WORKING PAPERS AND STUDIES

ISSN 1725-4825

2006 EDITION

Seasonal Adjustment of Aggregate Series using Univariate and Multivariate Basic Structural Models

Conference on seasonality, seasonal adjustment and their implications for short-term analysis and forecasting

Luxembourg, 10-12 May 2006

EUROPEAN COMMISSION

THEME
General
and regional
statistics

Europe Direct is a service to help you find answers to your questions about the European Union

Freephone number (*):

00 800 6 7 8 9 10 11

(*) Certain mobile telephone operators do not allow access to 00-800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server (http://europa.eu).

Luxembourg: Office for Official Publications of the European Communities, 2006

ISBN 92-79-03403-0
ISSN 1725-4825
Catalogue number: KS-DT-06-004-EN-N

Seasonal Adjustment of Aggregate Series using Univariate and Multivariate Basic Structural Models

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

Centre for Statistical and Survey Methodology
School of Mathematics and Applied Statistics
University of Wollongong,
NSW, Australia

December 2006

Abstract

Government statistical agencies are required to seasonally adjust non-stationary time series resulting from an aggregate of a number of cross-sectional time series. Traditionally, this has been achieved using the X-11 or X12-ARIMA process by using either direct or indirect seasonal adjustment. However, neither of these methods utilizes the multivariate system of time series which underlies the aggregate series. This paper compares a model-based univariate approach to seasonal adjustment with a model-based multivariate approach. Firstly, the univariate basic structural model (BSM) is applied directly to the aggregate series to obtain estimates of the seasonal components. Secondly, the multivariate basic structural model is applied to a transformed system of cross-sectional series to also obtain estimates of the seasonal components of the aggregate series. The prediction mean squared errors resulting from each method are compared by calculating their relative efficiency. An experimental study is conducted which fixes the parameters of the aggregate series and varies the stochastic structure of two cross-sectional series. It is shown that gains are achievable using the multivariate approach under certain conditions which rely on the relative values of the parameters within and between components of the cross-sectional series.

KEYWORDS: BSM, Kalman filter, multivariate time series, state space model, cross-sectional time series.

1 Introduction

Seasonally adjusted time series of economic and social data are important products of many official statistical agencies. It is imperative that the method employed produces accurate time series components such as the trend, seasonal and irregular components which make up the series. Seasonal adjustment involves estimating and removing the seasonal effects of the series. Data for a number of series is often collected, sometimes geographically or by industry, and then aggregated to obtain a total series. Seasonal adjustment of this aggregated series, as well as the cross-sectional series, is usually required for publication. The most widely used seasonal adjustment method is the X-11 filter-based method, which has progressed more recently to X-12-ARIMA. Model-based seasonal adjustment methods are recently gaining more attention and will be the focus of this paper.

This paper proposes a model-based approach for the estimation of seasonal effects for an aggregate series via a multivariate model. The aim is to examine if gains are achievable when estimating the seasonal component of the total series by jointly modeling the cross-sectional series. It is expected that under certain conditions, this procedure will enable a more accurate estimation of the seasonal component and hence the seasonally adjusted series compared to the estimation via a univariate approach.

When the seasonal component of a series is estimated from the aggregate series and then removed, the process is termed direct seasonal adjustment. Alternatively, if each of the cross-sectional series are seasonally adjusted separately, and then summed to obtain the aggregate seasonally adjusted series, the process is termed indirect seasonal adjustment. There is extensive debate (see Ghysels (1997), Hood and Findley (2003), Ladiray and Mazzi (2003), and Otranto and Triacca (2002)) as to whether to seasonally adjust directly or indirectly, usually focusing on filter-based methods. However, both direct and indirect seasonal adjustment use univariate time series methods. The indirect approach is not considered here, apart from taking into account the results from previous studies as a guide to designing this study.

Only a limited amount of research has been done using a multivariate model-based approach to seasonal adjustment. Previous research in this area has concentrated on ARIMA-based models which have restricting assumptions on the series and are also complex to implement (Planas and Campolongo (2001) and Ghysels (1997)). This study confirms and extends previous research by utilizing the flexibility and transparency of the basic structural model (BSM) in state space form, which can be applied to non-stationary time series. The multivariate state space model is a straightforward extension of the univariate model and due to its flexibility, multiple time series may be modeled jointly with little difficulty. The multivariate BSM can be applied to the modeling of cross-sectional time series with the aggregate series being the focus of attention.

A multivariate approach takes into account the links between the cross-sectional series in order to borrow strength for the estimation of the seasonal components of the total series. This paper will utilize the multivariate BSM for which the system of equations is referred to as a 'seemingly unrelated time series equations', (SUTSE) model, (Harvey, 1989, page 429). A model proposed by Marshall (1990), decomposes the disturbance terms into common and specific effects where the sub-series are linked by the variance parameter of the common effect. This model is applied to the each of the components in the multivariate BSM and then written in state space form.

The application of the Kalman filter and smoother to the state space model yields the estimates of the time series components: the trend, comprising the level and slope components, the seasonal factor and irregular (or measurement error). The Kalman filter and smoother also yield the mean squared errors of the estimated components and these may be utilized to compare the multivariate and univariate approaches. Since the sub-series are mostly collected routinely, the data are readily available. The estimates of the components for each of the sub-series are not of direct interest but are calculated in order to obtain a better estimate of the structural components of the total series.

In this paper, an experimental study is conducted which fixes the characteristics of the aggregate series and varies the stochastic structure of the sub-series. The resulting relative efficiency of the multivariate approach is examined. Section 2 gives a brief background of the seasonal adjustment approaches which involve cross-sectional series and also reviews some applications of the multivariate BSM. The BSM for the univariate and multivariate approaches and their corresponding state space forms are detailed in Section 3 and a measure of comparison is given in Section 4. The design of the study and parameter settings are outlined in Section 5. Results are presented in Section 6 with concluding remarks in Section 7.

2 Background

2.1 Seasonal Adjustment and Cross-sectional Series

A thorough discussion of the direct versus indirect methods is given in Hood and Findley (2003) with reference to the X-12-ARIMA and the SEATS (Signal Extraction in ARIMA Time Series) programs. In general, they comment that when the sub-series "have quite distinct seasonal patterns and have adjustments of good quality, indirect seasonal adjustment is usually of better quality than direct adjustment. On the other hand, when the component series have similar seasonal patterns, then summing the series may result in noise cancelation, and the direct seasonal adjustment is usually of better quality than the indirect adjustment" (Hood and Findley, 2003, p10).

Similarly, Ladiray and Mazzi (2003, p40) state that indirect adjustment should be preferred to direct adjustment if the "sub-components do not have similar characteristics or if the relative importance of the sub-series (in terms of weight) is changing very fast." They clarify the idea of similarity of sub-series by saying that the direct approach could be more suited to 'horizontal' or geographical aggregation (e.g. by country) and the indirect approach to 'vertical' or sectorial aggregation, such as by sector, branch or product.

There are a number of different criteria used in assessing the quality and adequacy of direct and indirect seasonal adjustment when using a filter-based method such as X-12-ARIMA, see Hood and Findley (2003), Ladiray and Mazzi (2003), and Otranto and Triacca (2002). In essence, the indirect adjustment is favoured when the sub-series have different characteristics and direct adjustment is favoured when the sub-series are similar.

For model-based seasonal adjustment, Planas and Campolongo (2001) used ARIMA models and found that when the stochastic properties of the two series were even slightly dissimilar, the indirect adjustment was more precise than the direct adjustment. The multivariate adjustment was found to be the most accurate estimation in terms of the final estimation error. However, when considering revision error, the indirect and direct methods performed better in many of the test cases than the multivariate method, thus giving no optimality if revision errors were considered (Planas and Campolongo, 2001). They also found multivariate estimation difficult to implement due to its complexity. Geweke (1978) found that the covariance structure between the series is crucial. He found that the joint

ARIMA model was advantageous when the sub-series are very heterogenous, or 'where the stochastic structure of the non-seasonal and seasonal components are dissimilar'.

The multivariate adjustment was found to be the most accurate estimation in terms of the final estimation error. However, when considering revision error, the indirect and direct methods performed better in many of the test cases than the multivariate method, thus giving no optimality if revision errors were considered.

2.2 The Use of the Multivariate BSM

Using the basic structural model and its multivariate state space form, a target series may be modeled jointly with one or more related series in order to obtain better estimates of the time series components of the target series. Harvey (2000) calculates the filtered estimates in a bivariate BSM model and discusses the improvement in the root mean squared error (RMSE) of the slope component relative to that obtained from using the univariate model. He shows that the gains achieved in the estimation of the slope component using the bivariate model came primarily from the high correlation between the slopes of the two series.

The U.S Bureau of Labor Statistics (BLS) apply state space models in estimating monthly employment and unemployment estimates for each of the 50 states and the District of Columbia. The models are fitted to the direct sample estimates obtained from the Current Population Survey (CPS) (Pfeffermann and Tiller, 2003). A filtering algorithm is developed for state space models with correlated measurement errors. A multivariate state space model with added constraints allows estimates to be obtained which borrow strength from both the past data and cross-sectional data (Pfeffermann and Tiller, 2003, p3).

Marshall (1992) applies the multivariate BSM to a variety of time series. He investigates the relative efficiency of the filtered estimates as a function of both time and the number of series. He concentrates on the estimation of the time-dependent means by using the Kalman filter applied to the multivariate local level model, and reports that there are gains for all the series.

Harvey and Durbin (1986) modeled the effect of seat belt legislation in Britain. They used intervention analysis based on a structural time series model to estimate the changes in casualty rates for different categories of road users following the legislation. Within the BSM they used the trigonometric approach to modeling the seasonality and incorporated an intervention variable to model the change in the level of the series after the introduction of the law. They comment on the flexibility of the structural model, which allows a considerable amount of complexity to be accommodated within the model such as explanatory and intervention variables. The model is written in state space form and the Kalman filter is applied for estimation of the time series components, the explanatory and intervention variables.

Durbin and Koopman (2001) also investigate the use of the structural model with intervention variables for the seat belt data. They extend the study to a bivariate model for the front seat passengers and rear seat passengers killed and seriously injured. The rear seat series is used as the related series but in these circumstances it is also used as a control group. The bivariate model results in a more precise measure of the effect of the seat belt law on front seat passengers. In fact, the results show that the root mean squared error is almost halved for the estimated intervention coefficient for the front seat series.

Sridharan, Vujic, and Koopman (2003) examine the impact of new legislation regarding sentencing of felony offenders on reported crime rates committed on or after January 1, 1995. The multivariate BSM allows simultaneous consideration of a set of time series where some series are considered as treatment series and others as control series. When correlations between the components are high, it means that components will be estimated with the combined use of more time series rather than just using the univariate series individually. Sridharan, Vujic, and Koopman (2003) found that this procedure leads to a more effective intervention analysis than other more traditional methods. The data only included four years of post-intervention observations and the intervention variable was modeled as a step variable to account for a level shift.

The flexibility of the multivariate structural time series model and the results of the applications described above, where gains have been achieved with the joint modeling of series motivates the current study.

3 Basic Structural Model

A structural time series model allows time series characteristics such as trend, seasonal and error components to be modeled specifically. The series of observations of the aggregate series, $Y_1, \ldots, Y_T$, will be modeled by a univariate additive BSM. If the aggregate series, denoted by Y_t, is a sum of K sub-series, for $t = 1 \ldots T$,

$$Y_t = \sum_{k=1}^{K} Y_{kt} \tag{1}$$

then $Y_{1t}, \ldots, Y_{Kt}$, may be modeled jointly with a multivariate BSM. The following subsections describe the univariate and multivariate BSM models which will be used in this study.

3.1 Univariate BSM

For a single additive time series, the population observations at time t denoted by Y_t, may be written as the sum of a local linear trend, L_t, a seasonal component, S_t, and an irregular or disturbance term, $\varepsilon_{U,\,t}$, where the U subscript denotes the univariate model. This basic structural model (BSM) may be written, in the notation adopted by Feder (2001), for $t = 1, \ldots, T$ as

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

The trend level may be assumed to evolve stochastically over time, and may or may not include a slope term, (for details see Harvey (1989, section 2.3)). For this study, the trend is represented by:

$$L_{t+1} = L_t + \eta_{U,\,t}, \qquad \eta_t \sim N(0, \sigma^2_{U,\,\eta}). \tag{3}$$

The seasonal component, S_t, may be a simple dummy variable constrained to add to zero over s seasons, $\sum_{j=1}^{s} S_j$, or over any s time periods, $\sum_{j=0}^{s-1} S_{t+1-j}$. If the seasonal effects are allowed to change stochastically over time then a disturbance may be introduced such that

$$\sum_{j=0}^{s-1} S_{t+1-j} = \omega_{U,\,t} \qquad \text{or} \qquad S_{t+1} = -\sum_{j=1}^{s-1} S_{t+1-j} + \omega_{U,\,t} \tag{4}$$

where $\omega_{U,\,t} \sim N(0, \sigma^2_{U,\,\omega})$. Since the disturbance term has an expectation of zero, equation (4) still allows the sum constraint of the seasonal effects to be zero over any s time periods.

Alternatively, the seasonal component may be modeled as a trigonometric variable (see Harvey (1989)). Other ways of expressing the seasonal component have been suggested. One method extends the dummy seasonal model (4) by letting each season evolve as a random walk, see Harrison and Stevens (1976). For a comprehensive review of modeling seasonal effects see Proietti (2000).

For this paper, the focus will be on the local level seasonal model, which is given by equations (2), (3) with (4) and will be the univariate model adopted for the aggregate series. The disturbance terms $\eta_{U,\,t}, \omega_{U,\,t}$ and $\varepsilon_{U,\,t}$, are assumed to be serially and mutually independent. The set of their variances $\{\sigma^2_{U,\,\eta},\ \sigma^2_{U,\,\omega},\ \sigma^2_{U,\,\varepsilon}\}$ are the parameters of the univariate model.

3.2 Multivariate BSM

A multivariate BSM will be applied to the K sub-series, $Y_{1t}, \ldots, Y_{Kt}$ for $t = 1 \ldots T$. The series may be linked via correlations of the disturbances driving each component. By modeling the sub-series jointly, these correlations are included as part of the structure of the covariance matrix for each component. Harvey (1989, Section 8.2) refers to this as 'contemporaneous correlation' and the model becomes a SUTSE model.

For a multivariate BSM, Marshall (1992) decomposes the disturbance terms into common effects, which are time specific, and time-unit specific effects, and relates these to the random error terms in a dynamic error components model. Applying this idea, 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 cross-sectional series with dummy seasonal components.

$$Y_{kt} = L_{kt} + S_{kt} + \varepsilon_t + \varepsilon^*_{kt} \tag{5}$$

where

$$L_{k,t+1} = L_{k,t} + \eta_t + \eta^*_{kt} \tag{6}$$

$$S_{k,t+1} = -\sum_{j=1}^{s-1} S_{k,t+1-j} + \omega_t + \omega^*_{kt} \tag{7}$$

The disturbance terms, $\varepsilon_t,\ \varepsilon^*_{kt},\ \eta_t,\ \eta^*_{kt},\ \omega_t,\ \omega^*_{kt}$ are assumed to be independent Normal random variables. The common effects are $\eta_t,\ \omega_t, \varepsilon_t$ and the time-unit specific effects are $\varepsilon^*_{kt},\ \eta^*_{kt},\ \omega^*_{kt}$.

The resulting three covariance matrices may have the following structure (Marshall, 1990):

$$\mathrm{Var}(x_t \mathbf{1}_K + \mathbf{x}^*_t) = \mathbf{\Sigma}_x = \sigma^2_x \mathbf{J}_K + \mathbf{D}_{x^*} \tag{8}$$

where

$$\begin{array}{rll} x & \text{stands for} & \eta,\ \omega,\ \text{or}\ \varepsilon \\ \mathbf{x}^*_t & \text{stands for} & (\eta^*_{1t}, \ldots, \eta^*_{Kt})',\ \ (\omega^*_{1t}, \ldots, \omega^*_{Kt})',\ \text{or}\ \ (\varepsilon^*_{1t}, \ldots, \varepsilon^*_{Kt})' \end{array}$$

$$\begin{array}{rcl} \mathbf{D}_{x^*} & = & \sigma^2_{x^*}\mathbf{I}_K \quad \text{or} \quad \mathbf{D}_{x^*} = \mathrm{diag}\big[\sigma^2_{1x^*}, \ldots, \sigma^2_{Kx^*}\big], \\ \mathbf{1}_K & \text{is} & \text{a } K \text{ dimensional unit vector,} \\ \mathbf{I}_K & \text{is} & \text{a } K \times K \text{ identity matrix,} \\ \mathbf{J}_K & = & \mathbf{1}_K \mathbf{1}'_K,\ \text{(a } K \times K \text{ matrix of all ones).} \end{array}$$

The two different structures for the covariance matrix $\mathbf{D}_{x^*}$ proposed above are named Model 1 and Model 2.

Model 1

The first and simplest covariance structure, $\mathbf{D}_{x^*} = \sigma^2_{x^*}\mathbf{I}_K$ has the unit-specific variances all equal to $\sigma^2_{x^*}$. For this model, the covariance matrix for each component has a compound symmetry structure, that is, the diagonal elements are all $(\sigma^2_x + \sigma^2_{x^*})$ and each off-diagonal element is σ^2_x.

Model 2

For Model 2, the K unit-specific variances are allowed to be different. There are $(K+1)$ unknown parameters for each of the three component covariance matrices (namely $\mathbf{\Sigma}_\eta$, $\mathbf{\Sigma}_\omega$, and $\mathbf{\Sigma}_\varepsilon$), giving a total of $3(K+1)$ unknown parameters for model. For example, when $K = 2$, the covariance matrix for the level component is

$$\mathbf{\Sigma}_\eta = \begin{pmatrix} \sigma^2_\eta + \sigma^2_{1\eta^*} & \sigma^2_\eta \\ \sigma^2_\eta & \sigma^2_\eta + \sigma^2_{2\eta^*} \end{pmatrix} \tag{9}$$

and similarly for $\mathbf{\Sigma}_\omega$, and $\mathbf{\Sigma}_\varepsilon$. Since Model 1 is a special case of Model 2, the focus will be on Model 2.

Since the aggregate series is the sum of the K cross-sectional series, the following constraint applies to the variance of the level component for the aggregate series given by the multivariate parameters:

$$\sigma^2_{tot,\eta} \quad = \quad K^2\sigma^2_\eta + \sum_{k=1}^{K} \sigma^2_{k\eta^*}. \tag{10}$$

Similarly, the constraints for the variance of the seasonal and error components are

$$\sigma^2_{tot,\omega} = K^2\sigma^2_\omega + \sum_{k=1}^{K} \sigma^2_{k\omega^*}, \qquad \sigma^2_{tot,\varepsilon} = K^2\sigma^2_\varepsilon + \sum_{k=1}^{K} \sigma^2_{k\varepsilon^*}. \tag{11}$$

The values of $\sigma^2_{tot,\eta}$, $\sigma^2_{tot,\omega}$ and $\sigma^2_{tot,\varepsilon}$ will be equal to $\sigma^2_{U,\,\eta}$, $\sigma^2_{U,\,\omega}$ and $\sigma^2_{U,\,\varepsilon}$ respectively.

3.3 State Space Form

Any BSM may be written more concisely in state space form (SSF). The Kalman filter and the Kalman smoother may then be applied to the model to obtain estimates of the components and their mean squared errors (MSE's) at each time point.

3.3.1 Univariate SSF

The state space form for the set of equations (2), (3) and (4) consists of a measurement (or observation) equation (12) and a transition (or state) equation (13). The measurement equation describes a linear combination of the unobserved components included in the state vector, α_t. The transition equation describes the development of the state vector.

The univariate state space model for the aggregate series may be written (Durbin and Koopman, 2001, section 3.1) as:

$$Y_t = \mathbf{Z}\alpha_t + \varepsilon_t \tag{12}$$
$$\alpha_{t+1} = \mathbf{T}\alpha_t + \mathbf{G}\gamma_t \tag{13}$$

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

$$\alpha_t = [L_t,\ S_t,\ S_{t-1},\ S_{t-2}]', \qquad \alpha_1 \sim N(\mathbf{a}_1, \mathbf{P}_1)$$
$$\gamma_t = [\eta_{U,\,t},\ \omega_{U,\,t}]', \qquad \gamma_t \sim N(0, \mathbf{Q})$$

$$\mathbf{Z} = \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix}, \qquad \varepsilon_{U,\,t} \sim N(0, \mathrm{H})$$
$$\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad \mathbf{G} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$
$$\mathrm{Var}\,(\mathbf{G}\gamma_t) = \mathbf{GQG}' = \begin{pmatrix} \sigma^2_{U,\,\eta} & 0 & 0 & 0 \\ 0 & \sigma^2_{U,\,\omega} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \mathrm{H} = \sigma^2_{U,\,\varepsilon} \tag{14}$$

It is assumed that the initial state vector, α_1, has a mean and variance given by $\mathrm{E}(\alpha_1) = \mathbf{a}_1$, and $\mathrm{Var}(\alpha_1) = \mathbf{P}_1$.

In general, α_t is a $p \times 1$ vector, where p is the number of unobserved components to be estimated. A local level seasonal BSM with a dummy seasonal component ($s = 4$) for quarterly data, will have $p = 1 + (s - 1) = 4$, as given above. If the data is monthly ($s = 12$), then the same model will have $p = 1 + s - 1 = 12$. The $u \times 1$ vector, γ_t, contains the disturbance terms which apply to the state vector, here $u = 2$. The $1 \times p$ matrix, $\mathbf{Z}$ is a selection matrix for the measurement equation whereas $\mathbf{T}$, ($p \times p$ matrix) and $\mathbf{G}$, ($p \times u$ matrix) apply to the transition equation.

3.3.2 Multivariate SSF

The multivariate BSM or SUTSE model (5) - (7) would usually be written in state space form in a similar way to the univariate SSF (refer to (12) and (13)), with the measurement errors separate to the state vector. This conventional format requires uncorrelated measurement errors, that is, the covariance matrix, $\boldsymbol{\Sigma}_\varepsilon$, is assumed to be diagonal. However, due to the common disturbance term, ε_t, the multivariate BSM contains correlated measurement errors, which cannot be handled by the standard Kalman filter or by standard software packages. To overcome this problem, Durbin and Koopman (2001, Section 6.4) suggest including the measurement errors in the state vector resulting in $\alpha_{(m),t}$, which has $p = 1 + (s - 1) + 1 = 5$ and $u = 3$.

The state space form is amended to allow for these different dimensions for the multivariate system of cross-sectional series $Y_{1t}, \ldots, Y_{Kt}$, as given below. The amended state space form has the (m) subscript to denote the multivariate model.

$$\mathbf{Y}_{(m),t} = (\mathbf{Z}_{(m)} \otimes \mathbf{I}_K)\alpha_{(m),t} \tag{15}$$
$$\alpha_{(m),t+1} = (\mathbf{T}_{(m)} \otimes \mathbf{I}_K)\alpha_{(m),t} + (\mathbf{G}_{(m)} \otimes \mathbf{I}_K)\gamma_{(m),t} \tag{16}$$

For quarterly data, and a dummy seasonal component in Model 2,

$$\begin{aligned}
\mathbf{Y}_{(m),t} &= [Y_{1t}, \ldots, Y_{Kt}]' \\
\alpha_{(m),t} &= [L_{1t}, \ldots, L_{Kt},\ S_{1t}, \ldots, S_{Kt},\ S_{1,t-1}, \ldots, S_{K,t-1}, \\
&\qquad S_{1,t-2}, \ldots, S_{K,t-2},\ (\varepsilon_t + \varepsilon^*_{1t}), \ldots, (\varepsilon_t + \varepsilon^*_{Kt})\ ]' \\
\gamma_{(m),t} &= [\ (\eta_t + \eta^*_{1t}), \ldots, (\eta_t + \eta^*_{Kt}),\ (\omega_t + \omega^*_{1t}), \ldots, (\omega_t + \omega^*_{Kt}), \\
&\qquad (\varepsilon_{t+1} + \varepsilon^*_{1,t+1}), \ldots, (\varepsilon_{t+1} + \varepsilon^*_{K,t+1})\ ]'
\end{aligned} \tag{17}$$

The system matrices are given by

$$\mathbf{Z}_{(m)} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{T}_{(m)} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \mathbf{G}_{(m)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{18}$$

$$Var\left((\mathbf{G}_{(m)} \otimes \mathbf{I}_K)\gamma_{(m),t}\right) = \begin{pmatrix} \mathbf{\Sigma}_\eta & 0 & 0 & 0 & 0 \\ 0 & \mathbf{\Sigma}_\omega & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{\Sigma}_\varepsilon \end{pmatrix} \tag{19}$$

If $K = 2$, the covariance matrix for the level component, $\mathbf{\Sigma}_\eta$, is given by (9) and similarly for $\mathbf{\Sigma}_\omega$ and $\mathbf{\Sigma}_\varepsilon$.

3.3.3 Applying a Transformation

The aggregate series is the series of interest here rather than the individual sub-series. In order to straightforwardly estimate the components of the aggregate series within the multivariate framework, a transformation on the multivariate state space model is required. The transformation allows the aggregate series to be included as one of the multivariate series. Let $\mathbf{A}$ be a $K \times K$ transformation matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 & 1 \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix} = \left(\begin{array}{c|c} 1\ 1 \quad \ldots & 1 \\ \hline & 0 \\ \mathbf{I}_{(K-1)} & \vdots \\ & 0 \end{array}\right) \tag{20}$$

Using $\mathbf{A}$ to obtain the transformed data, the aggregate series becomes augmented to the set comprising of series 1 to series $(K-1)$. This data set will be referred to as the 'transformed' data.

$$\mathbf{A}\left(Y_{1,t}, Y_{2,t}, \ldots, Y_{K,t}\right)' = \left(Y_{tot,t}, Y_{1,t}, \ldots, Y_{K-1,t}\right)'. \tag{21}$$

Applying the transformation to the state space model in (15) and (16), gives

$$\begin{aligned}
\mathbf{Y}_{(M),t} = \mathbf{A}\mathbf{Y}_{(m),t} &= \mathbf{A}(\mathbf{Z}_{(m)} \otimes \mathbf{I}_K)\alpha_{(m),t} \\
&= (\mathbf{Z}_{(m)} \otimes \mathbf{I}_K)(\mathbf{I}_p \otimes \mathbf{A})\alpha_{(m),t} \\
&= (\mathbf{Z}_{(m)} \otimes \mathbf{I}_K)\alpha_{(M),t}
\end{aligned} \tag{22}$$

$$\begin{aligned}
\alpha_{(M),t+1} &= (\mathbf{I}_p \otimes \mathbf{A})\alpha_{(m),t+1} \\
&= (\mathbf{T}_{(m)} \otimes \mathbf{I}_K)(\mathbf{I}_p \otimes \mathbf{A})\alpha_{(m),t} + (\mathbf{I}_p \otimes \mathbf{A})(\mathbf{G}_{(m)} \otimes \mathbf{I}_K)\gamma_{(m),t} \\
&= (\mathbf{T}_{(m)} \otimes \mathbf{I}_K)\alpha_{(M),t} + (\mathbf{I}_p\mathbf{G}_{(m)} \otimes \mathbf{A}\mathbf{I}_K)\gamma_{(m),t} \\
&= (\mathbf{T}_{(m)} \otimes \mathbf{I}_K)\alpha_{(M),t} + (\mathbf{G}_{(m)} \otimes \mathbf{I}_K)(\mathbf{I}_u \otimes \mathbf{A})\gamma_{(m),t} \\
&= (\mathbf{T}_{(m)} \otimes \mathbf{I}_K)\alpha_{(M),t} + (\mathbf{G}_{(m)} \otimes \mathbf{I}_K)\gamma_{(M),t}.
\end{aligned} \tag{23}$$

The matrices $\mathbf{Z}_{(m)}$, $\mathbf{T}_{(m)}$ and $\mathbf{G}_{(m)}$ from (18) remain unchanged. However $\alpha_{(m),t}$, and $\gamma_{(m),t}$ are renamed with the (M) subscript to allow for the transformed elements, see (26) - (27) below. The transformed model has the state space form:

$$\mathbf{Y}_{(M),t} = \mathbf{Z}_{(M)}\alpha_{(M),t} \tag{24}$$

$$\alpha_{(M),t+1} = \mathbf{T}_{(M)}\alpha_{(M),t} + \mathbf{G}_{(M)}\gamma_{(M),t} \tag{25}$$

where

$$\begin{aligned}
\mathbf{Z}_{(M)} &= \mathbf{Z}_{(m)} \otimes \mathbf{I}_K, \qquad \mathbf{T}_{(M)} = \mathbf{T}_{(m)} \otimes \mathbf{I}_K, \qquad \mathbf{G}_{(M)} = \mathbf{G}_{(m)} \otimes \mathbf{I}_K \\
\alpha_{(M),t} &= [L_{tot,t}, L_{1,t}, \ldots, L_{K-1,t},\ S_{tot,t}, S_{1,t}, \ldots, S_{K-1,t}, \\
&\quad S_{tot,t-1}, S_{1,t-1}, \ldots, S_{K-1,t-1},\ S_{tot,t-2}, S_{1,t-2}, \ldots, S_{K-1,t-2}, \\
&\quad \varepsilon_{tot,t}, (\varepsilon_t + \varepsilon^*_{1t}), \ldots, (\varepsilon_t + \varepsilon^*_{K-1,t})\ ]'
\end{aligned} \tag{26}$$

$$\begin{aligned}
\gamma_{(M),t} &= [\ \eta_{tot,t}, (\eta_t + \eta^*_{1t}), \ldots, (\eta_t + \eta^*_{K-1,t}), \\
&\quad \omega_{tot,t}, (\omega_t + \omega^*_{1t}), \ldots, (\omega_t + \omega^*_{K-1,t}), \\
&\quad \varepsilon_{tot,t+1}, (\varepsilon_{t+1} + \varepsilon^*_{1,t+1}), \ldots, (\varepsilon_{t+1} + \varepsilon^*_{K-1,t+1})\ ]'
\end{aligned} \tag{27}$$

where

$$\eta_{tot,t} = K\eta_t + \sum_{k=1}^{K} \eta^*_{kt} \qquad \omega_{tot,t} = K\omega_t + \sum_{k=1}^{K} \omega^*_{kt} \qquad \varepsilon_{tot,t} = K\varepsilon_t + \sum_{k=1}^{K} \varepsilon^*_{kt} \tag{28}$$

$$\gamma_{(M),t} \sim N(0, \mathbf{Q}_{(M)})$$

$$\text{Var}\left(\mathbf{G}_{(M)}\gamma_{(M),t}\right) = \mathbf{G}_{(M)}\mathbf{Q}_{(M)}\mathbf{G}'_{(M)} = \begin{pmatrix} \mathbf{\Sigma}_{(M),\eta} & 0 & 0 & 0 & 0 \\ 0 & \mathbf{\Sigma}_{(M),\omega} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{\Sigma}_{(M),\varepsilon} \end{pmatrix} \tag{29}$$

If $K = 2$, the covariance matrix for the level component is given by

$$\mathbf{\Sigma}_{(M),\eta} = \mathbf{A}\mathbf{\Sigma}_\eta\mathbf{A}' = \begin{pmatrix} \sigma^2_{tot,\eta} & 2\sigma^2_\eta + \sigma^2_{1\eta^*} \\ 2\sigma^2_\eta + \sigma^2_{1\eta^*} & \sigma^2_\eta + \sigma^2_{1\eta^*} \end{pmatrix} \tag{30}$$

with $\sigma^2_{tot,\eta} = 4\sigma^2_\eta + \sigma^2_{1\eta^*} + \sigma^2_{2\eta^*}$ from (10). Similarly for $\mathbf{\Sigma}_{(M),\omega}$ and $\mathbf{\Sigma}_{(M),\varepsilon}$.

3.4 Application of the Kalman Filter

A linear Gaussian state space model may be analysed by applying the Kalman filter and Kalman smoother to the observations. The Kalman filter provides the optimal estimator of the state vector at time t, taking into account observations up to time t, via a forward recursion. Denote the information provided by $Y_1, Y_2, \dots, Y_t$, as $\mathcal{F}_t$ when $t < T$. The Kalman smoother further improves the component estimates and provides the optimal estimator of the state vector at time $t < T$, taking into account all the observations in the sample, $Y_1, Y_2, \dots, Y_T$.

The properties of the components in the state vector, conditional on the set $\mathcal{F}_t$, are (Durbin and Koopman, 2001):

$$\mathbf{a}_{t+1} = \mathrm{E}(\alpha_{t+1}|\mathcal{F}_t) \qquad \text{and} \qquad \mathbf{P}_{t+1} = \mathrm{Var}(\alpha_{t+1}|\mathcal{F}_t). \tag{31}$$

3.4.1 Kalman Filter for the Univariate Model

The standard set of filtering equations may be found in Chapter 4 of Durbin and Koopman (2001). For the univariate local level seasonal BSM in state space form, as described in (12) and (13), with corresponding system matrices (14), these are given by:

$$\begin{aligned} \mathbf{a}_{t+1} &= \mathbf{T}\mathbf{a}_t + \mathbf{K}_t\nu_t \\ \mathbf{P}_{t+1} &= \mathbf{T}\mathbf{P}_t\mathbf{L}_t^{'} + \mathbf{G}\mathbf{Q}\mathbf{G}^{'} \end{aligned} \tag{32}$$

where

$$\begin{aligned} \nu_t &= Y_t - \mathbf{Z}\mathbf{a}_t = \mathbf{Z}\alpha_t + \varepsilon_{U,\,t} - \mathbf{Z}\mathbf{a}_t \\ \mathbf{F}_t &= \mathrm{Var}(\nu_t) = \mathbf{Z}\mathbf{P}_t\mathbf{Z}^{'} + \mathrm{H}, \qquad\qquad H = \mathrm{Var}(\varepsilon_t) \\ \mathbf{K}_t &= \mathbf{T}\mathbf{P}_t\mathbf{Z}^{'}\mathbf{F}_t^{-1} \\ \mathbf{L}_t &= \mathbf{T} - \mathbf{K}_t\mathbf{Z}. \end{aligned} \tag{33}$$

The state vector estimator, $\mathbf{a}_{t|t}$, and its corresponding error variance matrix, $\mathbf{P}_{t|t}$, are

$$\begin{aligned} \mathbf{a}_{t|t} &= \mathrm{E}(\alpha_t|\mathcal{F}_t) = \mathbf{a}_t + \mathbf{M}_t\mathbf{F}_t^{-1}\nu_t \\ \mathbf{P}_{t|t} &= \mathrm{Var}(\alpha_t|\mathcal{F}_t) = \mathbf{P}_t - \mathbf{M}_t\mathbf{F}_t^{-1}\mathbf{M}_t^{'} \end{aligned} \tag{34}$$

where $\mathbf{M}_t = \mathbf{P}_t\mathbf{Z}^{'}$, and then the updating equations are defined as

$$\begin{aligned} \mathbf{a}_{t+1} &= \mathbf{T}\mathbf{a}_{t|t} \\ \mathbf{P}_{t+1} &= \mathbf{T}\mathbf{P}_{t|t}\mathbf{T}^{'} + \mathbf{G}\mathbf{Q}\mathbf{G}^{'} \end{aligned} \tag{35}$$

The application of the Kalman filter is performed with the SPLUS FinMetrics software using the set of functions collectively called the SsfPack, (Koopman, Shephard, and Doornik, 1999). The variance matrix, $\mathbf{P}_1$, of the initial state vector α_1, is assumed to have the form

$$\mathbf{P}_1 = \kappa\mathbf{P}_\infty + \mathbf{P}_* \tag{36}$$

where κ is a scalar value, $\mathbf{P}_*$ is the covariance matrix of the stationary components in α_1 and $\mathbf{P}_\infty$ is the covariance matrix of the non-stationary components in α_1, (Zivot, Wang, and Koopman, 2004). Diffuse initial conditions are handled with the exact initial Kalman filter where the filter equations are derived as $\kappa \to \infty$. This approach is described in Durbin and Koopman (2001, Section 5.2).

In particular, for the univariate local level seasonal model, $\mathbf{a}_1 = E(\alpha_1)$ is a 4×1 zero vector, $\mathbf{P}_\infty$ is a 4×4 identity matrix and $\mathbf{P}_*$ is a 4×4 zero matrix.

3.4.2 Kalman Filter for the Transformed Model

The Kalman filter equations (32) to (35) need to be amended for the state space form given by (24) and (25), where the measurement error has been placed within the state vector, and corresponding system matrices $(\mathbf{Z}_{(M)}, \mathbf{T}_{(M)}, \mathbf{G}_{(M)}, \mathbf{Q}_{(M)})$ are included in (26) - (29). Only one equation listed in (33) requires modification, (apart from subscripts). That is the equation for $\mathbf{F}_t$, which becomes

$$\mathbf{F}_{(M),\,t} = \mathbf{Z}_{(M)}\mathbf{P}_{(M),\,t}\mathbf{Z}^{'}_{(M)} \tag{37}$$

since H=0, and the $\mathbf{Q}_{(M)} = \mathrm{Var}(\gamma_{(M),\,t})$ matrix now includes the $\mathbf{\Sigma}_{(M),\varepsilon}$ as shown in (29).

To compensate for this restructuring of the state vector, the set up of the exact initial conditions matrices described in Durbin and Koopman (2001, Section 5.2) are amended. The $\mathbf{P}_*$ matrix which holds the variance of the stationary part of α_1, instead of being a zero matrix, now includes the $\mathbf{\Sigma}_{(M),\varepsilon}$ covariance matrix in the lower right $(K \times K)$ block diagonal. All other elements of the $(5K \times 5K)$ matrix are zero. The $\mathbf{P}_\infty$ matrix (also of dimension $5K \times 5K$) therefore becomes zero in the corresponding position and the remaining main diagonal elements are 1. For further details of the exact initialisation of the filter see Koopman and Durbin (2000).

A simplification of the multivariate filtering process is described in Durbin and Koopman (2001, Section 6.4) where the elements of the observational vectors are brought into the analysis individually. This method basically converts the multivariate series into a univariate series and allows computational savings and simplifies the initialisation process. This method is applied in the SsfPack of functions in SPLUS.

4 Comparison of Univariate and Multivariate Methods

The main focus of this paper is to determine whether the use of the sub-series improves the estimates of the unobserved components of the aggregate series and hence the seasonally adjusted aggregate series. Therefore, we will compare results from the univariate model with those from the transformed multivariate model.

4.1 Variance of the seasonally adjusted series

The question arises as to how to calculate the accuracy of the seasonally adjusted series when a state space model is applied. Harvey (1989) explains that when the optimal estimator of the seasonal component is obtained by the smoothing algorithm and subtracted from the original series to give the seasonally adjusted series,

$$Y^a_{t|T} = Y_t - \hat{S}_{t|T} \qquad t = 1, \ldots, T \tag{38}$$

then, 'the RMSE of $\hat{S}_{t|T}$, and hence $Y^a_{t|T}$, is also given by the smoother' (Harvey, 1989, p303). An advantage of the structural model-based approach to seasonal adjustment is that it estimates the variance of the seasonally adjusted series as a by-product of estimating the seasonally adjusted series (Jain, 2001).

Burridge and Wallis (1985) answer this question for the Kalman filter formulation of signal extraction methods in more detail. They note that this is applicable to non-stationary time series and for stationary series it is equivalent to the Wiener-Kolmogorov theory as applied in Planas and Campolongo (2001). They state that the appropriate measure of the accuracy of the adjusted data is the error variance of the seasonal component estimate, conditional on the data. Thus, for the current-adjusted series, $Y^a_{t|t}$, the error variance of the seasonal component estimate, conditional on the data, is given by $\text{MSE}(\hat{S}_{t|t})$. Note that the current-adjusted series can be viewed as the preliminary seasonally adjusted series as it is conditional on observations up to time t. This is the error variance given by the Kalman filter as calculated in the matrix $\mathbf{P}_{t|t}$, given by (34), for the element pertaining to the seasonal component.

In the following study, the value of $\text{MSE}(\hat{S}_{tot,t|t})$ using the transformed multivariate model, given the multivariate parameters, is denoted by $\text{MSE}(\hat{S}^M_{t|t})$. For the univariate method, given the total (or univariate) parameters, $\text{MSE}(\hat{S}_{tot,t|t})$ is denoted by $\text{MSE}(\hat{S}^U_{t|t})$. The relative efficiency ratio for the filtered estimates is given by

$$RE_{kf}(M) = \frac{\text{MSE}(\hat{S}^U_{t|t})}{\text{MSE}(\hat{S}^M_{t|t})}, \qquad t = 1 \dots T \tag{39}$$

and can be considered as a preliminary estimate of the equivalent measure which uses the MSE of the smoothed seasonal component. The $RE_{kf}(M)$ measure over t is the quantity of interest in the comparison of the multivariate method with the univariate method.

5 Design of the Study

In the discussion on direct versus indirect adjustment in Section 2.1, a number of authors agreed that when the series have similar patterns, direct adjustment is favoured and alternatively, when the series have dissimilar patterns, indirect adjustment is favoured. From the work carried out by Geweke (1978) and Planas and Campolongo (2001), particular attention needs to be given to the relationship of parameters between the sub-series and between components.

5.1 Setting the Parameters

The parameters in the multivariate state space model are the variances of the disturbance terms. These parameters are found in the covariance matrices, $\mathbf{\Sigma}_\eta$, $\mathbf{\Sigma}_\omega$ and $\mathbf{\Sigma}_\varepsilon$. The specification of their structure will be called the 'design' of the sub-series.

In order to examine the behaviour of the relative efficiency of the multivariate method with two sub-series, $(K = 2)$, the design will be varied. The series 1 parameter values need to vary from those for series 2 within components, from being the same (as given by Model 1), to being very different (as given by Model 2), as defined in Section 3.2. Also, the structure of the covariance matrices for the non-seasonal components and the seasonal component need to be considered relative to one another.

Let c be the ratio of the variances between sub-series 1 and 2 within each component. The variance of the level component for series k is given by

$$\begin{aligned} \text{Var}(L_{kt}) &= \text{Var}(\eta_t + \eta^*_{kt}) \\ &= \sigma^2_\eta + \sigma^2_{k,\eta^*}, \end{aligned}$$

then the c-ratio for the level component is defined as

$$c_\eta = \frac{\text{Var}(L_{1t})}{\text{Var}(L_{2t})} = \frac{\sigma^2_\eta + \sigma^2_{1,\eta^*}}{\sigma^2_\eta + \sigma^2_{2,\eta^*}} \tag{40}$$

Similarly, the c-ratios for the seasonal and error components are given respectively by:

$$c_\omega = \frac{\sigma_\omega^2 + \sigma_{1,\omega^*}^2}{\sigma_\omega^2 + \sigma_{2,\omega^*}^2}, \quad c_\varepsilon = \frac{\sigma_\varepsilon^2 + \sigma_{1,\varepsilon^*}^2}{\sigma_\varepsilon^2 + \sigma_{2,\varepsilon^*}^2}. \tag{41}$$

If $c_\eta = c_\omega = c_\varepsilon = 1$ then Model 2 reverts back to Model 1, which is the compound symmetry case where each diagonal element has the same value. This would mean that for each component the same properties apply between series. For the seasonal component, it does not mean that the seasonal patterns are the same, but the degree of stability of the seasonal component is the same.

To emulate the 'slightly dissimilar' case to the 'strongly dissimilar' case, as in Planas and Campolongo (2001), the values of the c-ratios will be varied. The term 'dissimilar' here applies to the stochastic behaviour of the component, ranging from almost deterministic to very stochastic. For a dissimilar case, it would mean a value of c much greater than 1 or a small fractional value.

In this study, we let the c-ratios vary in the set {1, 5, 10, 20} or their reciprocals {1, 0.2, 0.1, 0.05}. Furthermore, to set a design where 'the stochastic structure of the non-seasonal and seasonal components are dissimilar', the c-ratios would need to be dissimilar between components, and so for one component, it could be greater than one, and for another component it could be less than one.

With this in mind, a set of designs for the sub-series may be formulated and are named in the following tables. Table 1 shows Design (a) where all c-ratios are greater than or equal to one. Note that c_η and c_ε have been set to the same value in each design thereby reducing the number of combinations considered and setting the focus on the seasonal c-ratio, c_ω. Design (b) given in Table 2, has $c_\omega > 1$ but has the reciprocal of these values for c_η and c_ε.

Design (a)		c_η and c_ε			
		1	5	10	20
c_ω	1	*a11*	*a12*	*a13*	*a14*
	5	*a21*	*a22*	*a23*	*a24*
	10	*a31*	*a32*	*a33*	*a34*
	20	*a41*	*a42*	*a43*	*a44*

Table 1: Sub-Series Design (a): $c_\omega \geq 1$, and $c_\eta, c_\varepsilon \geq 1$

Design (b)		c_η and c_ε		
		0.2	0.1	0.05
c_ω	5	*b22*	*b23*	*b24*
	10	*b32*	*b33*	*b34*
	20	*b42*	*b43*	*b44*

Table 2: Sub-Series Design (b): $c_\omega > 1$, and $c_\eta, c_\varepsilon < 1$

In addition to the setting of the c-ratios, the correlation between the series due to the common disturbance term, needs to be considered for each component. For this study, the correlation values for the seasonal component, ρ_ω, is set to one of the following values {0.1, 0.3, 0.5, 0.7, 0.9}. For the level and error components, the correlation values (ρ_η and ρ_ε) considered are {0.2, 0.4, 0.6, 0.8}. The choice of these values means that the homogeneous case, where the covariance matrices are proportional to one another, (Harvey, 1989, Section 8.3) is avoided for certain combinations of the c-ratios described above. The design table, naming the correlation combinations is given in Table 3.

Correlations		ρ_η and ρ_ε				
		0.2	0.4	0.6	0.8	1.0
ρ_ω	0.1	A1	B1	C1	D1	E1
	0.3	A2	B2	C2	D2	E2
	0.5	A3	B3	C3	D3	E3
	0.7	A4	B4	C4	D4	E4
	0.9	A5	B5	C5	D5	E5

Table 3: Correlation design combinations for ρ_ω, ρ_η, and ρ_ε

For example, the design 'A1*a23*' refers to when $c_\omega = 5$, $c_\eta, c_\varepsilon = 10$, $\rho_\omega = 0.1$, and $\rho_\eta, \rho_\varepsilon = 0.2$. Not all of the correlation designs will be possible for each of the c-ratio design combinations due to the restrictions and constraints on the multivariate variance parameters which are explained in the next section.

The 9 parameters in the multivariate model may be expressed in terms of the design of the model defined by $\{c_\eta, c_\omega, c_\varepsilon, \rho_\eta, \rho_\omega, \rho_\varepsilon, \sigma^2_{tot,\eta}, \sigma^2_{tot,\omega}, \sigma^2_{tot,\varepsilon}\}$.

5.2 Application of Constraints

The aggregate series will be fixed as a case study but the design of the underlying sub-series will be varied. The variance parameters for the total series are set with

$$\sigma^2_{tot,\eta} = 0.01, \quad \sigma^2_{tot,\omega} = 1, \quad \sigma^2_{tot,\varepsilon} = 1, \tag{42}$$

and will therefore be applied to the univariate method.

With these given univariate parameters, as well as the c-ratios, the correlations, and the constraints given in (10) and (11), the multivariate parameters for each component can be determined by solving a set of simultaneous equations. For example, the seasonal component equations are:

$$\sigma^2_{tot,\omega} = 4\sigma^2_\omega + \sigma^2_{1\omega^*} + \sigma^2_{2\omega^*}, \quad c_\omega = \frac{\sigma^2_\omega + \sigma^2_{1,\omega^*}}{\sigma^2_\omega + \sigma^2_{2,\omega^*}}, \quad \rho_\omega = \frac{\sigma^2_\omega}{\sqrt{(\sigma^2_\omega + \sigma^2_{1,\omega^*})(\sigma^2_\omega + \sigma^2_{2,\omega^*})}} \tag{43}$$

which, when solved in terms of $\sigma^2_{tot,\omega}$, c_ω and ρ_ω, and noting that each of these are positive, give

$$\sigma^2_\omega = \frac{\rho_\omega\sqrt{c_\omega}\sigma^2_{tot,\omega}}{1 + c_\omega + 2\rho_\omega\sqrt{c_\omega}}, \; \sigma^2_{1,\omega^*} = \frac{\sigma^2_{tot,\omega}(c_\omega - \rho_\omega\sqrt{c_\omega})}{1 + c_\omega + 2\rho_\omega\sqrt{c_\omega}}, \; \sigma^2_{2,\omega^*} = \frac{\sigma^2_{tot,\omega}(1 - \rho_\omega\sqrt{c_\omega})}{1 + c_\omega + 2\rho_\omega\sqrt{c_\omega}}. \tag{44}$$

Since $\sigma^2_{1\omega^*} \geq 0$, $\sigma^2_{2\omega^*} \geq 0$ and $\sigma^2_\omega > 0$ then the restrictions on the correlation values are such that

$$\begin{aligned} &\text{if} \quad c_\omega \geq 1, \quad \text{then} \quad 0 < \rho_\omega \leq \frac{1}{\sqrt{c_\omega}}, \\ &\text{and if} \quad c_\omega < 1, \quad \text{then} \quad 0 < \rho_\omega \leq \sqrt{c_\omega} \end{aligned} \tag{45}$$

and similarly for the level and error components.

Given the 9 exact multivariate parameters, the data for Y_{1t} and Y_{2t} are generated from the multivariate model equations for $t = 40 + T$ as described in (5) and (7), with starting values $L_1 = 5, S_1 = -1.5, S_0 = -1, S_{-1} = 0.5$ for both series. The first 40 data points of each series are discarded, leaving the $t = 1 \ldots T$ simulated quarterly observations required. For this study, T is set to 40, giving 10 years of quarterly data. The length of the series is therefore adequate to examine the behaviour of the relative efficiency ratio over time.

The series are summed contemporaneously, $Y_{tot,t} = Y_{1t} + Y_{2t}$ to obtain the simulated aggregate series. Note that since exact parameters are applied to the model, only one realisation of the data is required in order to obtain the MSE's of the seasonal component. This is because the theoretical expressions for the MSE values only rely on the parameter values and not on the observations. Exact parameters will be applied here so that the effect of the design on the relative efficiency ratio is not obscured by the values of the estimated parameters. The effect of estimation is included in a forthcoming study.

6 Results: Effect of the parameters of the sub-series

The relative efficiency of the seasonal component calculated with the univariate (or direct) method and with the multivariate method is given by $RE_{kf}(M)$. This measure is determined for each design specified in Tables 1 and 2 using the exact parameters. To obtain an overview of these results, the same correlation combination is chosen for each design, that is A1, where $\rho_\omega = 0.1$ and $\rho_\eta, \rho_\varepsilon = 0.2$.

Figure 1 shows the results over $t = 1 \ldots 40$ for the 16 different *'a'* designs. For $t = 1 \ldots 4$, the relative efficiency is exactly one. However, from $t = 5$, gains using the multivariate method are achievable for some, but not all, of the *'a'* designs and vary in magnitude and over time. For those designs which do achieve gains, the gains climb in the next few time points to reach a steady value. The amount of time it takes for this converged value to occur depends on the design. For example, design *a41* has the largest relative efficiency at $T = 40$, $RE_{kf}(M) = 1.29$, which translates to a gain of approximately 22.7%, but has the slowest rate of convergence.

The next few highest gains are for designs *a31*, *a14* and *a13* respectively. Note that *a41* has $c_\omega = 20$ and $c_\eta, c_\varepsilon = 1$, and *a31* has $c_\omega = 10$ and $c_\eta, c_\varepsilon = 1$, both with a high c-ratio for the seasonal component. For the design *a14*, $c_\omega = 1, c_\eta, c_\varepsilon = 20$ and for design *a13*, $c_\omega = 1, c_\eta, c_\varepsilon = 10$. Thus, the four *'a'* designs which give the highest $RE_{kf}(M)$ result, have either a high between-series c-ratio for the seasonal component or a high between-series c-ratio for the non-seasonal components, but not both. The result is higher if the two c-ratios defining the design are at opposite ends of the scale. So, even when the variances for the two series are the same for the seasonal component ($c_\omega = 1$), if the variances of the non-seasonal components are largely different (high c_η, c_ε), a gain is still achievable (although not as large) for the aggregate seasonal component.

To explore the differences among the 'a' designs in more detail, we focus on $RE_{kf}(M)$ at $T = 40$ for each design, found in Table 4. The lowest results for the relative efficiency belong to the designs which have $c_\omega = c_\eta = c_\varepsilon$, namely, *a11*, *a22*, *a33*, and *a44*. Note that for *a11*, $c_\omega = c_\eta = c_\varepsilon = 1$ and this design represents the compound symmetry case, (Model 1). Even when all the c-ratios are high, as in *a33* and *a44*, where the series are largely dissimilar for all components, the fact that they are equal, overrides the between-series effect. Thus, when the c-ratios are equal, the design of the covariance matrices become closer to a homogeneous state.

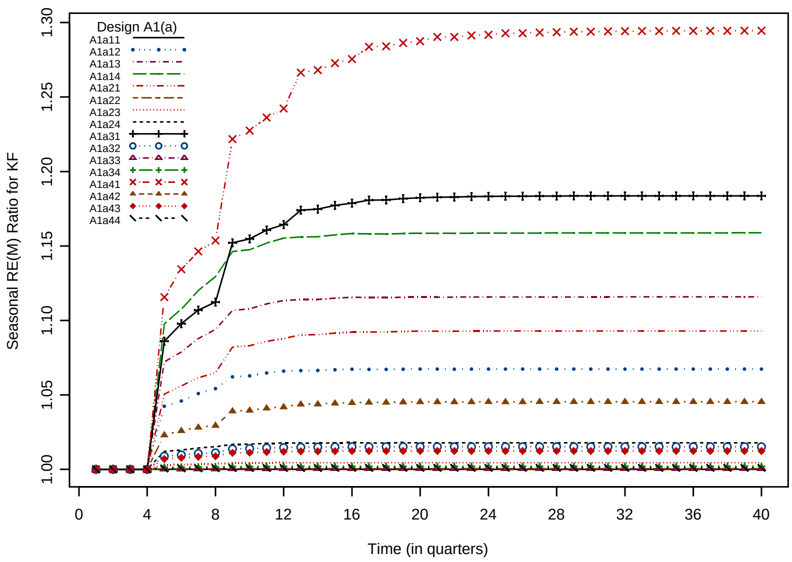


Figure 1: $RE_{kf}(M)$ for seasonal component with Design A1 for *a11* to *a44*.

Design (a)	c_η and c_ε			
	1	5	10	20
c_ω 1	*a11* 1.0000	*a12* 1.0674	*a13* 1.1158	*a14* 1.1588
5	*a21* 1.0929	*a22* 1.0005	*a23* 1.0045	*a24* 1.0178
10	*a31* 1.1837	*a32* 1.0152	*a33* 1.0008	*a34* 1.0021
20	*a41* 1.2945	*a42* 1.0454	*a43* 1.0124	*a44* 1.0010

Table 4: Results for $RE_{kf}(M)$ for seasonal component at T=40 for sub-series Design (a) with A1 ($\rho_\omega = 0.1$, and $\rho_\eta, \rho_\varepsilon = 0.2$)

The '*b*' designs use the reciprocal of the values of c_η, c_ε given in the '*a*' designs and the results over time are shown in Figure 2. The results show a similar pattern for $RE_{kf}(M)$ over time, however, the magnitude is much greater than for the '*a*' designs, with 9 designs giving an $RE_{kf}(M)$ of over 1.25. The largest gain is achieved by design *b44* ($c_\omega = 20,\ c_\eta, c_\varepsilon = 0.05$) , with $RE_{kf}(M) = 2.48$ at $T = 40$. This translates to a gain of almost 60% for the multivariate method. Again, it can be seen that the designs where c_ω is very different from c_η and c_ε, for example *b44, b43, b34*, give the highest gains. The numerical results for $T = 40$ for each '*b*' design are given in Table 5.

The correlation combination in all '*a*' and '*b*' designs discussed so far, is identical and labeled A1, where $\rho_\omega = 0.1$ and $\rho_\eta, \rho_\varepsilon = 0.2$. Therefore, even when the correlation values between the series are small, large gains are attainable with the size of the gain depending on the design structure.

Three designs have been chosen to determine the effect of increasing the seasonal correlation for the '*a*' design. Firstly, designs *a12*, *a13* and *a14* have been analysed with correlation combinations A1 to A5, which keep the non-seasonal correlation coefficient low at 0.2, while allowing the seasonal correlation to be one of $\{0.1,\ 0.3,\ 0.5,\ 0.7,\ 0.9\}$

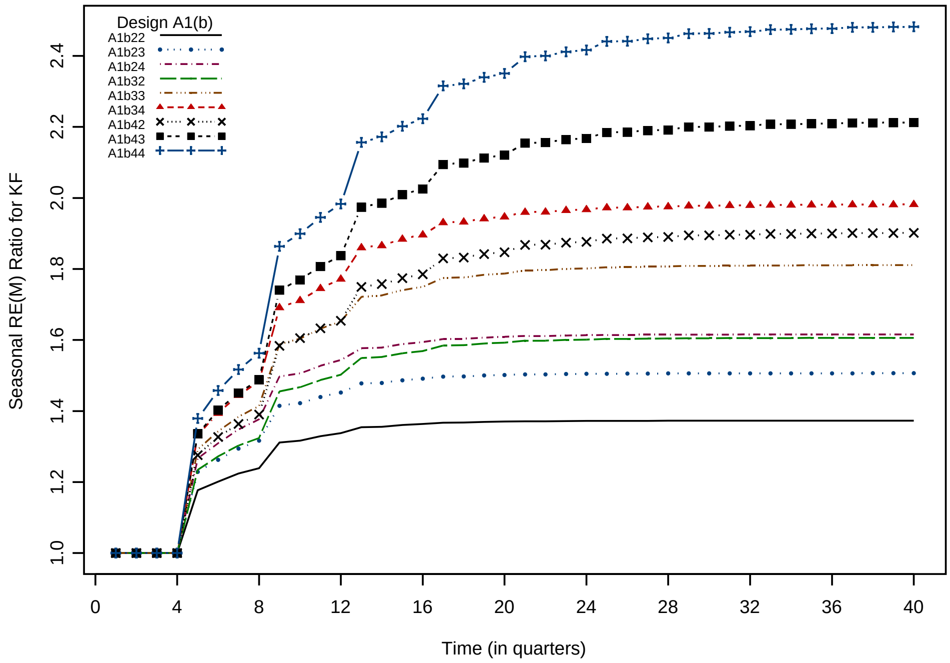
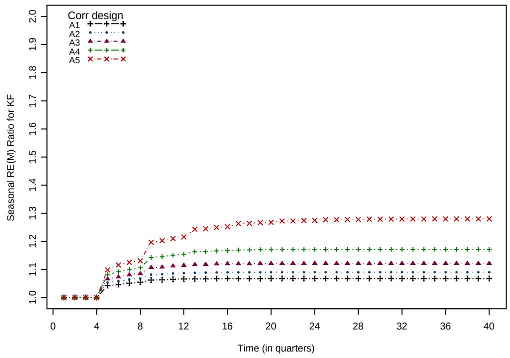


Figure 2: $RE_{kf}(M)$ for seasonal component with Design A1 for $b22$ to $b44$.

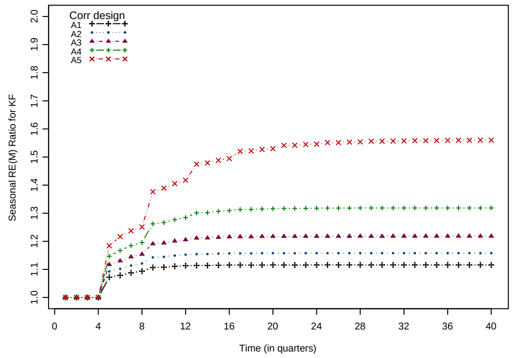
Design (b)	c_η and c_ε		
	0.2	0.1	0.05
c_ω 5	$b22$ 1.3728	$b23$ 1.5063	$b24$ 1.6158
10	$b32$ 1.6060	$b33$ 1.8108	$b34$ 1.9818
20	$b42$ 1.9014	$b43$ 2.2125	$b44$ 2.4820

Table 5: Results for $RE_{kf}(M)$ for seasonal component at T=40 for sub-series Design (b) with A1 ($\rho_\omega = 0.1$, and $\rho_\eta, \rho_\varepsilon = 0.2$)

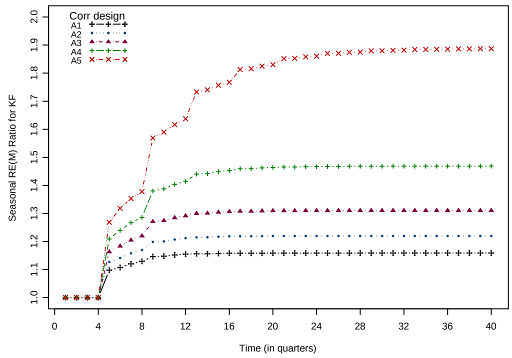
as defined in Table 3. Figure 3 shows these results for $t = 1 \ldots 40$ and with the same vertical scale. The seasonal correlation does impact the result for the $RE_{kf}(M)$ for the seasonal component as would be expected. The plots show that as the seasonal correlation increases, and the non-seasonal correlation is kept low, the relative efficiency also increases, but the increase is dependent upon the design structure, and does not appear to be in constant increments. As the non-seasonal c-ratio increases across the designs from 5 ($a12$) to 10 ($a13$) and then to 20 ($a14$), the effect of the seasonal correlation coefficient intensifies as shown in the three plots in Figure 3.



(a) Design *a12*



(b) Design *a13*



(c) Design *a14*

Figure 3: $RE_{kf}(M)$ for seasonal component for A1 - A5 for Designs *a12*, *a13* and *a14*.

By taking the results for the last time point (i.e.T=40) from each time series within each plot in Figure 3, the positive relationship between the seasonal correlation and the $RE_{kf}(M)$ value is shown more clearly. Figure 4 shows these results as well as those for design *a11*. Firstly, for the compound symmetry design *a11*, the result for the relative efficiency remains constant at 1 as the seasonal correlation increases. The gradient of the curve increases from design *a12* to design *a13* to the steepest curve for design *a14*, thus as the c_η, c_ε non-seasonal c-ratios increase from 5 to 10 to 20.

It is now appropriate to determine the effect of increasing the non-seasonal correlation whilst keeping the seasonal correlation constant. The designs *a21, a31, a41* will be analysed for correlation combinations A1 to E1. Hence, the seasonal correlation will be kept at $\rho_\omega = 0.1$, and the non-seasonal correlations ρ_η, ρ_ε will be one of $\{0.2, 0.4, 0.6, 0.8, 1.0\}$. Figure 5 shows the time series plots of the $RE_{kf}(M)$ value for the three designs *a21, a31, a41* respectively with the same scale on the vertical axis. Here it is the non-seasonal correlation which is increasing, and the relative efficiency of the seasonal component is still affected. For design *a21*, where the $c_\omega = 5$ and the $c_\eta, c_\varepsilon = 1$, the effect of increasing the non-seasonal correlation is quite small as shown in Figure 5(a). For design *a31*, where the $c_\omega = 10$ and the $c_\eta, c_\varepsilon = 1$, it can be seen in Figure 5(b) that the improvement is greater as the non-seasonal correlation increases, and then even more so for design *a41* where the

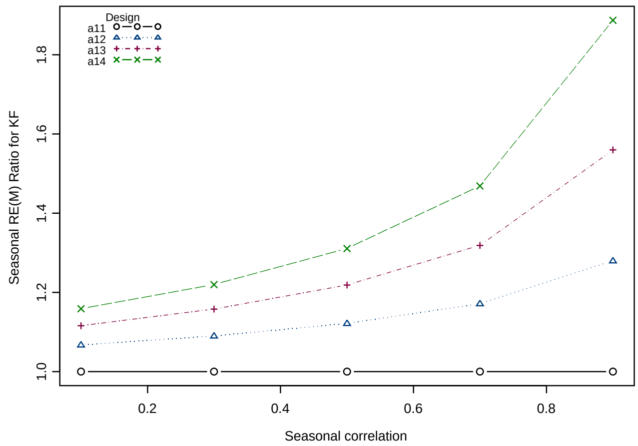
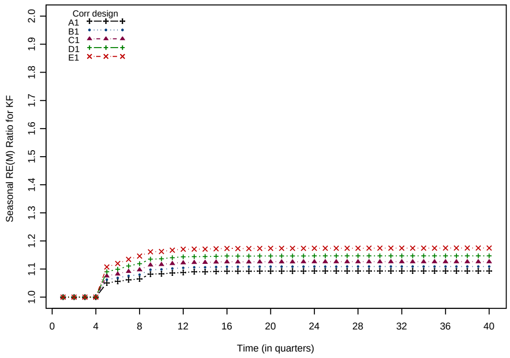


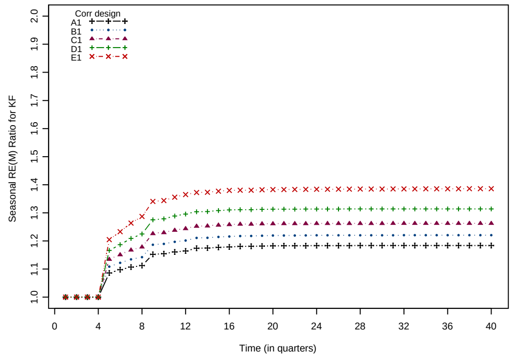
Figure 4: Seasonal correlation versus $RE_{kf}(M)$ at T=40 for Designs *a11*, *a12*, *a13*, *a14* for A1-A5.

$c_\omega = 20$ as shown in Figure 5(c).

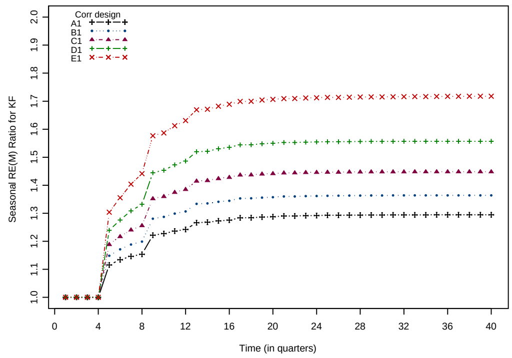
The results for T=40 from each of the series shown in Figure 5, have been plotted against the non-seasonal correlation and given in Figure 6. The plot also shows the results for design *a11*, and overall the results are similar to those shown in Figure 4. Again, it can be seen that the impact of the increasing non-seasonal correlation is dependent upon the design of the structure. There seems to be an interaction between the magnitude of c_ω and the non-seasonal correlation since the gradient of the curve increases as both c_ω and $\rho_\eta, \rho_\varepsilon$ increase.



(a) Design *a21*



(b) Design *a31*



(c) Design *a41*

Figure 5: $RE_{kf}(M)$ for seasonal component for A1 - E1 for Designs *a21*, *a31* and *a41*.

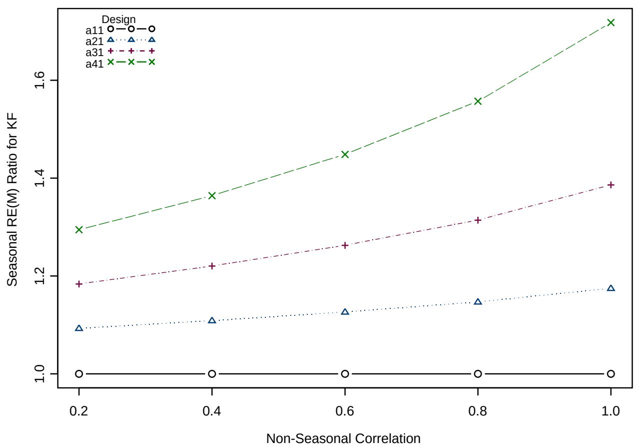


Figure 6: Non-seasonal correlation versus $RE_{kf}(M)$ at T=40 for Designs *a11*, *a21*, *a31*, *a41* for A1-E1.

7 Conclusion

This case study looks at the relative efficiency of the seasonally adjusted aggregate series by using a multivariate structural time series model in state space form applied to the non-stationary sub-series. It extends the work carried out by Geweke (1978) who studied the accuracy of the seasonally adjusted series for stationary time series using spectral densities. Planas and Campolongo (2001) apply some of the results for the multivariate case in Geweke (1978) but also used ARIMA models to describe the sub-series.

The time series considered here are non-stationary and the model structure for a BSM is very different to the structure of an ARIMA model, however, on the whole, the results from this study give similar conclusions and show that gains are possible with the use of the sub-series when a multivariate model is applied.

This study focuses on one particular seasonal local level aggregate series and determines a number of designs for two underlying sub-series. Keeping constraints for the aggregate parameters, the exact multivariate parameters are determined with reference to the ratios of the variances of the sub-series and also the correlations for each of the seasonal and non-seasonal components. Gains are attainable under certain conditions which all rely on the values of the parameters of not only the seasonal component but also the non-seasonal components. The between-series (i.e. within components) and the within-series (i.e. between components) relationships for the two series have been studied and both impact the relative efficiency. The results are best summarised under five main points.

Firstly, when the two sub-series have the same variance parameters for both the seasonal and non-seasonal components (c-ratios are all equal to 1), then there is no difference between the multivariate and the univariate methods. In addition, there is very little difference in the methods when the c-ratios are equal but have a value greater than 1. This is due to the design being close to the homogeneous system. This first point confirms the 'similar' patterns case studied by (Planas and Campolongo, 2001, p21) who found that "the direct and multivariate adjustments tend to coincide and yield nearly equal estimation errors" when using ARIMA-based models.

Secondly, the results for the relative efficiency measure are higher when the c-ratio for the seasonal component is very different to the the c-ratio for the non-seasonal component even if all c-ratios are greater than 1 as in design '*a*'. The magnitude of the relative efficiency measure becomes much greater if the c-ratio for one component (e.g. seasonal) is greater than one but for the other(s) (i.e. non-seasonal) component it is less than one, as in design '*b*'. This confirms the point made by Taylor in his comments to Geweke's paper (Geweke, 1978, p432) where he states that "where the stochastic structure of the non-seasonal and seasonal components are dissimilar, the relative efficiency of the optimal procedure is quite high". This study shows that even when the correlations between the series are low, this statement holds true.

Thirdly, if the c-ratios are held constant and the non-seasonal correlation remains fixed at a low value, then if the seasonal correlation is increased incrementally, the relative efficiency improves, but the extent of the increase depends on the design structure. If the series are described by Model 1, where c-ratios are all equal to 1, then increasing the correlation has no effect on the relative efficiency. Thus, the between components effect overrides the within components effect in regard to the common disturbance term. If the series are quite similar, the increasing seasonal correlation does increase the relative efficiency but this is magnified if the series have dissimilar c-ratios for the seasonal and non-seasonal components. Hence for Model 2, designs with opposing correlation levels perform better.

A similar result holds if the seasonal correlation remains fixed at a low value and the c-ratios are kept fixed, and then the non-seasonal correlation is increased. The relative efficiency for Model 2 increases as the non-seasonal correlation increases away from the value of the seasonal correlation. Thus, better gains are achieved when the seasonal and non-seasonal correlations are at opposite ends of the (positive) correlation scale.

The last two points extend the work done by Geweke (1978) and look more closely at the effect of the different correlation combinations with respect to the dissimilarity of seasonal and non-seasonal stochastic structures. The results plotted over time, show that there is definitely a dependency or interaction between these measures when determining the relative efficiency of the seasonally adjusted series for the multivariate and univariate methods.

Lastly, this study also looks at the relative efficiency over time which is not discussed by either of the previously mentioned authors. For the first 4 time points in this study, the multivariate method and univariate method yield exactly the same MSE's for the filtered estimates. As time progresses, the relative efficiency increases above 1 for each simulation carried out in this study. There are different rates of convergence but on the whole, each plot shows a time series which reaches a steady state. Those with higher c-ratios for the seasonal component tend to be slowest to converge. Increasing the seasonal correlation also has an impact on the rate of convergence. As the correlation increases, the rate becomes slower. However, when the non-seasonal correlation is increased, the rate of convergence for the relative efficiency seems to remain fairly constant.

This paper reports the results of a preliminary study with respect to the multivariate BSM and aggregate time series. Other related work in progress includes the development of an indicator measure, extension of the number of sub-series, a grouping criteria and shortened time spans.

References

Burridge, P., and K. Wallis (1985): "Calculating the variance of seasonally adjusted series," *Journal of the American Statistical Association*, 80(391), 541–552.

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

Feder, M. (2001): "Time series analysis of repeated surveys: the state-space approach," *Statistica Neerlandica*, 55(2), 182–199.

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

Ghysels, E. (1997): "Seasonal adjustment and other data transformations," *Journal of Business and Economic Statistics*, 15(4), 410–418.

Harrison, J., and C. Stevens (1976): "Bayesian Forecasting (with discussion)," *Journal of the Royal Statistical Society*, 38, 205–247.

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

Harvey, A., and J. Durbin (1986): "The effects of seat belt legislation on British road casualties: A case study in structural time series modelling," *Journal of the Royal Statistical Society*, 149(3), 187–227.

Hood, C., and D. F. Findley (2003): "Comparing direct and indirect seasonal adjustments of aggregate series," in *Seasonal Adjustment*, ed. by M. Manna, and R. Peronaci, chap. 1, pp. 9–21. European Central Bank, Frankfurt am Main, Germany.

Jain, R. K. (2001): "A state space model-based method of seasonal adjustment," *Monthly Labour Review*, July, 37–45.

Koopman, S., and J. Durbin (2000): "Fast Filtering and smoothing for multivariate state space models," *Journal of Time Series Analysis*, 21(3), 281–296.

Koopman, S., N. Shephard, and J. Doornik (1999): "Statistical algorithms for models in state space using SsfPack 2.2," *The Econometrics Journal*, 2, 113–166.

Ladiray, D., and G. Mazzi (2003): "Seasonal adjustment of European aggregates: direct versus indirect approach," in *Seasonal Adjustment*, ed. by M. Manna, and R. Peronaci, chap. 3, pp. 37–65. European Central Bank, Frankfurt am Main, Germany.

Marshall, P. (1990): "Analysis of a cross-section of time series using structural time series models," Ph.d, London School of Economics and Political Science, University of London.

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

Otranto, E., and U. Triacca (2002): "Measures to evaluate the discrepancy between direct and indirect model-based seasonal adjustment," *Journal of Official Statistics*, 18(4), 511–530.

Pfeffermann, D., and R. Tiller (2003): "State-space modelling with correlated measurements with application to small area estimation under benchmark constraints," Working Paper M03/11, Southampton Statistical Sciences Research Institute, University of Southampton.

Planas, C., and F. Campolongo (2001): "The seasonal adjustment of contemporaneously aggregated series," Working Paper 8221011/1-Lot3, EUROSTAT, Luxembourg.

Proietti, T. (2000): "Comparing seasonal components for structural time series models," *International Journal of Forecasting*, 16, 247–260.

Sridharan, S., S. Vujic, and S. Koopman (2003): "Intervention time series analysis of crime rates," *Timbergen Institute Discussion Paper*, (TI 2003 -040/4).

Zivot, E., J. Wang, and S. Koopman (2004): "State space modelling in macroeconomics and finance using SsfPack in S+Finmetrics," in *State Space and Unobserved Component Models*, ed. by A. Harvey, S. Koopman, and N. Shephard, chap. 13, pp. 284–335. Cambridge University Press.