2008

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

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

D. G. Steel
University of Wollongong, dsteel@uow.edu.au

Y. X. Lin
University of Wollongong, yanxia@uow.edu.au

Centre for Statistical and Survey Methodology

The University of Wollongong

Working Paper

01-08

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

Carole Birrell, David G Steel and Yan-Xia Lin

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

Seasonal Adjustment of Aggregated 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

April 2008

Abstract

Government statistical agencies are required to seasonally adjust non-stationary time series resulting from aggregation 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 aggregated 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 aggregated 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 aggregated series. The prediction mean squared errors resulting from each method are compared by calculating their relative efficiency. Results indicate that gains are achievable using the multivariate approach according to the relative values of the parameters of the cross-sectional series.

1 Introduction

Seasonally adjusted time series of economic and social data are important products of many official statistical agencies. 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 (or sub-series), is usually required for publication. Given that seasonal adjustment involves estimating and removing the seasonal effects of the series, it is important that the method employed produces accurate estimates of the seasonal components.

When the seasonal component of a series is estimated from the aggregated series and then removed, the process is called direct seasonal adjustment. Alternatively, if each of the sub-series are seasonally adjusted separately, and then summed to obtain the aggregated seasonally adjusted series, the process is called indirect seasonal adjustment. However, both direct and indirect seasonal adjustment use univariate time series methods.

This paper considers a model-based approach for the estimation of seasonal effects for an aggregated series via a multivariate model. The variance of the seasonally adjusted series given by the univariate and multivariate models will be compared using their relative efficiency. The aim is to examine if gains are achievable for the variance of the seasonally adjusted aggregated series by jointly modelling the sub-series. An empirical study will thoroughly investigate the conditions which affect relative efficiency. This will be carried out by fixing the known parameters of an aggregated series, but varying the parameters of the sub-series.

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 are detailed in Section 3 and a measure to compare them 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 conclusions in Section 7.

2 Background

There is extensive debate on whether to use the indirect or the direct approach to seasonal adjustment (see Ghysels, 1997; Hood and Findley, 2003; Ladiray and Mazzi, 2003; and Otranto and Triacca, 2002). Most of the discussion focuses on filter-based methods such as X-11 (Shiskin, Young, and Musgrave, 1967), and its subsequent variants, due to the fact that the seasonally adjusted series resulting from the two methods can, and usually do, differ (Hood and Findley, 2003). Questions arise as to which method produces the more accurate estimates, how to compare the methods, and, if one method is not always better than the other, under what conditions each method should be employed. 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, Geweke (1978) found that the covariance structure between the series is crucial. Planas and Campolongo (2001) used ARIMA models to confirm and extend the results in Geweke (1978). They studied the seasonal adjustment of contemporaneously aggregated series and compared the relative accuracy of the direct method with the indirect and multivariate methods. They confirmed Geweke's result that when the stochastic properties of the two series are even slightly dissimilar, the indirect adjustment is more precise than the direct adjustment (see also Ghysels, 1997). The multivariate adjustment was found to be the most accurate estimation in terms of the final estimation error. However, multivariate estimation was difficult to implement due to its complexity.

Due to the flexibility of the basic structural model and its state space form, multiple time series can be modelled jointly with little difficulty. Extending this idea, a target series can be modelled jointly with one or more related series in order to obtain better estimates of the time series components of the target series. Harvey and Chung (2000) calculated the filtered estimates in a bivariate BSM model and discussed the improvement in the root mean squared error (RMSE) of the slope component over that obtained from just using the univariate model. They found 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.

Other applications of the multivariate BSM can be found in Pfeffermann and Tiller (2003), Durbin and Koopman (2001) and Sridharan, Vujic, and Koopman (2003). The flexibility of the multivariate structural time series model and the results described above, where gains have been achieved with the joint modelling 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 modelled specifically. The series of observations of the aggregated series, $Y_1, \ldots, Y_T$, will be modelled by a univariate additive BSM. If the aggregated series, denoted by Y_t, is a sum of K sub-series, for $t = 1 \ldots T$,

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

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

3.1 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 seasonal component, S_t, and an irregular or disturbance term, $\varepsilon_{U,\,t}$. This BSM may be written, in the notation adopted by Feder (2001), for $t = 1, \ldots, T$ as

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

where the U subscript, in the serially independent $\varepsilon_{U,\,t}$ denotes the univariate model.

The trend 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:

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

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 expected value of the sum of the seasonal effects to be zero over any s time periods.

The focus for this paper 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 aggregated series. 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.

3.2 Multivariate BSM

If an univariate time series is disaggregated such that the sum of the K sub-series is the aggregated (or total) series (1), then a multivariate BSM can be applied to the sub-series, $Y_{1t}, \ldots, Y_{Kt}$. The series may be linked via correlations of the disturbances driving each component. By modelling 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 'seemingly unrelated time series equations' (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. 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.

$$\begin{aligned} Y_{kt} &= L_{kt} + S_{kt} + \varepsilon_t + \varepsilon^*_{kt} & (5)\\ L_{k,t+1} &= L_{kt} + \eta_t + \eta^*_{kt} & (6)\\ S_{k,t+1} &= -\sum_{j=1}^{s-1} S_{k,t+1-j} + \omega_t + \omega^*_{kt} & (7) \end{aligned}$$

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):

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

where

x stands for $\eta,\ \omega,$ or ε

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

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

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

Model 1
The simplest covariance structure, $\mathbf{D}_{x^*} = \sigma^2_{x^*}\mathbf{I}_K$ has all the unit-specific variances 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 the variances which are specific to the K sub-series. These are allowed to differ as given by the definition for Model 2. The alternative structure for $\mathbf{D}^*_x$, has K different values on the diagonal, and hence would have $(K+1)$ 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 example, if $K = 2$, then the covariance matrix for the level component would be

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

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 sub-series:

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

When using the exact parameters, the values of the parameters obtained from the multivariate model for the total series, namely $\sigma^2_{tot,\,\eta}$, $\sigma^2_{tot,\,\omega}$ and $\sigma^2_{tot,\,\varepsilon}$ will be equal to the values of the parameters for the univariate model $\sigma^2_{U,\,\eta}$, $\sigma^2_{U,\,\omega}$ and $\sigma^2_{U,\,\varepsilon}$ respectively. When the parameters are estimated, this property will not necessarily hold.

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 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 (11) and a transition (or state) equation (12). The measurement equation describes the observation at time t as a linear combination of the unobserved components included in the state vector, α_t. The transition equation describes the development of the state vector from one time point to the next. The univariate state space model for the aggregate series may be written (Durbin and Koopman, 2001, section 3.1) as:

$$Y_t = \mathbf{Z}\alpha_t + \varepsilon_{U,\,t} \tag{11}$$

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

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

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

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

$$\mathrm{Var}\,(\mathbf{G}\gamma_t) = \mathbf{GQG}' = \begin{pmatrix} \sigma^2_{U,\,\eta} & 0 & 0 & 0 \\ 0 & \sigma^2_{U,\,\omega} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \mathrm{H} = \sigma^2_{U,\,\varepsilon}. \tag{13}$$

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 (11) and (12)), with the measurement errors separate to the state vector. This conventional format requires uncorrelated measurement errors, that is, the covariance matrix, $\mathbf{\Sigma}_\varepsilon$, 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.

$$\mathbf{Y}_{(m),t} = (\mathbf{Z}_{(m)} \otimes \mathbf{I}_K)\alpha_{(m),t} \tag{14}$$

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

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

$$\begin{aligned}
\mathbf{Y}_{(m),t} &= [Y_{1t}, \ldots, Y_{Kt}]' \\
\alpha_{(m),t} &= [L_{1t}, \ldots, L_{Kt},\ S_{1t}, \ldots, S_{Kt},\ S_{1,t-1}, \ldots, S_{K,t-1}, \\
&\quad S_{1,t-2}, \ldots, S_{K,t-2},\ (\varepsilon_t + \varepsilon^*_{1t}), \ldots, (\varepsilon_t + \varepsilon^*_{Kt})\ ]' \\
\gamma_{(m),t} &= [\ (\eta_t + \eta^*_{1t}), \ldots, (\eta_t + \eta^*_{Kt}),\ (\omega_t + \omega^*_{1t}), \ldots, (\omega_t + \omega^*_{Kt}), \\
&\quad (\varepsilon_{t+1} + \varepsilon^*_{1,t+1}), \ldots, (\varepsilon_{t+1} + \varepsilon^*_{K,t+1})\ ]'
\end{aligned} \tag{16}$$

The system matrices are given by

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

$$\mathbf{T}_{(m)} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \mathbf{G}_{(m)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{17}$$

$$Var\left((\mathbf{G}_{(m)} \otimes \mathbf{I}_K)\gamma_{(m),t}\right) = \begin{pmatrix} \mathbf{\Sigma}_\eta & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{\Sigma}_\omega & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{\Sigma}_\varepsilon \end{pmatrix} \tag{18}$$

where $\mathbf{0}_K$ is a $K \times K$ matrix of zeroes. 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 aggregated series is the series of interest here rather than the individual sub-series. In order to straightforwardly estimate the components of the aggregated series within the multivariate framework and with standard software, a transformation on the multivariate state space model is required. The transformation allows the aggregated series to be included as one of the multivariate series. This means that the estimates of the components of the aggregated series and their mean squared errors are available from the output directly.

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

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 & 1 \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix} = \left(\begin{array}{c|c} 1\ 1\ \ \ldots & 1 \\ \hline & 0 \\ \mathbf{I}_{(K-1)} & \vdots \\ & 0 \end{array}\right) \tag{19}$$

Applying $\mathbf{A}$ 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.

$$\mathbf{A}\left(Y_{1t}, Y_{2t}, \ldots, Y_{Kt}\right)' = \left(Y_{tot,t}, Y_{1t}, \ldots, Y_{K-1,t}\right)'. \tag{20}$$

Applying the transformation to the state space model in (14) and (15), gives

$$\begin{aligned}
\mathbf{Y}_{(M),t} = \mathbf{A}\mathbf{Y}_{(m),t} &= \mathbf{A}(\mathbf{Z}_{(m)} \otimes \mathbf{I}_K)\alpha_{(m),t} \\
&= (\mathbf{Z}_{(m)} \otimes \mathbf{I}_K)(\mathbf{I}_p \otimes \mathbf{A})\alpha_{(m),t} \\
&= (\mathbf{Z}_{(m)} \otimes \mathbf{I}_K)\alpha_{(M),t}
\end{aligned} \tag{21}$$

$$\begin{aligned}
\alpha_{(M),t+1} &= (\mathbf{I}_p \otimes \mathbf{A})\alpha_{(m),t+1} \\
&= (\mathbf{T}_{(m)} \otimes \mathbf{I}_K)(\mathbf{I}_p \otimes \mathbf{A})\alpha_{(m),t} + (\mathbf{I}_p \otimes \mathbf{A})(\mathbf{G}_{(m)} \otimes \mathbf{I}_K)\gamma_{(m),t} \\
&= (\mathbf{T}_{(m)} \otimes \mathbf{I}_K)\alpha_{(M),t} + (\mathbf{I}_p\mathbf{G}_{(m)} \otimes \mathbf{A}\mathbf{I}_K)\gamma_{(m),t} \\
&= (\mathbf{T}_{(m)} \otimes \mathbf{I}_K)\alpha_{(M),t} + (\mathbf{G}_{(m)} \otimes \mathbf{I}_K)(\mathbf{I}_u \otimes \mathbf{A})\gamma_{(m),t} \\
&= (\mathbf{T}_{(m)} \otimes \mathbf{I}_K)\alpha_{(M),t} + (\mathbf{G}_{(m)} \otimes \mathbf{I}_K)\gamma_{(M),t}.
\end{aligned} \tag{22}$$

The matrices $\mathbf{Z}_{(m)}$, $\mathbf{T}_{(m)}$ and $\mathbf{G}_{(m)}$ from (17) remain unchanged. However $\alpha_{(m),t}$, and $\gamma_{(m),t}$ are renamed with the (M) subscript to allow for the transformed elements, see (25) - (26) below. The transformed model has the state space form:

$$\mathbf{Y}_{(M),t} = \mathbf{Z}_{(M)}\alpha_{(M),t} \tag{23}$$
$$\alpha_{(M),t+1} = \mathbf{T}_{(M)}\alpha_{(M),t} + \mathbf{G}_{(M)}\gamma_{(M),t} \tag{24}$$

where

$$\begin{aligned}
\mathbf{Z}_{(M)} &= \mathbf{Z}_{(m)} \otimes \mathbf{I}_K, \qquad \mathbf{T}_{(M)} = \mathbf{T}_{(m)} \otimes \mathbf{I}_K, \qquad \mathbf{G}_{(M)} = \mathbf{G}_{(m)} \otimes \mathbf{I}_K \\
\alpha_{(M),t} &= [L_{tot,t}, L_{1t}, \ldots, L_{K-1,t},\ S_{tot,t}, S_{1t}, \ldots, S_{K-1,t}, \\
&\quad S_{tot,t-1}, S_{1,t-1}, \ldots, S_{K-1,t-1},\ S_{tot,t-2}, S_{1,t-2}, \ldots, S_{K-1,t-2}, \\
&\quad \varepsilon_{tot,t}, (\varepsilon_t + \varepsilon^*_{1t}), \ldots, (\varepsilon_t + \varepsilon^*_{K-1,t})\ ]'
\end{aligned} \tag{25}$$

$$\begin{aligned}
\gamma_{(M),t} = &\left[\ \eta_{tot,t}, (\eta_t + \eta^*_{1t}), \ldots, (\eta_t + \eta^*_{K-1,t}), \right. \\
&\ \omega_{tot,t}, (\omega_t + \omega^*_{1t}), \ldots, (\omega_t + \omega^*_{K-1,t}), \\
&\ \left. \varepsilon_{tot,t+1}, (\varepsilon_{t+1} + \varepsilon^*_{1,t+1}), \ldots, (\varepsilon_{t+1} + \varepsilon^*_{K-1,t+1})\ \right]'
\end{aligned} \tag{26}$$

where

$$\eta_{tot,t} = K\eta_t + \sum_{k=1}^{K}\eta^*_{kt} \qquad \omega_{tot,t} = K\omega_t + \sum_{k=1}^{K}\omega^*_{kt} \qquad \varepsilon_{tot,t} = K\varepsilon_t + \sum_{k=1}^{K}\varepsilon^*_{kt} \tag{27}$$
$$\gamma_{(M),t} \sim N(0, \mathbf{Q}_{(M)})$$

$$\mathrm{Var}\left(\mathbf{G}_{(M)}\gamma_{(M),t}\right) = \mathbf{G}_{(M)}\mathbf{Q}_{(M)}\mathbf{G}'_{(M)} = \begin{pmatrix} \mathbf{\Sigma}_{(M),\eta} & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{\Sigma}_{(M),\omega} & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{\Sigma}_{(M),\varepsilon} \end{pmatrix} \tag{28}$$

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

$$\mathbf{\Sigma}_{(M),\eta} = \mathbf{A}\mathbf{\Sigma}_{\eta}\mathbf{A}' = \begin{pmatrix} \sigma^2_{tot,\eta} & 2\sigma^2_{\eta} + \sigma^2_{1\eta^*} \\ 2\sigma^2_{\eta} + \sigma^2_{1\eta^*} & \sigma^2_{\eta} + \sigma^2_{1\eta^*} \end{pmatrix} \tag{29}$$

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, α_{t+1}, taking into account observations up to time t, via a forward recursion. Denote the information provided by $Y_1, Y_2, \ldots, Y_t$, as $\mathbb{Y}_t$ 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, $Y_1, Y_2, \ldots, Y_T$.

Let the vector, $\mathbf{a}_{t+1|t}$, denote the conditional mean of the state vector, α_{t+1}, based on information available up to time t. Also, let the matrix, $\mathbf{P}_{t+1|t}$, denote the conditional variance for the estimation error of α_{t+1}, based on information available up to time t. $\mathbf{P}_{t+1|t}$ can also be referred to as the mean squared error (MSE) of the estimator $\mathbf{a}_{t+1|t}$ (Harvey, 1989). Therefore, the notation is given as:

$$\begin{aligned} \mathbf{a}_{t+1|t} &= \mathrm{E}(\alpha_{t+1}|\mathbb{Y}_t), \\ \mathbf{P}_{t+1|t} &= \mathrm{Var}(\alpha_{t+1}|\mathbb{Y}_t) \\ &= \mathrm{E}\left[(\alpha_{t+1} - \mathbf{a}_{t+1|t})(\alpha_{t+1} - \mathbf{a}_{t+1|t})^{'}|\mathbb{Y}_t\right]. \end{aligned} \tag{30}$$

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 (11) and (12), with corresponding system matrices (13), these are given by:

$$\begin{aligned} \mathbf{a}_{t+1|t} &= \mathbf{T}\mathbf{a}_{t|t-1} + \mathbf{K}_t\nu_t, \\ \mathbf{P}_{t+1|t} &= \mathbf{T}\mathbf{P}_{t|t-1}\mathbf{L}_t^{'} + \mathbf{G}\mathbf{Q}\mathbf{G}^{'}, \end{aligned} \tag{31}$$

where

$$\begin{aligned} \nu_t &= Y_t - \mathbf{Z}\mathbf{a}_{t|t-1} = \mathbf{Z}\alpha_t + \varepsilon_{U,\,t} - \mathbf{Z}\mathbf{a}_{t|t-1}, \\ F_t &= \mathrm{Var}(\nu_t) = \mathbf{Z}\mathbf{P}_{t|t-1}\mathbf{Z}^{'} + \mathrm{H}, \qquad \mathrm{H} = \mathrm{Var}(\varepsilon_{U,\,t}), \\ \mathbf{K}_t &= \mathbf{T}\mathbf{P}_{t|t-1}\mathbf{Z}^{'}F_t^{-1}, \\ \mathbf{L}_t &= \mathbf{T} - \mathbf{K}_t\mathbf{Z}. \end{aligned} \tag{32}$$

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

$$\begin{aligned} \mathbf{a}_{t|t} &= \mathrm{E}(\alpha_t|\mathbb{Y}_t) = \mathbf{a}_{t|t-1} + \mathbf{M}_tF_t^{-1}\nu_t, \\ \mathbf{P}_{t|t} &= \mathrm{Var}(\alpha_t|\mathbb{Y}_t) = \mathbf{P}_{t|t-1} - \mathbf{M}_tF_t^{-1}\mathbf{M}_t^{'}, \end{aligned} \tag{33}$$

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

$$\begin{aligned} \mathbf{a}_{t+1|t} &= \mathbf{T}\mathbf{a}_{t|t} \\ \mathbf{P}_{t+1|t} &= \mathbf{T}\mathbf{P}_{t|t}\mathbf{T}' + \mathbf{G}\mathbf{Q}\mathbf{G}' \end{aligned} \tag{34}$$

The Kalman filter can be applied using the `S+FinMetrics` software, in particular 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

$$\mathbf{P}_1 = \kappa\mathbf{P}_\infty + \mathbf{P}_* \tag{35}$$

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 (31) to (34) need to be amended for the state space form given by (23) and (24), 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 (25) - (28). Only one equation listed in (32) requires modification, (apart from subscripts). That is the equation for $\mathbf{F}_t$, which becomes

$$\mathbf{F}_{(M),\,t} = \mathbf{Z}_{(M)}\mathbf{P}_{(M),\,t}\mathbf{Z}'_{(M)} \tag{36}$$

with H now a zero matrix, denoted by $\mathbf{H}$. The $\mathbf{Q}_{(M)} = \mathrm{Var}(\gamma_{(M),\,t})$ matrix now includes the variance matrix for the measurement error terms, $\mathbf{\Sigma}_{(M),\varepsilon}$, as shown in (28).

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 $5K \times 5K$) is no longer an identity matrix. The lower right $K \times K$ block diagonal is replaced by a $K \times K$ zero matrix. All other elements remain the same. 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 `S+FinMetrics`.

4 Comparison of Univariate and Multivariate Methods

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

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,

$$Y^a_{t|T} = Y_t - \hat{S}_{t|T} \qquad t = 1, \ldots, T, \tag{37}$$

then, 'the root mean square error (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 the Kalman filter formulation 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 $\text{MSE}(\hat{S}_{t|t})$. 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 $\mathbf{P}_{t|t}$ (33), for the element pertaining to the seasonal component.

In this study, the value of $\text{MSE}(\hat{S}_{tot,t|t})$ using the transformed multivariate model, given the multivariate parameters, will be denoted by $\text{MSE}(\hat{S}^M_{t|t})$. For the univariate method, given the total (or univariate) parameters, $\text{MSE}(\hat{S}_{tot,t|t})$ will be denoted by $\text{MSE}(\hat{S}^U_{t|t})$. To compare the two values, the relative efficiency of the MSE obtained by the univariate method to that of the multivariate method will be calculated. This ratio is defined by:

$$RE_t(M) = \frac{\text{MSE}(\hat{S}^U_{t|t})}{\text{MSE}(\hat{S}^M_{t|t})}, \qquad t = 1 \ldots T \tag{38}$$

and can be considered as a preliminary estimate of the equivalent measure which uses the MSE of the smoothed seasonal component. The $RE_t(M)$ measure 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, various authors agreed that when the series have similar patterns, direct adjustment is favoured and when the series have dissimilar patterns, indirect adjustment is favoured. In model-based seasonal adjustment, particular attention needs to be given to the relationship of parameters between the sub-series and between components, as shown in Geweke (1978) and Planas and Campolongo (2001).

5.1 Setting the Parameters

The parameters in the multivariate state space model are the variances and covariances of the disturbance terms, 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 parameter settings will be varied. The within components need to vary from being the same (as given by Model 1), to being very different (as given by Model 2). Also, the structure of the covariance matrices for the trend component and the seasonal component need to be considered in relation to one another.

The variance of the level component for series k is given by

$$\begin{aligned}\mathrm{Var}(L_{k,t+1} - L_{kt}) &= \mathrm{Var}(\eta_t + \eta^*_{kt})\\ &= \sigma^2_\eta + \sigma^2_{k\eta^*}.\end{aligned}$$

A measure of between-series similarity, c, of the stochastic properties of the series is defined here to help quantify the comments above. If c is defined to be the ratio of the variances between sub-series 1 and 2 within each component, then the c-ratio for the level component is defined as

$$c_\eta = \frac{\mathrm{Var}(L_{1,t+1} - L_{1t})}{\mathrm{Var}(L_{2,t+1} - L_{2t})} = \frac{\sigma^2_\eta + \sigma^2_{1\eta^*}}{\sigma^2_\eta + \sigma^2_{2\eta^*}} \tag{39}$$

Hence, analogously for the other components, the c-ratios for the seasonal and error components are given respectively by:

$$c_\omega = \frac{\sigma^2_\omega + \sigma^2_{1\omega^*}}{\sigma^2_\omega + \sigma^2_{2\omega^*}}, \quad c_\varepsilon = \frac{\sigma^2_\varepsilon + \sigma^2_{1\varepsilon^*}}{\sigma^2_\varepsilon + \sigma^2_{2\varepsilon^*}}. \tag{40}$$

If $c_\eta = c_\omega = c_\varepsilon = 1$ then Model 2 reverts to Model 1, the compound symmetry case, in which all diagonal elements have the same value and the same properties between series apply for each component. For the seasonal component, it does not mean that the set of seasonal factors is the same for sub-series 1 and sub-series 2, but the degree of stability of the seasonal component is the same.

In this study, the c-ratios vary in the set {1, 5, 10, 20} and their reciprocals {1, 0.2, 0.1, 0.05}. Furthermore, to set a design where the stochastic structures of the non-seasonal and seasonal components are different, the c-ratios need to differ 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, combinations of the c-ratios for the components are formulated and are labelled in the following table. 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'* has $c_\omega > 1$ but has the reciprocal of these values for c_η and c_ε. These are also shown in Table 1.

In addition to 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}. These values have been chosen to avoid certain combinations of the c-ratios which would result in the homogeneous case. That is, where the covariance

Table 1: Labels for sub-series design *'a'*: $c_\omega \geq 1$, and $c_\eta, c_\varepsilon \geq 1$, and design *'b'*: $c_\omega \geq 1$, and $c_\eta, c_\varepsilon < 1$

		c_η and c_ε						
		Design *'a'*				Design *'b'*		
		1	5	10	20	0.2	0.1	0.05
c_ω	1	*a11*	*a12*	*a13*	*a14*	-	-	-
	5	*a21*	*a22*	*a23*	*a24*	*b22*	*b23*	*b24*
	10	*a31*	*a32*	*a33*	*a34*	*b32*	*b33*	*b34*
	20	*a41*	*a42*	*a43*	*a44*	*b42*	*b43*	*b44*

Table 2: Correlation design combinations for ρ_ω, ρ_η, and ρ_ε

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

matrices are proportional to one another, (Harvey, 1989, Section 8.3). The design table, labelling the correlation combinations is given in Table 2. 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 constraints on the multivariate variance parameters which are explained in the next section.

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

5.2 Application of Constraints

In this study, the total series remains fixed but the properties of the underlying sub-series vary. The variance parameters for the total series are set with

$$\sigma^2_{tot,\eta} = 0.01, \quad \sigma^2_{tot,\omega} = 1, \quad \sigma^2_{tot,\varepsilon} = 1. \tag{41}$$

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

$$\sigma^2_{tot,\omega} = 4\sigma^2_\omega + \sigma^2_{1\omega^*} + \sigma^2_{2\omega^*}, \quad c_\omega = \frac{\sigma^2_\omega + \sigma^2_{1\omega^*}}{\sigma^2_\omega + \sigma^2_{2\omega^*}}, \quad \rho_\omega = \frac{\sigma^2_\omega}{\sqrt{(\sigma^2_\omega + \sigma^2_{1\omega^*})(\sigma^2_\omega + \sigma^2_{2\omega^*})}}. \tag{42}$$

Solved in terms of $\sigma^2_{tot,\omega}$, c_ω and ρ_ω, the seasonal parameters are

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

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 $0 < \rho_\omega \leq \frac{1}{\sqrt{c_\omega}}$, and if $c_\omega < 1$, then $0 < \rho_\omega \leq \sqrt{c_\omega}$. Similar constraints apply to the level and error components.

Given the nine 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), (6) 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 over time.

The series are summed contemporaneously, $Y_{tot,t} = Y_{1t} + Y_{2t}$ to obtain the simulated aggregated series. Since exact parameters are applied to the model, no data are actually required to obtain the MSE of the seasonal component. For ease of computation however, one realisation of the data has been used with the exact parameters in the software `S+FinMetrics`. It is also possible to calculate the value of $RE_t(M)$ by substituting the parameter settings directly into its algebraic expression for a given value of t. 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 considered in a forthcoming study.

6 Results: Effect of design of the sub-series

The relative efficiency, $RE_t(M)$, is determined for each c-ratio combination specified in Table 1 using the exact parameters. To obtain an overview of these results, the same correlation combination is chosen for each design, with $\rho_\omega = 0.1$ and $\rho_\eta = \rho_\varepsilon = 0.2$. This combination is labelled *A1* in Table 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. These gains vary in magnitude and over time. For those designs which achieve gains, the gains climb in the next few time points to reach a steady value. The time until approximate convergence depends on the design. For example, design *a41* has the largest relative efficiency, $RE_{40}(M) = 1.29$, 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 $RE_t(M)$ 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 very different ($c_\eta \geq 10, c_\varepsilon \geq 10$), a gain is still achievable (although not as large as when the variances differ) for the total seasonal component.

To explore the differences among the 'a' designs in more detail, the numerical results for $T = 40$ for each design are extracted. These results, which are equivalent to $RE_{40}(M)$, are found in Table 3. The lowest results for the relative efficiency belong to the designs

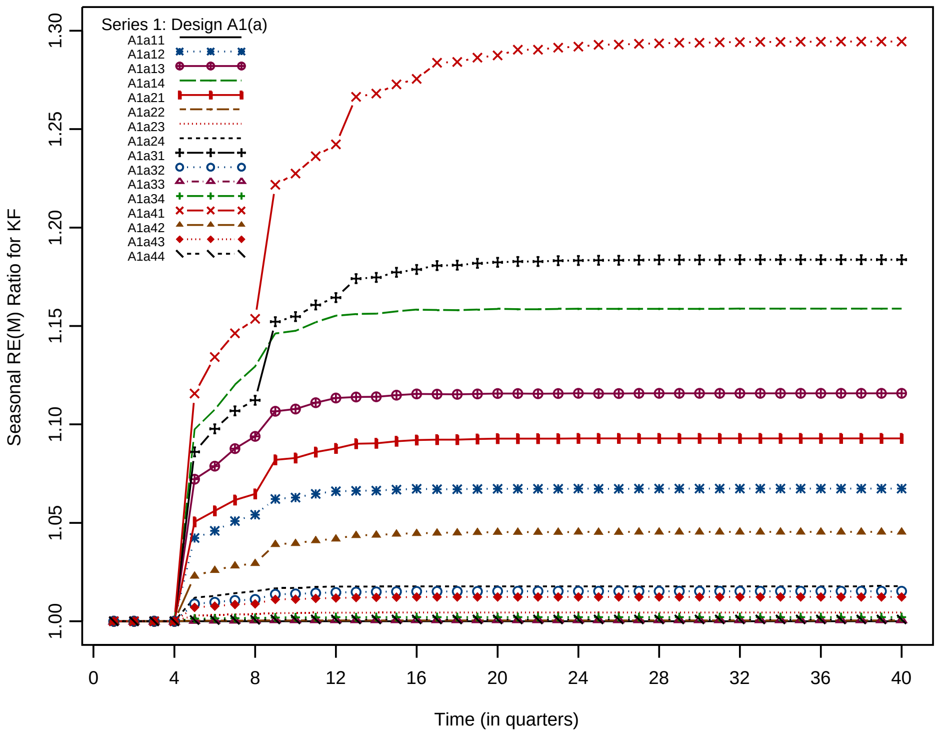


Figure 1: Results of $RE_t(M)$ for sub-series design 'a' with correlation settings *A1*.

Table 3: Results of $RE_{40}(M)$ for sub-series design 'a' and 'b' with correlation settings *A1*.

		c_η and c_ε						
		1	5	10	20	0.2	0.1	0.05
c_ω	1	*a11* 1.000	*a12* 1.067	*a13* 1.116	*a14* 1.159	-	-	-
	5	*a21* 1.093	*a22* 1.001	*a23* 1.005	*a24* 1.018	*b22* 1.373	*b23* 1.506	*b24* 1.616
	10	*a31* 1.184	*a32* 1.015	*a33* 1.001	*a34* 1.002	*b32* 1.606	*b33* 1.811	*b34* 1.982
	20	*a41* 1.295	*a42* 1.045	*a43* 1.012	*a44* 1.001	*b42* 1.901	*b43* 2.213	*b44* 2.482

which have $c_\omega = c_\eta = c_\varepsilon$, namely, *a11*, *a22*, *a33*, and *a44*. Note that for *a11*, where $c_\omega = c_\eta = c_\varepsilon = 1$, 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 structure of the covariance matrices become closer to a homogeneous state.

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 $RE_t(M)$. However, the magnitude is much greater than for the 'a' designs, with nine designs giving an $RE_{40}(M)$ of over 1.25. The largest gain is achieved by design *b44* ($c_\omega = 20,\ c_\eta, c_\varepsilon = 0.05$), with $RE_{40}(M) = 2.48$. 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 3.

When the correlation settings are varied, as given in Table 2, the $RE_t(M)$ is also

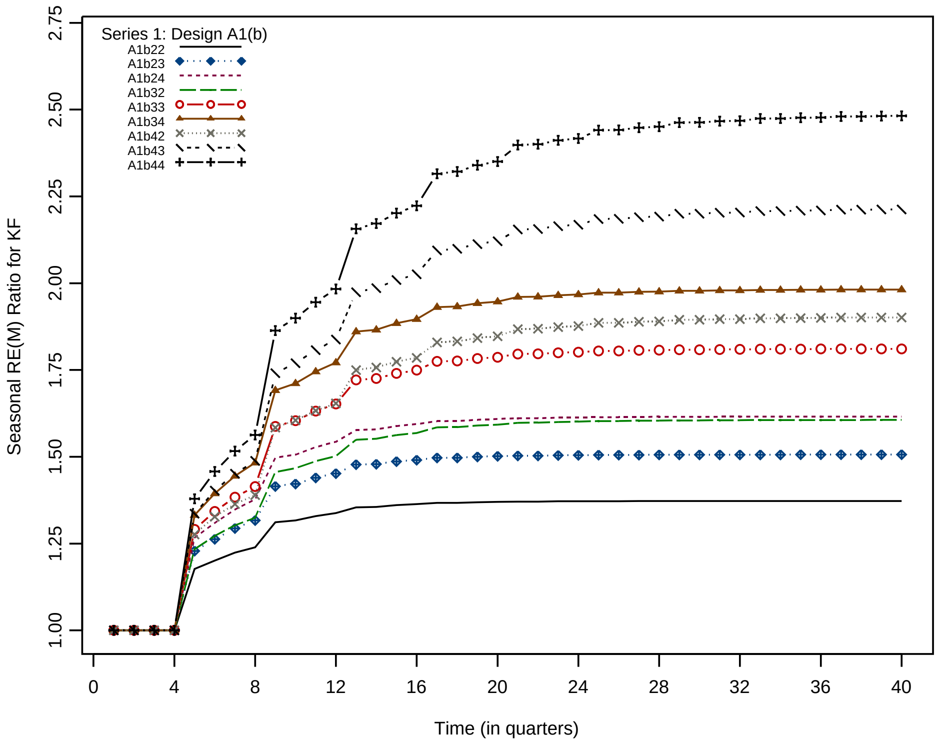
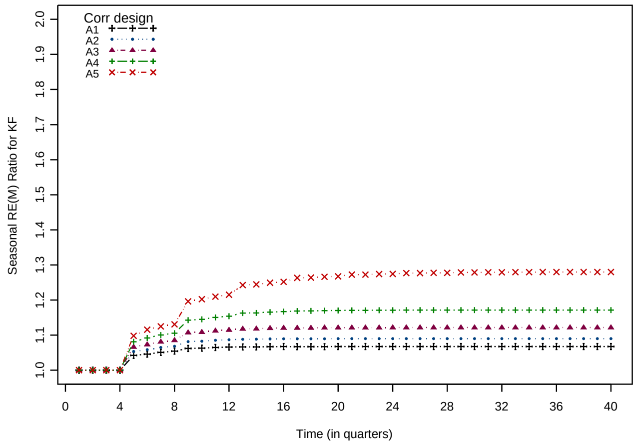


Figure 2: Results of $RE_t(M)$ for sub-series design 'b' with correlation settings *A1*.

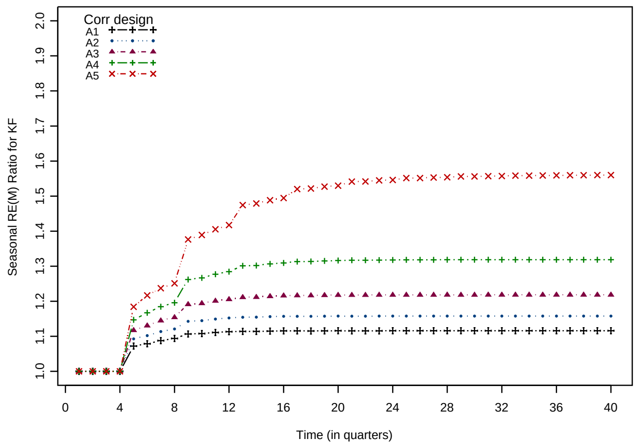
affected. The correlation combination in all 'a' and 'b' designs discussed so far, is identical, with $\rho_\omega = 0.1$ and $\rho_\eta = \rho_\varepsilon = 0.2$ (labelled *A1*). The results show that, even when the correlations between sub-series are small, large gains are attainable, with the size 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 $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ as defined in Table 2. Figure 3 shows the time series plots of $RE_t(M)$ for $t = 1, \ldots, 40$ for the three designs *a12*, *a13* and *a14*, with each plot having the same vertical scale. The seasonal correlation affects $RE_t(M)$ for the seasonal component, as would be expected. The plots show that, as the seasonal correlation increases, the relative efficiency also increases, but the increase is dependent upon the design structure. As the non-seasonal c-ratio increases through designs, from 5 (*a12*) to 10 (*a13*) and then to 20 (*a14*), the effect of the seasonal correlation coefficient intensifies.

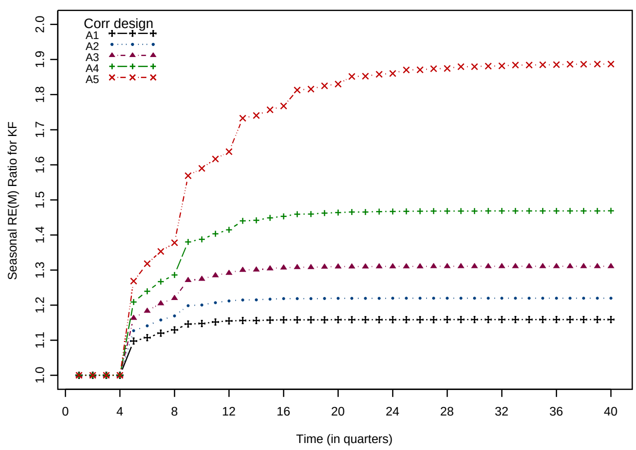
By taking the results for the last time point (i.e. $T = 40$) from each time series within each plot of Figure 3, the positive relationship between the correlation of the seasonal effects and the $RE_t(M)$ value is shown more clearly. Figure 4 shows these results, as well as those for design *a11*. Firstly, for the compound symmetry design (*a11*), the result for the relative efficiency remains constant at one 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 non-seasonal c-ratios c_η, c_ε increase from 5 to 10 to 20.



(a) Design *a12*



(b) Design *a13*



(c) Design *a14*

Figure 3: $RE_t(M)$ for sub-series designs *a12*, *a13* and *a14* with *A1* - *A5*.

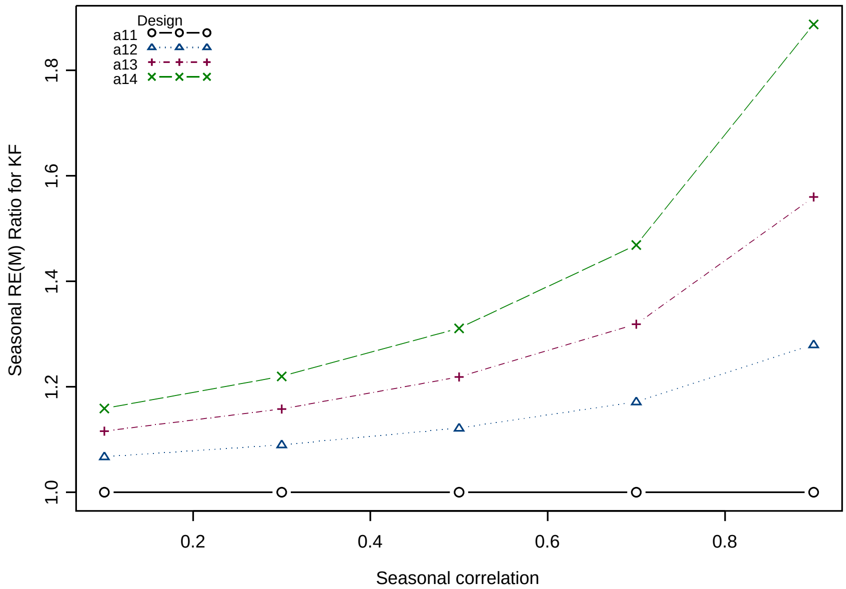
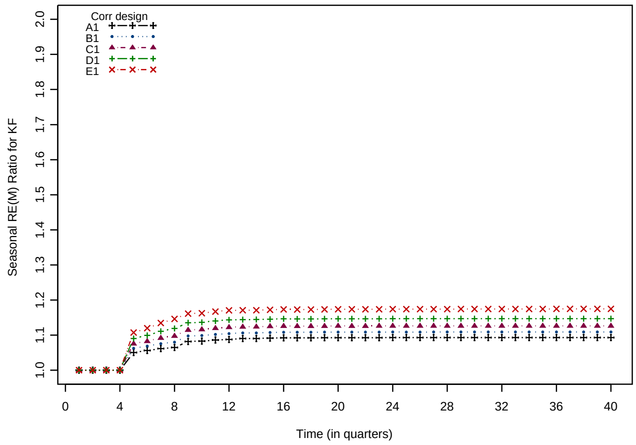


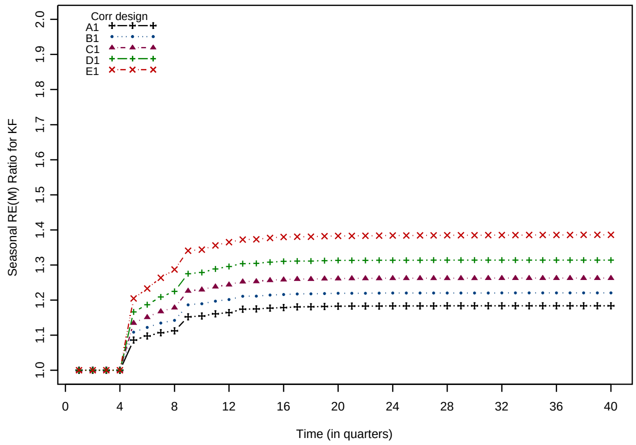
Figure 4: Seasonal correlation versus $RE_{40}(M)$ for designs *a11*, *a12*, *a13*, *a14* with *A1* - *A5*.

The effect of increasing the non-seasonal correlation whilst keeping the seasonal correlation constant is now investigated. The designs *a21, a31, a41* are analysed for correlation combinations *A1* - *E1*. This means that the seasonal correlation is kept at $\rho_\omega = 0.1$, and the non-seasonal correlations ρ_η, ρ_ε are one of $\{0.2,\ 0.4,\ 0.6,\ 0.8,\ 1.0\}$. Figure 5 shows the time series plots of $RE_t(M)$ for the three designs *a21, a31, a41* with the same scale on the vertical axis. Note that here it is the non-seasonal correlation which is increasing, and the relative efficiency of the seasonal component is still affected. For design *a21*, for which $c_\omega = 5$ and $c_\eta = c_\varepsilon = 1$, the effect of increasing the non-seasonal correlation is quite small, as shown in Figure 5(a). For design *a31*, for which $c_\omega = 10$ and $c_\eta = c_\varepsilon = 1$, it can be seen from Figure 5(b) that the improvement is greater as the non-seasonal correlation increases. This becomes even more evident for design *a41*, for which $c_\omega = 20$ as shown in Figure 5(c).

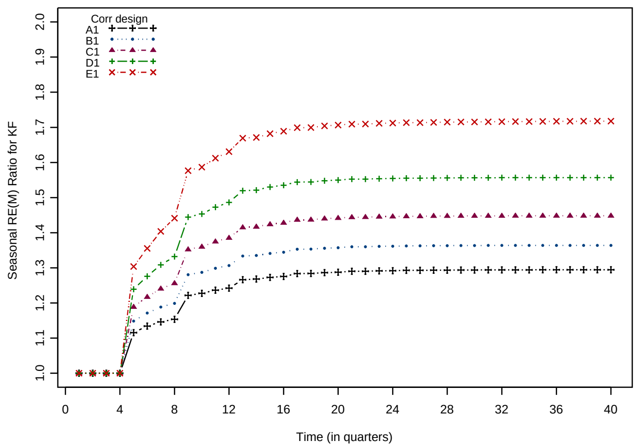
The results for $T = 40$ from each of the series shown in Figure 5 have been plotted against the non-seasonal correlation and given in Figure 6. The plot also shows the results for design *a11*. Overall, the results are similar to those shown in Figure 4. Again, it can be seen that the impact of the increasing non-seasonal correlation is dependent upon the parameter settings. 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: $RE_t(M)$ for sub-series designs *a21*, *a31* and *a41* with *A1* - *E1*.

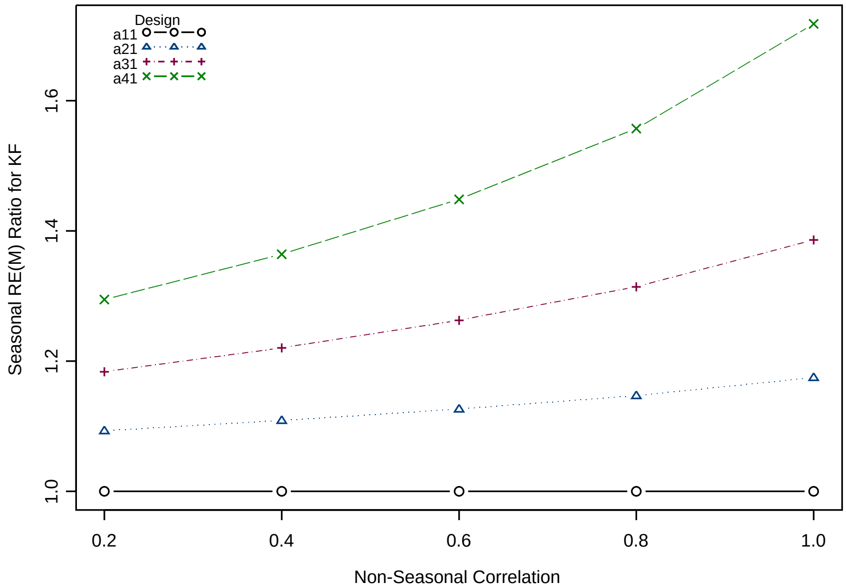


Figure 6: Non-seasonal correlation versus $RE_{40}(M)$ for designs *a11*, *a21*, *a31*, *a41* with *A1-E1*.

7 Conclusion

In this empirical study, the relative efficiency of the seasonally adjusted aggregated series has been investigated by using a multivariate structural time series model applied to the non-stationary sub-series. It is therefore an extension to the work of Geweke (1978), who used spectral densities in studying the accuracy of the seasonally adjusted series in reference to stationary time series. More recently, Planas and Campolongo (2001) applied some of the results for the multivariate case in Geweke (1978) but also used ARIMA models to describe the sub-series.

Although non-stationary time series have been considered here, and a different model structure has been used, the results from this study reflect quite similar conclusions to those made by Geweke (1978) and Planas and Campolongo (2001). It has been shown that gains in the accuracy of the seasonally adjusted series are possible by joint modelling of the sub-series.

This study focuses on one particular local level seasonal aggregated series and utilises a selection of designs for two sub-series. Keeping constraints for the aggregated 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 conditions which rely on the values of the parameters of the seasonal component and 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 affect 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 one), 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 high, meaning that the series have very different variance parameters within components only if the c-ratios are equally high for both the seasonal and non-seasonal components. 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 relative efficiency is higher when the c-ratio for the seasonal component is very different to the c-ratio for the non-seasonal component, even if all c-ratios are greater than one, as in design 'a'. The magnitude of the relative efficiency becomes much greater if the c-ratio is greater than one for one component (e.g. seasonal) but is less than one for other (i.e. non-seasonal) components, as in design 'b'. This confirms the point made by Taylor in his comments on Geweke's paper (Geweke, 1978, p432): "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 with non-seasonal correlation kept constant and low, when the seasonal correlation is increased incrementally, the relative efficiency improves, but the extent of the increase depends on the design structure. If the sub-series are described by Model 1, where c-ratios are all equal to one, then increasing the correlation has no effect on the relative efficiency. If the sub-series are quite similar, increasing seasonal correlation increases the relative efficiency, and this is magnified if the series have dissimilar c-ratios for the seasonal and non-seasonal components. A similar result holds if the seasonal correlation is kept constant and low, the c-ratios are held constant and the non-seasonal correlation is increased. Thus, the relative efficiency increases if the non-seasonal correlation increases away from the value of the seasonal correlation.

The last two points extend the work of Geweke (1978) and look more closely at the effect of different correlation combinations in addition to the effect of seasonal and non-seasonal stochastic structures. Fourthly, the results plotted over time show that the impact of increasing either the seasonal or non-seasonal correlation is greatest for the designs with very different c-ratios.

Lastly, this study also examines the evolution of relative efficiency over time, an aspect not discussed by either of the previously mentioned authors. For the first 4 time points, the multivariate method and univariate method yield exactly the same MSEs for the filtered estimates. This is due to the application of the exact initial Kalman filter. For exact parameters, the theoretical expressions for the MSE of the seasonal component of the total series, for the univariate and multivariate methods, are equal for $t = 1, \ldots, 4$. As time progresses, the relative efficiency increases above one 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, convergence becomes slower. However, when the non-seasonal correlation is increased, the rate of convergence for relative efficiency seems to remain fairly constant.

This paper reports the results of modelling two sub-series of a particular aggregated time series with known parameters. Other aggregated series with different parameters have shown to produce similar results. The method proposed here may be extended to aggregated series with more than two sub-series.

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.

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

Harvey, A., and C.-H. Chung (2000): "Estimating the underlying change in unemployment in the UK," *Journal of the Royal Statistical Society*, 163(3), 303–339.

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, 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. thesis, 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.

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

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.