2_{\eta},\ \sigma^2_{1\eta^*},\ \sigma^2_{2\eta^*},\ \sigma^2_{\omega},\ \sigma^2_{1\omega^*},\ \sigma^2_{2\omega^*},\ \sigma^2_{\varepsilon},\ \sigma^2_{1\varepsilon^*},\ \sigma^2_{2\varepsilon^*}\right)^{\top}.$$

The details of the state space form for the multivariate model is given in the Appendix. It is not possible to concentrate any of the covariance matrices out of the likelihood function as a block. However, it is possible to concentrate out one of the diagonal elements from one of the component covariance matrices (Harvey, 1989, Section 8.2.2). The concentrated parameter is set as the first diagonal element of $\mathbf{\Sigma}_{(m),\,\varepsilon}$, from (16) found in the Appendix. The concentrated parameter for the multivariate model is given by $\sigma^2_c = \sigma^2_{\varepsilon} + \sigma^2_{1\varepsilon^*}$.

The standard Kalman filter equations are given an (m) subscript to denote the multivariate model and amended such that

$$\begin{aligned}
\mathbf{F}_{(m),t} &= \sigma^2_c \mathbf{F}^c_{(m),t} \\
\mathbf{F}^c_{(m),t} &= \mathbf{Z}_{(m)} \mathbf{P}^c_{(m),t|t-1} \mathbf{Z}^{\top}_{(m)} \\
\mathbf{P}_{(m),t|t-1} &= \sigma^2_c \mathbf{P}^c_{(m),t|t-1}.
\end{aligned}$$

The concentrated diffuse log-likelihood function for the multivariate model is given by the following:

$$\begin{aligned}
\ln \mathrm{L}_{(m),dc} &= -T\ln(2\pi) - 4\ln 2 - (T-4)\ln(\sigma^2_c) - \frac{1}{2}\sum_{t=5}^{T} \ln|\mathbf{F}^c_{(m),t}| \\
&\quad - \frac{1}{2\sigma^2_c}\sum_{t=5}^{T} \nu^{\top}_t \left(\mathbf{F}^c_{(m),t}\right)^{-1} \nu_t. \qquad (7)
\end{aligned}$$

Note that for the multivariate model, there are Kq elements in the state vector which are non-stationary. For the univariate model, $q = 4$, hence if $K = 2$, there are 8 non-stationary elements in the state vector $\alpha_{(m),t}$ (details are in the Appendix).

When ln $\mathrm{L}_{(m),c}$ (7) is differentiated with respect to σ_c^2 and set to zero, the result is

$$\hat{\sigma}_c^2 = \frac{1}{2(T-4)} \sum_{t=5}^{T} \nu_t^\top \left(\mathbf{F}_{(m),t}^c\right)^{-1} \nu_t.$$

For the derivation of the general case (see Koopman et al., 1999).

For estimation, the parametrization of the component covariance matrices ($\mathbf{\Sigma}_{(m),\,\eta}$, $\mathbf{\Sigma}_{(m),\,\omega}$ and $\mathbf{\Sigma}_{(m),\,\varepsilon}$) requires two important characteristics. Firstly, each of the estimated component covariance matrices needs to be positive semi-definite, and secondly, the relationships between the elements within each matrix needs to be retained. Also, the off-diagonal elements here are required to be positive.

To satisfy these conditions, the concentrated log-likelihood is maximised with respect to the elements of $\psi_c^{(m)} = (x_1 \ldots x_8)'$ which are:

$$x_1 = \sqrt{\frac{\sigma_\eta^2}{\sigma_c^2}}, \quad x_2 = \sqrt{\frac{\sigma_{1\eta^*}^2}{\sigma_c^2}}, \quad x_3 = \sqrt{\frac{\sigma_{2\eta^*}^2}{\sigma_c^2}} \quad x_4 = \sqrt{\frac{\sigma_\omega^2}{\sigma_c^2}},$$
$$x_5 = \sqrt{\frac{\sigma_{1\omega^*}^2}{\sigma_c^2}}, \quad x_6 = \sqrt{\frac{\sigma_{2\omega^*}^2}{\sigma_c^2}} \quad x_7 = \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_c^2}}, \quad x_8 = \sqrt{\frac{\sigma_{2\varepsilon^*}^2}{\sigma_c^2}}.$$

To ensure that $\hat{\sigma}_{1\varepsilon^*}^2$ is non-negative, the maximum likelihood estimate for x_7 will need to be constrained so that $0 \leq x_7^2 \leq 1$. This can be achieved in the `S-PLUS` software by setting a lower bound of -1 and an upper bound of 1 on the estimate for x_7. With this constraint in place, the component covariance matrices will be positive semi-definite.

To obtain the values corresponding to the estimated univariate parameters contained in $\hat{\psi}^{(U)}$, the expressions for the total parameter values are calculated from the estimated multivariate parameters using (5).

4. Results of Parameter Estimates for Varying Series Length

The parameter estimates for 1000 realisations for each length of T were calculated for the univariate and multivariate methods for *Set 1* and *Set 2*. The aim is to compare the results of the univariate parameter estimates ($\hat{\sigma}^2_{U,\eta}$, $\hat{\sigma}^2_{U,\omega}$, $\hat{\sigma}^2_{U,\varepsilon}$) with the corresponding parameter estimates for the aggregate series ($\hat{\sigma}^2_{tot,\eta}$, $\hat{\sigma}^2_{tot,\omega}$, $\hat{\sigma}^2_{tot,\varepsilon}$) calculated using (5) with the multivariate estimates.

The results for all values of T studied are given in Table 2 for *Set 1* and in Table 3 for *Set 2*. The tables show the median, mean and standard errors for the estimated parameters of each component over the 1000 series generated. Not surprisingly, the estimates closest to the true parameter values are for $T = 240$, the longest series. However, the results for the median show that the parameter estimates given by the multivariate method are closer to the exact value for every pair (univariate vs multivariate) for each value of T except for the level component for $T \geq 40$ for *Set 1* and $T > 40$ for *Set 2*. This is an interesting result which suggests there may be a series length, $T = T_L$, at which this 'cross-over' occurs. Thus, for $T > T_L$, the estimator of the level parameter has greater precision using the univariate model.

In general, there is a marked improvement in the median of the estimates when the sub-series are used in the estimation process. For the level and seasonal parameters, there is an underestimate of the true value, whereas the estimates for the measurement error parameter show an overestimate for each value of T. The means fluctuate slightly as T increases, probably due to the presence of outliers but become closer overall to their true parameter value as T reaches 240. As expected, the standard errors generally decrease as T increases. They are generally smaller for the multivariate method than for the univariate method.

For the seasonal parameter, the results are shown in Figure 1. The exact parameter for the seasonal component is 1.0, as shown by the horizontal dotted line. It can be

seen that the median of the estimates is an underestimate of the true value for each length, T, for both the univariate and multivariate methods. However, the median for the multivariate method is greater, and hence closer to the true value for every length. The distributions are close to symmetrical for $T \geq 120$, and show that asymptotically the estimate approaches the true value. Again, as T decreases, the distributions become more positively skewed with outliers sometimes more extreme for the multivariate method. The range and interquartile range are smaller for most T for the multivariate method. In particular for *Set 1*, the distribution of the parameter estimates for the multivariate method for $T = 24$ is similar to that for the univariate method for $T = 40$. Notably for *Set 2*, when $T = 28$, 24, and 20, the lower quartile of the multivariate method is approximately equal to the median for the corresponding T for the univariate method.

To compare the variability of the estimates for the two methods, the relative efficiencies of the variances are calculated (using RE= V(univ)/V(mult)) from the results shown in Tables 2 and 3 and are given in Table 4. This table highlights the gains which are achieved with the multivariate model. For *Set 1* the relative efficiency varies between 1.20 and 1.38 for the level parameter. For the seasonal parameter, the relative efficiency shows greater gains than for the level parameter, with values ranging from 1.30 for $T = 20$ to 1.67 for $T = 80$. The gains are higher again for the measurement error parameter.

For *Set 2*, the relative efficiencies of the level parameter estimates range from 1.06 for $T = 24$ to 1.47 for $T = 120$. For the seasonal parameter, the gains are much greater, with values ranging from 1.73 for $T = 24$ to 2.28 for $T = 240$. For the estimate of the seasonal parameter, the relative efficiency of the variances are higher for *Set 2* than for *Set 1*, although all values are above unity. For the measurement error, there is not a distinctive pattern across T.

These results show that for the two examples studied, the use of the multivariate approach in the estimation of parameters improves the accuracy of the parameter estimates

of the aggregated series.

TABLE 2 ABOUT HERE

TABLE 3 ABOUT HERE

TABLE 4 ABOUT HERE

FIGURE 1 ABOUT HERE

5. Effect of Parameter Estimation on MSE of Model Components

The second stage of the estimation of the state vector components and their corresponding mean squared errors are discussed in this section. When considering the variance of the seasonally adjusted series, the prediction mean squared error (PMSE) of the seasonal component is required. The second stage of the estimation process involves the substitution of the estimated parameters in the theoretical expressions for the PMSE of the state vector components. This process is known to produce an underestimate of the true PMSE in univariate models. The bias is often referred to as the 'naïve bias' and results from not accounting for the variability resulting from the estimation of the model parameters (Durbin & Koopman, 2001; Quenneville & Singh, 2000; Pfeffermann & Tiller, 2005). In this section, the naïve bias will be examined for different series lengths for the univariate model and for the multivariate model. It is not known how the naïve bias behaves in multivariate models, as previous literature has focussed on univariate models such as the local level model as discussed in Quenneville & Singh (2000) and Pfeffermann & Tiller (2005). A trigonometric seasonal model is also considered in Pfeffermann & Tiller (2005).

The smoothed estimate of the state vector, α_t (defined in (15) in the Appendix), is conditional on all observations, $(\mathbb{Y}_T)$, and is denoted by $\mathbf{a}_{t|T}$. It is the mean of the distribution of α_t given $\mathbb{Y}_T$ and is defined in Harvey (1989, Section 3.6). The matrix $\mathbf{V}_{t|T}$ is the MSE matrix of the estimator, $\mathbf{a}_{t|T}$.

If the estimated parameters, $\hat{\psi}^{(U)}$, are used in the Kalman smoother, then $\mathbf{a}_{t|T}$ and $\mathbf{V}_{t|T}$ are estimated with:

$$\begin{aligned}
\hat{\mathbf{a}}_{t|T} &= \mathrm{E}\left(\alpha_t |\, \mathbb{Y}_T,\ \hat{\psi}^{(U)}\right) \\
\hat{\mathbf{V}}_{t|T} &= \mathrm{Var}\left(\alpha_t |\, \mathbb{Y}_T,\ \hat{\psi}^{(U)}\right) \\
&= \mathrm{E}_T\left[\left(\alpha_t - \hat{\mathbf{a}}_{t|T}\right)\left(\alpha_t - \hat{\mathbf{a}}_{t|T}\right)^{\top}\right].
\end{aligned}$$

The main focus of this paper is to compare the accuracy of the seasonally adjusted aggregate series of the direct approach with that of the multivariate approach. A simple transformation of the multivariate model allows explicit analysis of $Y_{tot,t}$ and hence a method of comparison of the accuracy of the seasonally adjusted aggregate series between the two approaches.

Let $\mathbf{A}$ be a $K \times K$ matrix:

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix} = \left(\begin{array}{c|c} 1\ 1 \quad \dots & 1 \\ \hline & 0 \\ \mathbf{I}_{(K-1)} & \vdots \\ & 0 \end{array}\right).$$

Applying $\mathbf{A}$ to obtain the transformed data, the aggregate series becomes augmented to the set comprising of series 1 to series $(K-1)$. Define $\mathbf{Y}_{(M),\,t}$ such that

$$\mathbf{Y}_{(M),\,t} = \mathbf{A}\left(Y_{1t}, Y_{2t}, \dots, Y_{Kt}\right)^{\top} = \left(Y_{tot,t}, Y_{1t}, \dots, Y_{K-1,t}\right)^{\top}. \tag{8}$$

For the multivariate model, the smoothed state vector and its MSE matrix have an (M) subscript:

$$\begin{aligned} \hat{\mathbf{a}}_{(M),\,t|T} &= \mathrm{E}\left(\alpha_{(M),\,t} |\, \mathbb{Y}_{(M),\,T},\ \hat{\psi}^{(M)}\right) \\ \hat{\mathbf{V}}_{(M),\,t|T} &= \mathrm{Var}\left(\alpha_{(M),\,t} |\, \mathbb{Y}_{(M),\,T},\ \hat{\psi}^{(M)}\right). \end{aligned}$$

where $\hat{\psi}^{(M)}$, and is given by

$$\hat{\psi}^{(M)} = \left(\hat{\sigma}^2_{tot,\eta},\ \hat{\sigma}^2_{\eta},\ \hat{\sigma}^2_{1\eta^*},\ \hat{\sigma}^2_{tot,\omega},\ \hat{\sigma}^2_{\omega},\ \hat{\sigma}^2_{1\omega^*},\ \hat{\sigma}^2_{tot,\varepsilon},\ \hat{\sigma}^2_{\varepsilon},\ \hat{\sigma}^2_{1\varepsilon^*}\right)^{\top}. \tag{9}$$

The estimate of the seasonal component given by the univariate approach is denoted by $\hat{\mathcal{S}}^U_{t|T}$, which is an element of $\hat{\mathbf{a}}_{t|T}$. Its mean squared error, $MSE\left(\hat{\mathcal{S}}^U_{t|T}\right)$ is an element of $\hat{\mathbf{V}}_{t|T}$. Similarly for the multivariate approach, the estimate of the seasonal component is denoted by $\hat{\mathcal{S}}^M_{t|T}$, and is an element of $\hat{\mathbf{a}}_{(M),\,t|T}$. Its mean squared error is given by $MSE\left(\hat{\mathcal{S}}^M_{t|T}\right)$ and is an element of $\hat{\mathbf{V}}_{(M),\,t|T}$. More detail on $\alpha_{(M),\,t}$ is given in the Appendix.

5.1 Naïve Bias

The naïve bias is the bias in the prediction error variance resulting from the substitution of the parameter estimates into the Kalman smoother. To calculate the naïve bias, the 'true' MSE is approximated using the simulated series component and the estimated series component, as given in Pfeffermann & Tiller (2005, p903). The calculations here are based on 1000 simulated series for each series length. For the seasonal component, it is

$$MSE^U_t = \sum_{i=1}^{1000} \frac{\left(\hat{\mathcal{S}}^U_{t|T,\,i} - S_{t,\,i}\right)^2}{1000}, \qquad i = 1, \ldots, 1000 \tag{10}$$

where $S_{t,i}$ is the simulated value of the seasonal component at time t for the ith generated series and $\hat{\mathcal{S}}^U_{t|T,\,i}$ is the estimated seasonal component at time t determined by the Kalman smoother, given the vector of estimated parameters, $\hat{\psi}^{(U)}$, for the ith generated data series of length T.

The form of (10) for the multivariate approach is given by

$$MSE^M_t = \sum_{i=1}^{1000} \frac{\left(\hat{\mathcal{S}}^M_{t|T,\,i} - S_{t,\,i}\right)^2}{1000}, \quad i = 1, \ldots, 1000.$$

For the naïve bias, let $\widehat{MSE}_{i,t}^{U}$ denote the state variance for the seasonal component calculated by the Kalman smoother with the estimated parameters, $\hat{\psi}^{(U)}$, given by the univariate model for the ith generated series. Similarly, $\widehat{MSE}_{i,t}^{M}$ is the notation adopted for the multivariate model. The mean bias over the 1000 realisations and its associated MSE are determined by

$$\text{Univ.:} \quad \bar{d}_t^U = \sum_{i=1}^{1000} \frac{\widehat{MSE}_{i,t}^{U} - MSE_t^U}{1000}, \quad \bar{d}_t^{U(2)} = \sum_{i=1}^{1000} \frac{(\widehat{MSE}_{i,t}^{U} - MSE_t^U)^2}{1000}$$

$$\text{Mult.:} \quad \bar{d}_t^M = \sum_{i=1}^{1000} \frac{\widehat{MSE}_{i,t}^{M} - MSE_t^M}{1000}, \quad \bar{d}_t^{M(2)} = \sum_{i=1}^{1000} \frac{(\widehat{MSE}_{i,t}^{M} - MSE_t^M)^2}{1000}.$$

The bias may be considered relative to the true value for each $t = 1, \ldots, T$. The mean over T can then be evaluated and expressed as a percentage. It is termed the 'mean percent relative bias', and together with the 'mean percent relative root mean squared error' are defined in Pfeffermann & Tiller (2005, p904). They are given here for both approaches.

Univariate:

$$\text{Rel-Bias} = \frac{100}{T} \sum_{t=1}^{T} \left(\frac{\bar{d}_t^U}{MSE_t^U} \right), \text{ Rel-RMSE} = \frac{100}{T} \sum_{t=1}^{T} \left(\frac{\sqrt{\bar{d}_t^{U(2)}}}{MSE_t^U} \right); \tag{11}$$

Multivariate:

$$\text{Rel-Bias} = \frac{100}{T} \sum_{t=1}^{T} \left(\frac{\bar{d}_t^M}{MSE_t^M} \right), \text{ Rel-RMSE} = \frac{100}{T} \sum_{t=1}^{T} \left(\frac{\sqrt{\bar{d}_t^{M(2)}}}{MSE_t^M} \right). \tag{12}$$

The naïve bias is calculated by Quenneville & Singh (2000) for a local level model (which does not include a seasonal component) for series of lengths $T = 40$ and $T = 100$.

They found that the naïve bias is greater in absolute terms for the moderate length ($T = 40$) than for the longer series ($T = 100$). For the moderate length, the mean percent relative bias for 1000 realisations was reported to be -21.2%, whereas for the series with $T = 100$, the relative bias was -9.0%.

Pfeffermann & Tiller (2005) reproduce the experiment carried out by Quenneville & Singh (2000) with series of lengths $T = 40$ and $T = 100$ but extend the number of realisations to 5000. Their results for the naïve bias are slightly smaller, with the mean percent relative naïve bias for $T = 40$ reported as -18.5% and for $T = 100$, it is -7.6%.

Although these results are for a simpler model than is studied here, they show the effect of using the estimated parameters in the calculation of the predicted mean squared error of the state vector. It is expected that for this study, the naïve bias will be large and negative and that it will decrease (in absolute terms) as the length of the series is increased. The univariate model being investigated here is the local level seasonal model which differs from a local level model in that it also has a dummy seasonal component. The multivariate model is also considered.

5.1.1 Results for Naïve Bias

The results for the naïve bias for the seasonal component for *Set 1* and *Set 2* for 1000 realisations of each series length are given in Table 5. The Rel-Bias and Rel-RMSE for the univariate model and multivariate model are calculated with the equations in (11) and (12) respectively.

TABLE 5 ABOUT HERE

For *Set 1*, Table 5 shows that the relative bias calculated for the univariate model ranges from -27.97 for $T = 24$, down to -5.75 for $T = 240$. Although this model is different to the local level model, the result for $T = 40$ of -19.38% is of a similar magnitude to the result of -21.2% reported in Quenneville & Singh (2000) and the result of -18.5% reported in Pfeffermann & Tiller (2005).

The relative bias calculated for the multivariate model of *Set 1* ranges from -27.57 for $T = 24$, down to -6.33 for $T = 240$. From $T = 20$ to $T = 28$ the results for the relative bias are very similar for the two approaches. It is not until $T = 40$ and $T = 120$ that a difference becomes more noticeable. That is, there is a slight increase in bias (in absolute terms) for the multivariate model compared to the univariate model. However, for *Set 1*, the Rel-RMSE is smaller for each value of T for the multivariate model. Thus, for values of T less than 40, the multivariate model performs slightly better than the univariate model.

For *Set 2*, the results in Table 5 show a relative bias which ranges from -27.04 for $T = 20$ down (in absolute terms) to -4.48 for $T = 240$. For this set, there is considerable improvement in the Rel-Bias for the multivariate model especially for $T \leq 80$. For $T < 40$, the Rel-RMSE is smaller for each value of T for the multivariate model. For $T \geq 40$, the Rel-RMSE is larger for the multivariate model, although the Rel-Bias is slightly smaller. For *Set 2*, the naïve bias is greatly improved by applying the multivariate model to series with short to moderate length.

6. Conclusion

In this paper, a simulation study showed the effects of estimating the aggregate series parameters with the univariate and multivariate methods for different series lengths. For the two examples studied, it was shown that the bias of the estimated parameters was much less for the multivariate model than for the univariate model. This was especially

the case for short to medium length time series. The relative efficiencies of the seasonally adjusted aggregated series also showed good gains for the multivariate model.

In the second stage, the relative naïve bias and the relative root mean squared error were calculated for the two sets for both the univariate and multivariate models. The results showed that the naïve bias of the variance of the seasonally adjusted series could be decreased substantially with the use of the multivariate model. This was evident from the results for *Set 2*, where the relative naïve bias was almost half of that obtained with the univariate model for short to medium length time series. There was also a slight decrease in the relative RMSE.

From this study, it can be concluded that very good gains for the accuracy of the seasonally adjusted series are possible through using the multivariate approach and that the conditions rely on the relative parameters of the seasonal and non-seasonal components of the sub-series.

Appendix

State space form for the univariate model

The state space form of the univariate LLS model as described in Section 2.1 is given by:

$$Y_t = \mathbf{Z}\alpha_t + \varepsilon_{U,\,t}, \tag{13}$$

$$\alpha_{t+1} = \mathbf{T}\alpha_t + \mathbf{G}\gamma_t, \tag{14}$$

where, for quarterly data (s=4), and a dummy seasonal component,

$$\alpha_t = [L_t,\ S_t,\ S_{t-1},\ S_{t-2}]^\top, \qquad \alpha_1 \sim N(\mathbf{a}_1, \mathbf{P}_1),$$

$$\gamma_t = [\eta_{U,\,t},\ \omega_{U,\,t}]', \qquad \gamma_t \sim N(0, \mathbf{Q}),$$

$$\mathbf{Z} = \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix}, \qquad \varepsilon_{U,\,t} \sim N(0, \mathrm{H}),$$

$$\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} \sigma^2_{U,\,\eta} & 0 \\ 0 & \sigma^2_{U,\,\omega} \end{pmatrix},$$

$$\mathrm{Var}\,(\mathbf{G}\gamma_t) = \mathbf{G}\mathbf{Q}\mathbf{G}^\top = \begin{pmatrix} \sigma^2_{U,\,\eta} & 0 & 0 & 0 \\ 0 & \sigma^2_{U,\,\omega} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathrm{H} = \sigma^2_{U,\,\varepsilon}. \tag{15}$$

The standard set of filtering equations which make up the Kalman filter may be found in Chapter 4 of Durbin & Koopman (2001). These equations are applied to the aggregate

series data and yield the estimates of the components for each time point $\mathbf{a}_{t|t}$ and their MSEs, $\mathbf{P}_{t|t}$.

The variance matrix, $\mathbf{P}_1$, of the initial state vector α_1, is assumed to have the form:

$$\mathbf{P}_1 = \kappa \mathbf{P}_{\infty,1} + \mathbf{P}_{*,1},$$

where κ is a large scalar value, $\mathbf{P}_{*,1}$ is the covariance matrix of the stationary components in α_1 and $\mathbf{P}_{\infty,1}$ is the covariance matrix of the non-stationary components in α_1 (Zivot & Wang, 2006).

In particular, for the univariate local level seasonal BSM with a dummy seasonal component model, $\mathbf{a}_1 = E(\alpha_1)$ is a 4×1 zero vector, $\mathbf{P}_{\infty,1}$ is a 4×4 identity matrix and $\mathbf{P}_{*,1}$ is a 4×4 zero matrix. The term 'zero vector' is a vector in which each element is zero, and 'zero matrix' is used to describe a matrix in which each element is zero. The exact initial Kalman filter as described in detail in Koopman & Durbin (2000) can be applied using the `S+FinMetrics` software, in the *SsfPack* set of functions Koopman et al. (1999).

State space form for the multivariate model

The state space form for the multivariate BSM allows for the unconventional format of having correlated measurement errors. This is due to the common disturbance term, ε_t in the model. To enable use of standard software packages, Durbin & Koopman (2001) suggest including the measurement errors in the state vector. The state space form of the multivariate LLS model as described in Section 2.2 is given by:

$$\begin{aligned}
\mathbf{Y}_{(m),\,t} &= (\mathbf{Z}_{(m)} \otimes \mathbf{I}_K)\alpha_{(m),\,t}, \\
\alpha_{(m),\,t+1} &= (\mathbf{T}_{(m)} \otimes \mathbf{I}_K)\alpha_{(m),\,t} + (\mathbf{G}_{(m)} \otimes \mathbf{I}_K)\gamma_{(m),\,t},
\end{aligned}$$

where $\otimes$ is the Kronecker product. For quarterly data (s=4), and a dummy seasonal component,

$$
\begin{aligned}
\mathbf{Y}_{(m),\,t} &= [Y_{1t}, \ldots, Y_{Kt}]^{\top}, \\
\alpha_{(m),\,t} &= [L_{1t}, \ldots, L_{Kt},\ S_{1t}, \ldots, S_{Kt},\ S_{1,\,t-1}, \ldots, S_{K,\,t-1},\ S_{1,\,t-2}, \ldots, S_{K,\,t-2}, \\
&\qquad (\varepsilon_t + \varepsilon^*_{1t}), \ldots, (\varepsilon_t + \varepsilon^*_{Kt})\ ]^{\top}, \\
\gamma_{(m),\,t} &= [\ (\eta_t + \eta^*_{1t}), \ldots, (\eta_t + \eta^*_{Kt}),\ (\omega_t + \omega^*_{1t}), \ldots, (\omega_t + \omega^*_{Kt}), \\
&\qquad (\varepsilon_{t+1} + \varepsilon^*_{1,\,t+1}), \ldots, (\varepsilon_{t+1} + \varepsilon^*_{K,\,t+1})\ ]^{\top}.
\end{aligned}
$$

The system matrices are given by

$$
\mathbf{Z}_{(m)} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \end{pmatrix}, \qquad \mathbf{G}_{(m)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \mathbf{T}_{(m)} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},
$$

and hence the covariance matrix of the multivariate system is

$$
\mathrm{Var}\left((\mathbf{G}_{(m)} \otimes \mathbf{I}_K)\gamma_{(m),\,t}\right) = \begin{pmatrix} \boldsymbol{\Sigma}_\eta & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \boldsymbol{\Sigma}_\omega & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \boldsymbol{\Sigma}_\varepsilon \end{pmatrix},
$$

where $\mathbf{0}_K$ represents a $K \times K$ matrix of zeros. If $K = 2$, the covariance matrix for the

level component, $\mathbf{\Sigma}_{(m),\,\varepsilon}$, is given by (16),

$$\mathbf{\Sigma}_{(m),\,\varepsilon} = \begin{pmatrix} \sigma_\varepsilon^2 + \sigma_{1\varepsilon^*}^2 & \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 & \sigma_\varepsilon^2 + \sigma_{2\varepsilon^*}^2 \end{pmatrix} \tag{16}$$

and similarly for $\mathbf{\Sigma}_{(m),\,\eta}$ and $\mathbf{\Sigma}_{(m),\,\omega}$, where (m) denotes the multivariate model.

The state space model for the transformed system, $\mathbf{Y}_{(M),\,t}$ (8), with $K = 2$ may be specified as follows, given that the measurement errors are placed within the state vector:

$$\begin{aligned} \mathbf{Y}_{(M),\,t} &= (\mathbf{Z}_{(m)} \otimes \mathbf{I}_2)\alpha_{(M),\,t}, \\ \alpha_{(M),\,t+1} &= (\mathbf{T}_{(m)} \otimes \mathbf{I}_2)\alpha_{(M),\,t} + (\mathbf{G}_{(m)} \otimes \mathbf{I}_2)\gamma_{(M),\,t}, \end{aligned} \tag{17}$$

where $\mathbf{I}_2$ is a 2×2 identity matrix and

$$\begin{aligned} \alpha_{(M),\,t} &= [L_{tot,\,t}, L_{1t}, S_{tot,\,t}, S_{1t}, S_{tot,\,t-1}, S_{1,\,t-1}, S_{tot,\,t-2}, S_{1,\,t-2}, \varepsilon_{tot,\,t}, (\varepsilon_t + \varepsilon_{1t}^*)]^\top, \\ \gamma_{(M),\,t} &= \left[\, \eta_{tot,\,t}, (\eta_t + \eta_{1t}^*),\ \omega_{tot,\,t}, (\omega_t + \omega_{1t}^*), \varepsilon_{tot,\,t+1},\ (\varepsilon_{t+1} + \varepsilon_{1,\,t+1}^*)\right]^\top, \end{aligned} \tag{18}$$

with $\alpha_{(M),\,1} \sim N(\mathbf{a}_{(M),\,1}, \mathbf{P}_{(M),\,1})$.

The transformed covariance matrix for the error component is denoted by $\mathbf{\Sigma}_{(M),\,\varepsilon}$:

$$\mathbf{\Sigma}_{(M),\,\varepsilon} = \begin{pmatrix} \sigma_{tot,\varepsilon}^2 & 2\sigma_\varepsilon^2 + \sigma_{1\varepsilon^*}^2 \\ 2\sigma_\varepsilon^2 + \sigma_{1\varepsilon^*}^2 & \sigma_\varepsilon^2 + \sigma_{1\varepsilon^*}^2 \end{pmatrix}, \tag{19}$$

and similarly for $\mathbf{\Sigma}_{(M),\,\eta}$ and $\mathbf{\Sigma}_{(M),\,\omega}$.

To compensate for the restructuring of the state vector, the set up of the exact initial conditions matrices described in Durbin & Koopman (2001, Section 5.2) is amended. The

variance matrix, $\mathbf{P}_{(M),\,1}$, of the initial state vector, $\alpha_{(M),\,1}$, is given by

$$\mathbf{P}_{(M),1} = \kappa \mathbf{P}_{(M)\infty,1} + \mathbf{P}_{(M)*,1}, \tag{20}$$

The $\mathbf{P}_{(M)*,1}$ matrix in (20) holds the variance of the stationary part of $\alpha_{(M),\,1}$. It is a 10×10 zero matrix with the lower right 2×2 block diagonal replaced by the $\mathbf{\Sigma}_{(M),\,\varepsilon}$ covariance matrix. The $\mathbf{P}_{(M)\infty,1}$ matrix (also of dimension 10×10 here since $K = 2$) is an identity matrix but with the lower right 2×2 block diagonal replaced by a 2×2 zero matrix. Also, $\mathbf{a}_{(M),\,1}$ is a 10×1 zero vector. For further details of the exact initialisation of the filter refer to Koopman & Durbin (2000).

Kalman Filter

The standard set of filtering equations may be found in Chapter 4 of Durbin & Koopman (2001). For the univariate BSM in state space form, as described in (13) and (14), with corresponding system matrices (15), these are given by

$$\begin{aligned}
\mathbf{a}_{t+1|t} &= \mathbf{T}\mathbf{a}_{t|t-1} + \mathbf{K}_t \nu_t, \\
\mathbf{P}_{t+1|t} &= \mathbf{T}\mathbf{P}_{t|t-1}\mathbf{L}_t^\top + \mathbf{G}\mathbf{Q}\mathbf{G}^\top,
\end{aligned}$$

where

$$\begin{aligned}
\nu_t &= Y_t - \mathbf{Z}\mathbf{a}_{t|t-1} = \mathbf{Z}\alpha_t + \varepsilon_{U,\,t} - \mathbf{Z}\mathbf{a}_{t|t-1}, \\
F_t &= \mathrm{Var}(\nu_t) = \mathbf{Z}\mathbf{P}_{t|t-1}\mathbf{Z}^\top + \mathrm{H}, \qquad\qquad \mathrm{H} = \mathrm{Var}(\varepsilon_{U,\,t}), \\
\mathbf{K}_t &= \mathbf{T}\mathbf{P}_{t|t-1}\mathbf{Z}^\top F_t^{-1}, \\
\mathbf{L}_t &= \mathbf{T} - \mathbf{K}_t\mathbf{Z}.
\end{aligned}$$

References

Birrell, C. L., Steel, D. & Lin, Y-X. (2010). Seasonal adjustment of an aggregate series using univariate and multivariate basic structural models. *CSSM Working paper 01-10, http://cssm.uow.edu.au/publications/UOW074724.html.*

Durbin, J. & Koopman, S.J. (2001). *Time Series Analysis by State Space Methods*, vol. 24 of *Oxford Statistical Science Series.* New York: Oxford University Press.

Geweke, J. (1978). The temporal and sectoral aggregation of seasonally adjusted time series. In *Seasonal Analysis of Economic Time Series*, ed. A. Zellner, pp. 411-427. U.S. Government Printing Office.

Ghysels, E. (1997). Seasonal adjustment and other data transformations. *J. Bus. Econom. Statist.* **15**, 410-418.

Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge: Cambridge University Press.

Harvey, A.C. & Chung, C-H. (2000). Estimating the underlying change in unemployment in the UK. *J. Roy. Statist. Soc. Ser. A.* **163**, 303-339.

Harvey, A.C. & Durbin, J. (1986). The effects of seat belt legislation on British road casualties: A case study in structural time series modelling. *J. Roy. Statist. Soc. Ser. A.* **149**, 187-227.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems.

Journal of Basic Engineering, Transactions, ASMA, Series D. **82**, 35-45.

Koopman, S.J. & Durbin, J. (2000). Fast filtering and smoothing for multivariate state space models. *J. Time Ser. Anal.* **21**, 281-296.

Koopman, S.J., Shephard, N. & Doornik, J.A. (1999). Statistical algorithms for models in state space using SsfPack 2.2. *Econom. J.* **2**, 113-166.

Marshall , P. (1992). Estimating time-dependent means in dynamic models for cross-sections of time series. *Empirical Economics.* **17**, 25-33.

Pfeffermann, D. & Tiller, R. (2005). Bootstrap approximation to prediction MSE for state space models with estimated parameters. *J. Time Ser. Anal.* **26**, 893-916.

Planas, C. & Campolongo, F. (2001). The seasonal adjustment of contemporaneously aggregated series. *EUROSTAT Working Paper no. 8221011/1-Lot3.*

Quenneville, B. & Singh, A.C. (2000). Bayesian prediction mean squared error for state space models with estimated parameters. *J. Time Ser. Anal.* **21**, 219-236.

Zivot, E. & Wang, J. (2006). *Modeling Financial Time Series with S-PLUS*, 2nd edn. New York: Springer Science+Business Media Inc.

Table 1: Parameters for sub-series *Set 1* and *Set 2*.

	σ^2_{η}	$\sigma^2_{1\eta^*}$	$\sigma^2_{2\eta^*}$	σ^2_{ω}	$\sigma^2_{1\omega^*}$	$\sigma^2_{2\omega^*}$	σ^2_{ε}	$\sigma^2_{1\varepsilon^*}$	$\sigma^2_{2\varepsilon^*}$
Set 1	0.0196	0.4192	0.0023	0.0455	0.4091	0.4091	0.0393	0.8384	0.0046
Set 2	0.0258	0.0150	0.3819	0.0272	0.8325	0.0588	0.0516	0.0300	0.7638

Table 2: Medians, Means and standard errors (in italics) of parameter estimates for *Set 1* for 1000 realisations.

	Set 1	Level		Seasonal		Error	
		True=0.5		True=1.0		True=1.0	
Median	T	Mult.	Univ.	Mult.	Univ.	Mult.	Univ.
	20	0.210	0.194	0.726	0.477	1.410	1.680
	24	0.234	0.210	0.704	0.523	1.442	1.641
	28	0.261	0.238	0.768	0.576	1.438	1.649
	40	0.313	0.344	0.846	0.716	1.332	1.413
	80	0.434	0.457	0.983	0.937	1.019	1.026
	120	0.426	0.446	0.962	0.908	1.088	1.130
	240	0.472	0.481	0.999	0.984	1.003	1.020
Mean	T	Mult.	Univ.	Mult.	Univ.	Mult.	Univ.
	20	0.314	0.308	0.862	0.639	1.499	1.804
		0.011	*0.012*	*0.019*	*0.021*	*0.024*	*0.035*
	24	0.313	0.314	0.844	0.650	1.499	1.777
		0.010	*0.011*	*0.017*	*0.020*	*0.023*	*0.032*
	28	0.324	0.322	0.843	0.673	1.494	1.747
		0.009	*0.010*	*0.015*	*0.018*	*0.022*	*0.031*
	40	0.355	0.372	0.906	0.784	1.379	1.516
		0.008	*0.009*	*0.013*	*0.016*	*0.021*	*0.027*
	80	0.457	0.481	1.017	0.962	1.041	1.093
		0.007	*0.007*	*0.009*	*0.012*	*0.014*	*0.020*
	120	0.439	0.460	0.981	0.936	1.102	1.148
		0.005	*0.006*	*0.008*	*0.010*	*0.011*	*0.016*
	240	0.476	0.494	1.011	0.999	1.004	1.013
		0.004	*0.004*	*0.005*	*0.007*	*0.008*	*0.012*

Table 3: Medians, Means and standard errors (in italics) of parameter estimates for *Set 2* for 1000 realisations.

	Set 2	Level True=0.5		Seasonal True=1.0		Error True=1.0	
Median	T	Mult.	Univ.	Mult.	Univ.	Mult.	Univ.
	20	0.294	0.196	0.818	0.486	1.183	1.598
	24	0.312	0.239	0.829	0.556	1.189	1.607
	28	0.332	0.269	0.827	0.603	1.234	1.560
	40	0.370	0.349	0.863	0.740	1.181	1.462
	80	0.445	0.449	0.936	0.918	1.112	1.130
	120	0.458	0.470	0.969	0.942	1.043	1.103
	240	0.489	0.489	0.990	0.980	0.996	1.022
Mean	T	Mult.	Univ.	Mult.	Univ.	Mult.	Univ.
	20	0.400	0.336	0.911	0.679	1.271	1.766
		0.012	*0.013*	*0.017*	*0.023*	*0.022*	*0.035*
	24	0.396	0.339	0.886	0.665	1.273	1.754
		0.011	*0.011*	*0.015*	*0.019*	*0.021*	*0.032*
	28	0.398	0.354	0.875	0.688	1.281	1.694
		0.010	*0.011*	*0.013*	*0.018*	*0.019*	*0.031*
	40	0.413	0.390	0.913	0.785	1.226	1.523
		0.008	*0.009*	*0.011*	*0.016*	*0.017*	*0.028*
	80	0.456	0.474	0.952	0.934	1.132	1.150
		0.006	*0.007*	*0.008*	*0.012*	*0.012*	*0.020*
	120	0.469	0.477	0.971	0.957	1.073	1.114
		0.005	*0.006*	*0.006*	*0.009*	*0.010*	*0.016*
	240	0.486	0.496	0.995	0.992	1.022	1.022
		0.004	*0.004*	*0.004*	*0.007*	*0.07*	*0.012*

Table 4: Relative efficiency of the variances of the parameter estimates for *Set 1* and *Set 2* for 1000 realisations.

	Set 1			*Set 2*		
T	Level	Seasonal	Error	Level	Seasonal	Error
20	1.20	1.30	2.14	1.13	1.79	2.50
24	1.29	1.35	2.05	1.06	1.73	2.48
28	1.25	1.43	1.98	1.19	1.89	2.61
40	1.23	1.49	1.78	1.30	2.00	2.63
80	1.33	1.67	1.95	1.44	2.27	2.70
120	1.35	1.63	1.99	1.47	2.17	2.41
240	1.38	1.57	1.89	1.37	2.28	2.68

Table 5: Percent mean relative naïve bias and relative root mean squared error of smoothed seasonal component MSE for *Set 1* and *Set 2* for 1000 realisations with $T = 20$, 24, 28, 40, 80, 120, 240.

Naïve Bias of Seasonal component		Multivariate		Univariate	
	T	Rel-Bias	Rel-RMSE	Rel-Bias	Rel-RMSE
Set 1	20	-26.97	41.14	-27.69	44.65
	24	-27.57	40.09	-27.97	42.82
	28	-25.78	38.58	-25.53	40.74
	40	-21.84	34.15	-19.38	35.20
	80	-13.84	24.12	-14.29	27.96
	120	-7.19	18.32	-6.71	21.10
	240	-6.33	14.46	-5.75	17.01
Set 2	20	-14.34	42.09	-27.04	43.79
	24	-15.92	40.02	-26.88	42.55
	28	-14.92	39.80	-26.30	41.53
	40	-13.02	37.88	-18.60	35.15
	80	-5.83	27.78	-12.66	27.42
	120	-4.53	25.75	- 5.57	21.25
	240	-1.14	18.51	-4.48	16.55

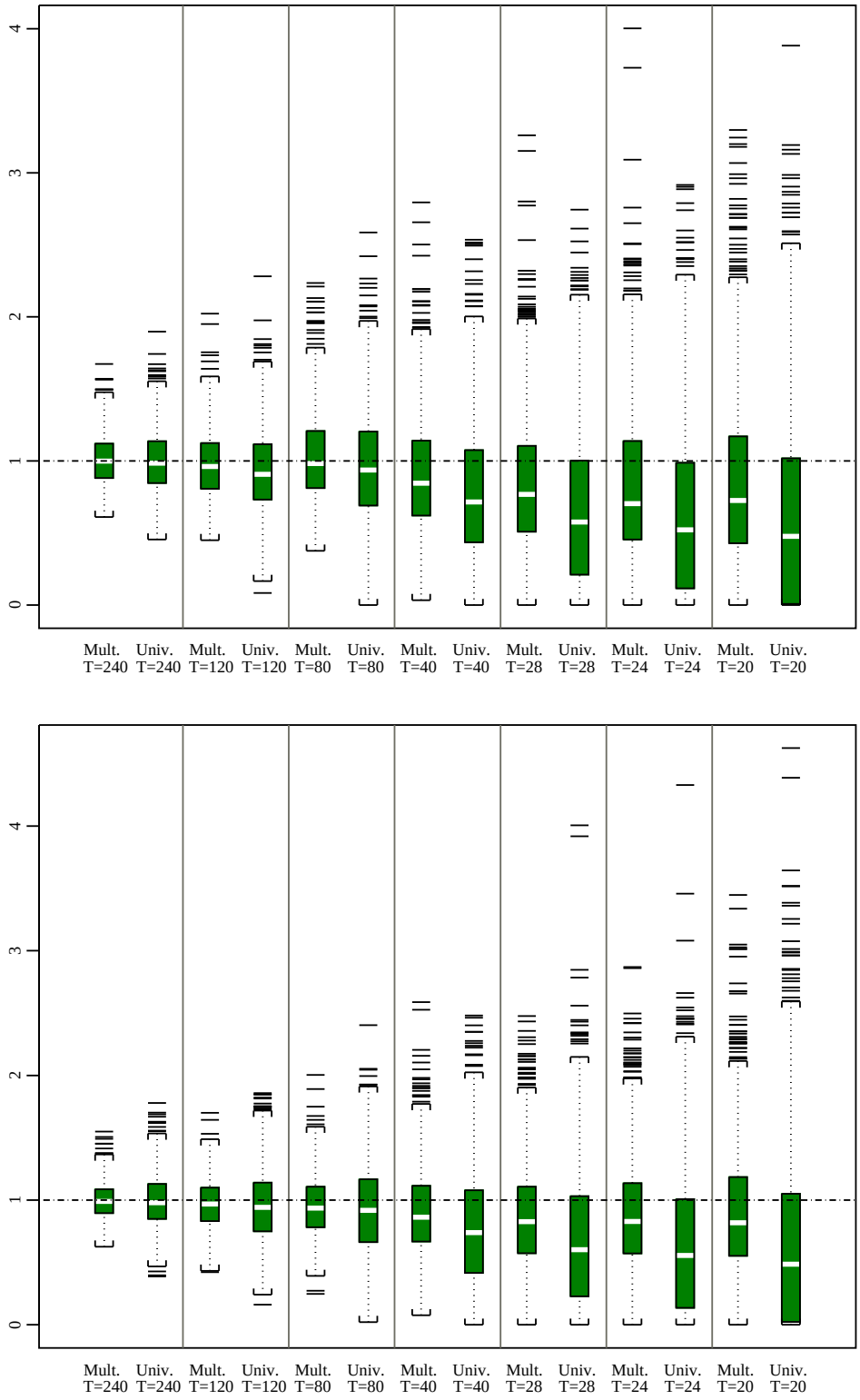

Figure 1: Distribution of the seasonal parameter estimate using the univariate and multivariate methods for 1000 realisations of *Set 1* (top) and *Set 2* (bottom) with $T = 240$, 120, 80, 40, 28, 24, 20.