2010

# Seasonal Adjustment of an Aggregate 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

01-10

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

Carole L. 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 an Aggregate Series using Univariate and Multivariate Basic Structural Models

**Carole L. Birrell**, *Centre for Statistical and Survey Methodology*
*School of Mathematics and Applied Statistics*
*University of Wollongong, 2522 NSW, Australia*
*Email: cbirrell@uow.edu.au*

**David G. Steel** *Centre for Statistical and Survey Methodology*
*School of Mathematics and Applied Statistics*
*University of Wollongong, 2522 NSW, Australia*
*Email: dsteel@uow.edu.au*

**Yan-Xia Lin** *Centre for Statistical and Survey Methodology*
*School of Mathematics and Applied Statistics*
*University of Wollongong, 2522 NSW, Australia*
*Email: yanxia@uow.edu.au*

---

**Abstract**

Time series resulting from aggregation of several sub-series can be seasonally adjusted directly or indirectly. With model-based seasonal adjustment, the sub-series may also be considered as a multivariate system of series and the analysis may be done jointly. This approach has considerable advantage over the indirect method, as it utilises the covariance structure between the sub-series.

This paper compares a model-based univariate and multivariate approach to seasonal adjustment. Firstly, the univariate basic structural model (BSM) is applied directly to the aggregate series. Secondly, the multivariate BSM is applied to a transformed system of sub-series. The prediction mean squared errors of the seasonally adjusted aggregate series 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 sub-series.

*AMS Subject Classification:* 62M10, 91B84

**Keywords**: Seasonal adjustment, Basic structural model, Kalman filter, multivariate time series, state space model.

---

## 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 sub-series (or cross-sectional 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 is seasonally adjusted separately, and then summed to obtain the aggregated seasonally adjusted series, the process is called indirect seasonal adjustment. Both direct and indirect seasonal adjustment employ univariate analyses. Although the indirect method utilises all the sub-series, it does not do so jointly and ignores the relationships between the sub-series.

The 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. By using a structural time series model, seasonal adjustment may be performed for the aggregate series using all the information in the sub-series by borrowing strength from the connections between the sub-series. 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 aggregate 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 and varying the parameters of the sub-series.

Section 2 gives a brief background of the seasonal adjustment approaches and also reviews some applications of the multivariate basic structural model (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 empirical study and parameter settings are outlined in Section 5. Results are presented in Sections 6 and 7 with conclusions in Section 8.

## 2 Background

In general, there are two main approaches to seasonal adjustment, namely a filter-based approach and a model-based approach. Filter-based methods estimate time series components, such as the trend and seasonal factors, by application of a set of filters to the original series. This iterative technique is applied by the widely used X-11 package (Shiskin *et al.*, 1967). X-11 evolved to include the use of ARIMA (Autoregressive Integrated Moving Average) models to forecast the series in order to improve the estimation of components at the end of the series, thereby reducing revisions. Developments include X-11-ARIMA and X-11-ARIMA-88 developed by Dagum (1988) and X-12-ARIMA (U.S. Bureau of Census).

Model-based seasonal adjustment requires modelling the observed time series and the unobserved components such as the trend, seasonal and irregular components. There are two main approaches to modelling time series: the ARIMA model-based (AMB) approach (see Burman, 1980; Hillmer and Tiao, 1982) and the structural time series (STS) approach (see Engle, 1978; Harvey and Todd, 1983; Harvey, 1989). The ARIMA model-based approach to seasonal adjustment involves fitting a seasonal ARIMA model to the overall series and then decomposing it into appropriate models for each of the components (Mar-

avall, 1995). This approach is often called 'signal extraction' (Whittle, 1963 and Burman, 1980). It is implemented with the TRAMO-SEATS (Time series Regression with ARIMA noise, Missing values, and Outliers - Signal Extraction in ARIMA Time Series) software (Gomez and Maravall, 1996).

Structural time series models are models where the components of the series are modelled separately. They are appropriate for modelling time series from many disciplines such as economics, sociology, engineering and geography (Harvey, 1989, p xi). The Kalman filter (Kalman, 1960), first developed in the field of engineering, is the algorithm used to estimate the structural components of the model. Programs such as STAMP (Structural Time series Analyser, Modeller and Predictor) and others such as the set of functions collectively called the *SsfPack* (Koopman *et al.*, 1999) in the module `S+FinMetrics` may be used for the analysis. The AMB approach is generally not suitable for multivariate analysis, whereas in the structural time series model known as the basic structural model (BSM) (Harvey, 1989), multivariate analysis is a natural extension to the univariate model.

A comparison of three seasonal adjustment methods with respect to both temporal and sectoral aggregation is discussed in Geweke (1978). Using spectral densities, Geweke calculated the mean squared error (MSE) using the multivariate, indirect and direct methods of seasonal adjustment for stationary series. For model-based seasonal adjustment, Geweke (1978) concluded that the covariance structure between the series is crucial. He found that the joint ARIMA model was advantageous, as summarised by Taylor, when the sub-series are "*very heterogeneous, or where the stochastic structure of the non-seasonal and seasonal components are dissimilar*" (Taylor, 1978, p432)

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 of the trend 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.

The main advantage of applying a BSM is that multivariate series may be modelled as an extension to the univariate model. Other advantages are that the structural components are a direct result of the estimation using the Kalman filter and the estimate of the variance of the seasonally adjusted series is a by-product of estimating the components (Jain, 2001). This paper parallels the work by Geweke (1978) and Planas and Campolongo (2001) but within the

basic structural model framework.

## 3 Basic structural model

A structural time series model allows time series characteristics such as trend, seasonal and error components to be modelled explicitly. The series of observations of the aggregated series, $Y_1, \ldots, Y_T$, can be directly modelled by a univariate additive BSM. If the aggregated series, denoted by $Y_{tot,t}$, is a sum of $K$ sub-series, for $t = 1 \ldots T$,

$$Y_{tot,t} = \sum_{k=1}^{K} Y_{kt} \tag{3.1}$$

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

For the experimental study in this paper, the local level seasonal (LLS) model is chosen. The local level seasonal model has the restriction that the trend does not include a slope component (for more details see Harvey, 1989, Section 2.3).

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

$$\begin{aligned} Y_t &= L_t + S_t + \varepsilon_{U,\,t}, & \varepsilon_{U,\,t} &\sim N(0, \sigma^2_{U,\,\varepsilon}) && (3.2)\\ L_{t+1} &= L_t + \eta_{U,\,t}, & \eta_{U,\,t} &\sim N(0, \sigma^2_{U,\,\eta}) && (3.3)\\ S_{t+1} &= -\sum_{j=1}^{s-1} S_{t+1-j} + \omega_{U,\,t} & \omega_{U,\,t} &\sim N(0, \sigma^2_{U,\,\omega}). && (3.4) \end{aligned}$$

The local level seasonal model is given by (3.2), (3.3) with (3.4) 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, 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 a univariate time series is disaggregated such that the sum of the $K$ sub-series is the aggregated (or total) series (3.1), then a multivariate BSM can be used for the sub-series, $Y_{1t}, \ldots, Y_{Kt}$. The series may be linked by the 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.

$$\begin{aligned} Y_{kt} &= L_{kt} + S_{kt} + \varepsilon_t + \varepsilon^*_{kt} && (3.5)\\ L_{k,t+1} &= L_{kt} + \eta_t + \eta^*_{kt} && (3.6)\\ S_{k,t+1} &= -\sum_{j=1}^{s-1} S_{k,t+1-j} + \omega_t + \omega^*_{kt} && (3.7) \end{aligned}$$

The disturbance terms, $\varepsilon_t,\ \varepsilon^*_{kt},\ \eta_t,\ \eta^*_{kt},\ \omega_t,\ \omega^*_{kt}$ are assumed to be mutually 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):

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

where $x$ stands for $\eta$, $\omega$, or $\varepsilon$ and $\mathbf{x}^*_t$ stands for $(\eta^*_{1t}, \ldots, \eta^*_{Kt})'$, $(\omega^*_{1t}, \ldots, \omega^*_{Kt})'$, or $(\varepsilon^*_{1t}, \ldots, \varepsilon^*_{Kt})'$. $\mathbf{1}_K$ is a $K$ dimensional vector of one's, and $\mathbf{J}_K = \mathbf{1}_K \mathbf{1}_K^{'}$. The matrix $\mathbf{D}_{x^*}$ can be defined as $\mathbf{D}_{x^*} = \mathrm{diag}\left[\sigma^2_{1x^*}, \ldots, \sigma^2_{Kx^*}\right]$. Note that if the diagonal elements of $\mathbf{D}_{x^*}$ are equal, the resulting covariance matrix has a compound symmetry structure.

The $K$ unit-specific variances are the variances that are specific to the $K$ sub-series. The matrix $\mathbf{D}^*_x$, has $K$ different values on the diagonal, and hence each of the three component covariance matrices (namely $\mathbf{\Sigma}_\eta$, $\mathbf{\Sigma}_\omega$, and $\mathbf{\Sigma}_\varepsilon$) would have $(K+1)$ unknown parameters, giving a total of $3(K+1)$ unknown parameters. For example, if $K = 2$, then the covariance matrix for the level component is

$$\mathbf{\Sigma}_\eta = \begin{pmatrix} \sigma^2_\eta + \sigma^2_{1\eta^*} & \sigma^2_\eta \\ \sigma^2_\eta & \sigma^2_\eta + \sigma^2_{2\eta^*} \end{pmatrix} \qquad (3.8)$$

and similarly for $\mathbf{\Sigma}_\omega$, and $\mathbf{\Sigma}_\varepsilon$.

Since the aggregate series given by (3.1) is the sum of the $K$ sub-series, the parameters of the aggregate series can be defined in terms of the parameters of the sub-series:

$$\begin{aligned} \sigma^2_{tot,\eta} &= K^2\sigma^2_\eta + \sum_{k=1}^{K} \sigma^2_{k\eta^*}, \quad \sigma^2_{tot,\omega} = K^2\sigma^2_\omega + \sum_{k=1}^{K} \sigma^2_{k\omega^*}, \\ \sigma^2_{tot,\varepsilon} &= K^2\sigma^2_\varepsilon + \sum_{k=1}^{K} \sigma^2_{k\varepsilon^*}. \qquad (3.9) \end{aligned}$$

The univariate and multivariate BSMs may be written more concisely in state space form (SSF) and then analysis is carried out with the Kalman filter. The Kalman filter provides the optimal estimator of the state vector, $\alpha_{t+1}$, taking into account the observations up to time $t$ using a forward recursion. The

elements of the state vector are the unobserved components such as the level and seasonal components. The mean squared errors (MSEs) of the estimates of the unobserved components are also provided by the Kalman filter. Details of state space form for the univariate and multivariate models are given in the Appendix. For more general information on state space models and the Kalman filter refer to Kalman (1960), Harvey (1989), and Durbin and Koopman (2001).

### 3.3 Application of the Kalman filter

A BSM can be written as a linear Gaussian state space model and analysed by applying the Kalman filter and Kalman smoother to the observations. In a state space model, the state vector, denoted by $\alpha_t$ (see Appendix for details) contains all the series components. 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 $Y_1, Y_2, \ldots, Y_t$, 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, $Y_1, Y_2, \ldots, Y_T$.

In state space form, the state vector estimator which includes the estimates for all the series components may be written as $\mathbf{a}_{t|t}$. Its corresponding error variance matrix is $\mathbf{P}_{t|t}$. They are defined as (Durbin and Koopman, 2001):

$$\begin{aligned} \mathbf{a}_{t|t} &= \mathrm{E}(\alpha_t|\mathbb{Y}_t) \\ \mathbf{P}_{t|t} &= \mathrm{Var}(\alpha_t|\mathbb{Y}_t) \end{aligned} \tag{3.10}$$

The Kalman filter can be applied using the `S+FinMetrics` software, in particular the set of functions collectively called the *SsfPack* (Koopman *et al.*, 1999). Further details of $\alpha_t$ are given in the appendix.

## 4 Comparison of univariate and multivariate methods

A direct approach to seasonal adjustment is based on $Y_t$ using the univariate model (3.2) to (3.4), an indirect approach would be (but not carried out in this paper) based on each individual sub-series $Y_{kt}$, $k = 1, 2, \cdots, K$, and a multivariate approach is based on $Y_{kt}$ using the multivariate model (3.5) to (3.7). The main focus of this paper is to compare the accuracy of the seasonally adjusted aggregate series of the direct approach with that of the multivariate approach. This section explains how a simple transformation of the multivariate model allows explicit analysis of $Y_{tot,t}$ and hence a method of comparison of the accuracy of the seasonally adjusted aggregate series.

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

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

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

that

$$\mathbf{Y}_{(M),\,t} = \mathbf{A}\left(Y_{1t}, Y_{2t}, \ldots, Y_{Kt}\right)' = \left(Y_{tot,t}, Y_{1t}, \ldots, Y_{K-1,t}\right)'. \tag{4.1}$$

In general, if the current-adjusted series is given by $Y^a_{t|t} = Y_t - \hat{S}_{t|t}$, where $\hat{S}_{t|t}$ is the estimate of the seasonal component, an element of $\mathbf{a}_{t|t}$ (3.10), then the appropriate measure of the accuracy of the adjusted data is the error variance of the seasonal component estimate, conditional on the data (see Harvey, 1989; Burridge and Wallis, 1985). This is the error variance given by the Kalman filter as calculated in the matrix $\mathbf{P}_{t|t}$ (3.10), for the element pertaining to the seasonal component. The current-adjusted series can be viewed as the preliminary seasonally adjusted series as it is conditional on observations up to time $t$ ($\mathbb{Y}_t$).

To compare the accuracy of the two approaches, their relative efficiency is calculated by:

$$RE_t(M) = \frac{\mathrm{MSE}(\hat{S}^U_{t|t})}{\mathrm{MSE}(\hat{S}^M_{t|t})}, \qquad t = 1 \ldots T \tag{4.2}$$

where $\mathrm{MSE}(\hat{S}^U_{t|t})$ and $\mathrm{MSE}(\hat{S}^M_{t|t})$ denote the value of $\mathrm{MSE}(\hat{S}_{t|t})$ using the univariate model and transformed multivariate model respectively.

## 5 Design of the study

To examine the question of whether, and under what conditions, the multivariate method is beneficial compared with the univariate method in determining a seasonally adjusted aggregate series, an empirical study using a range of parameter values is employed. A particular aggregate series with known parameters is taken and then disaggregated into two sub-series. The two sub-series are determined by setting the parameters for each component and for each series. These parameters are used in determining the elements of the covariance matrices $\mathbf{\Sigma}_\eta$, $\mathbf{\Sigma}_\omega$ and $\mathbf{\Sigma}_\varepsilon$. The specification of their structure is referred to as the design of the sub-series. By fixing the parameters for the total series, the parameters of the sub-series may be varied. This top-down approach allows a controlled way of measuring the effect of the parameters of the sub-series on the accuracy of the seasonally adjusted total series.

In the literature there is extensive discussion on direct versus indirect adjustment (see Ghysels, 1997; Hood and Findley, 2003; Ladiray and Mazzi, 2003; Otranto and Triacca, 2002). In general, it is 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). Their conclusions have been used to guide the setting of the sub-series parameters.

### 5.1 Setting the parameters

In this case study, a range of parameter values is set with reference to two concepts. Firstly, the relationship of the parameters between-series (i.e. within

components) is studied. For example, the parameter values for the level component for series 1 as compared to the parameter values for the level component for series 2. Secondly, the relationship of the parameters within-series (i.e. between components) is also studied. Thus, the structure of the covariance matrix for the level component and that for the seasonal component are considered relative to one another.

A measure of between-series similarity, $c$, of the stochastic properties of the series is defined here to help quantify the comments above. Let $c_\eta$, $c_\omega$ and $c_\varepsilon$ be the ratios of the variances of sub-series 1 and 2 for the level, seasonal and error components respectively. For example, the variance of the level component for sub-series $k$ is given by:

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

Then the c-ratio for the level component is defined to be:

$$c_\eta = \frac{\mathrm{Var}(L_{1,t+1} - L_{1t})}{\mathrm{Var}(L_{2,t+1} - L_{2t})} = \frac{\sigma^2_\eta + \sigma^2_{1\eta^*}}{\sigma^2_\eta + \sigma^2_{2\eta^*}} \tag{5.1}$$

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

$$c_\omega = \frac{\sigma^2_\omega + \sigma^2_{1\omega^*}}{\sigma^2_\omega + \sigma^2_{2\omega^*}}, \quad c_\varepsilon = \frac{\sigma^2_\varepsilon + \sigma^2_{1\varepsilon^*}}{\sigma^2_\varepsilon + \sigma^2_{2\varepsilon^*}}. \tag{5.2}$$

If $c_\eta = c_\omega = c_\varepsilon = 1$ then for each covariance matrix (namely $\mathbf{\Sigma}_\eta$, $\mathbf{\Sigma}_\omega$, and $\mathbf{\Sigma}_\varepsilon$), the diagonal elements have the same value. This means that the same properties between series apply for each component, which corresponds to a compound symmetry structure for each matrix. 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 Table 1. Table 1 shows design *'a'* where all c-ratios are greater than or equal to one. Note that $c_\eta$ and $c_\varepsilon$ 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, $c_\omega$. Design *'b'*, also shown in Table 1, has $c_\omega > 1$ but has the reciprocal of these values for $c_\eta$ and $c_\varepsilon$.

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, $\rho_\omega$, 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 ($\rho_\eta$ and $\rho_\varepsilon$) 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 matrices are proportional to one another, (Harvey, 1989, Section 8.3). The design table,

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 $c_\eta, c_\varepsilon < 1$

| | | $c_\eta$ and $c_\varepsilon$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Design *'a'* | | | | Design *'b'* | | |
| | | 1 | 5 | 10 | 20 | 0.2 | 0.1 | 0.05 |
| $c_\omega$ | 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 $\rho_\omega$, $\rho_\eta$, and $\rho_\varepsilon$

| | | $\rho_\eta$ and $\rho_\varepsilon$ | | | | |
|---|---|---|---|---|---|---|
| | | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 |
| $\rho_\omega$ | 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 |

labelling the correlation combinations is given in Table 2. For example, the design 'A1*a23*' refers to the case $c_\omega = 5$, $c_\eta, c_\varepsilon = 10$, $\rho_\omega = 0.1$, 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 nine 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

$$\sigma^2_{tot,\eta} = 0.01, \quad \sigma^2_{tot,\omega} = 1, \quad \sigma^2_{tot,\varepsilon} = 1. \tag{5.3}$$

With these given univariate parameters, as well as the constraints given in (3.9), 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:

$$\begin{aligned} \sigma^2_{tot,\omega} &= 4\sigma^2_\omega + \sigma^2_{1\omega^*} + \sigma^2_{2\omega^*}, \ \ c_\omega = \frac{\sigma^2_\omega + \sigma^2_{1\omega^*}}{\sigma^2_\omega + \sigma^2_{2\omega^*}}, \\ \rho_\omega &= \frac{\sigma^2_\omega}{\sqrt{(\sigma^2_\omega + \sigma^2_{1\omega^*})(\sigma^2_\omega + \sigma^2_{2\omega^*})}}. \end{aligned} \tag{5.4}$$

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

$$\begin{aligned} \sigma^2_\omega &= \frac{\rho_\omega\sqrt{c_\omega}\sigma^2_{tot,\omega}}{1+c_\omega+2\rho_\omega\sqrt{c_\omega}}, \quad \sigma^2_{1\omega^*} = \frac{\sigma^2_{tot,\omega}(c_\omega-\rho_\omega\sqrt{c_\omega})}{1+c_\omega+2\rho_\omega\sqrt{c_\omega}}, \\ \sigma^2_{2\omega^*} &= \frac{\sigma^2_{tot,\omega}(1-\rho_\omega\sqrt{c_\omega})}{1+c_\omega+2\rho_\omega\sqrt{c_\omega}}. \end{aligned} \tag{5.5}$$

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 multivariate parameters, the data for $Y_{1t}$ and $Y_{2t}$ are generated from the multivariate model equations for $t = 40+T$ as described in (3.5), (3.6) and (3.7), 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 transformation described in Section 4 is applied to obtain the required series: $Y_{tot,t}$ and $Y_{1t}$.

Using the *SsfPack* functions (Koopman *et al.*, 1999) in `S+FinMetrics`, the univariate state space model (Section A.1) is applied to the aggregate series $Y_{tot,t}$, and the multivariate state space model (Section A.2) is applied to the two series $Y_{tot,t}$ and $Y_{1t}$. This yields MSE($\hat{S}^U_{t|t}$) and MSE($\hat{S}^M_{t|t}$) respectively for $t = 1 \ldots 40$. The relative efficiency, $RE_t(M)$, can then be calculated for each time point.

Known parameters are applied here so that the effect of the design on the relative efficiency ratio is not obscured by the values of any estimated parameters. The effect of estimation of the parameters is considered in a forthcoming study.

## 6 Results: effect of the parameters of the sub-series

The relative efficiency, $RE_t(M)$, is determined for each c-ratio combination specified in Table 1 using the known parameters. To obtain an overview of these results, the same correlation combination (*A1*)is chosen for each design, with $\rho_\omega = 0.1$ and $\rho_\eta = \rho_\varepsilon = 0.2$.

Figure 1 shows the results over $t = 1 \ldots 40$ for the 16 different *'a'* designs given in Table 1. For $t = 1 \ldots 4$, the relative efficiency is exactly one. This result is due to the exact initial Kalman filter, as described in detail in Koopman (1997). 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 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. 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. As there is little or no gain in using the multivariate approach for these four designs, it is recommended to use the univariate approach.

The '*b*' designs use the reciprocal of the values of $c_\eta, c_\varepsilon$ 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 $c_\omega$ is very different from $c_\eta$ and $c_\varepsilon$, for example *b44, b43, b34*, give the highest gains. The numerical results for $T = 40$ for each '*b*' design are given in Table 3.

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.

By taking the results for the last time point ($T = 40$) from each $RE_t(M)$ time series, the positive relationship between the seasonal correlation and the $RE_t(M)$ value is shown. Plot(a) in Figure 3 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 ($c_\eta, c_\varepsilon$) increase from 5 to 10 to 20.

The effect of increasing the non-seasonal correlation whilst keeping the seasonal correlation constant is analysed for designs *a21, a31, a41* with correlation combinations *A1* to *E1*. From Table 2, this means that the seasonal correlation is kept at $\rho_\omega = 0.1$, and the non-seasonal correlations $\rho_\eta$, $\rho_\varepsilon$ are one of {0.2, 0.4, 0.6, 0.8, 1.0}. The results for $T = 40$ from each of the $RE_t(M)$ time series have been plotted against the non-seasonal correlation and given in Plot (b) in

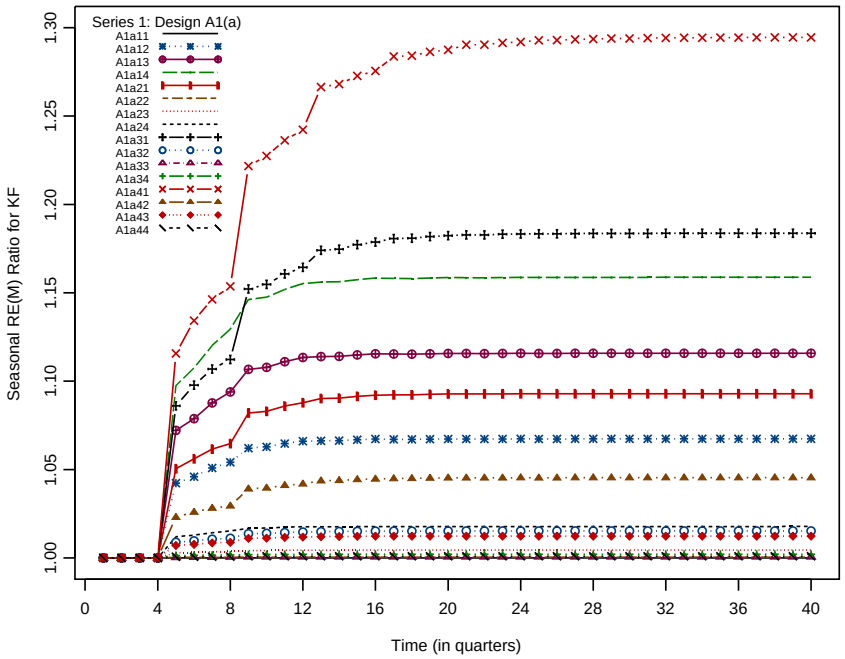



Figure 1: Results of $RE_t(M)$ for sub-series design '$a$' with correlation settings *A1* ($\rho_\omega = 0.1$, and $\rho_\eta = \rho_\varepsilon = 0.2$).

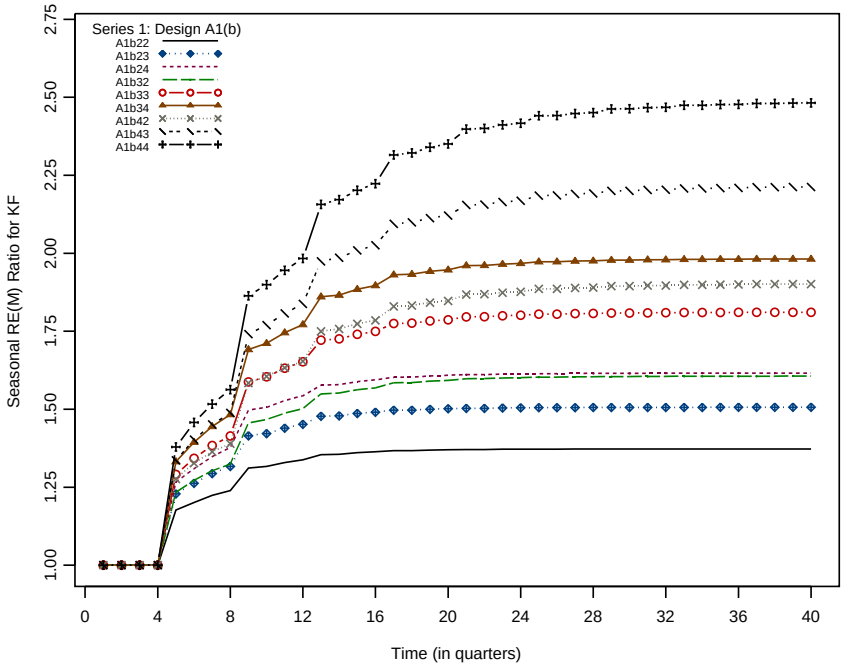
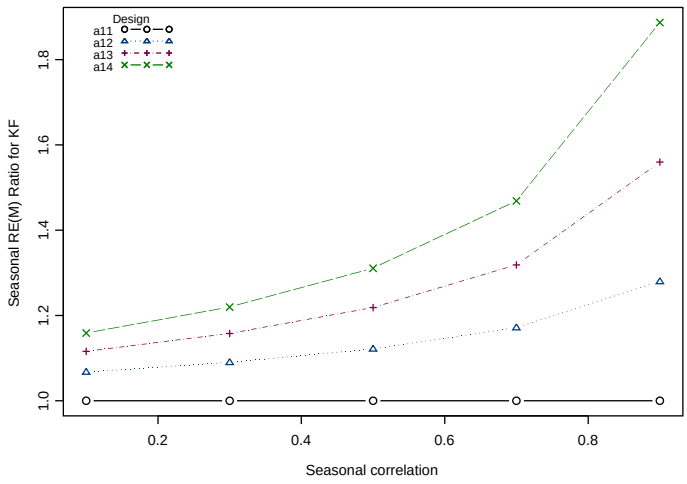


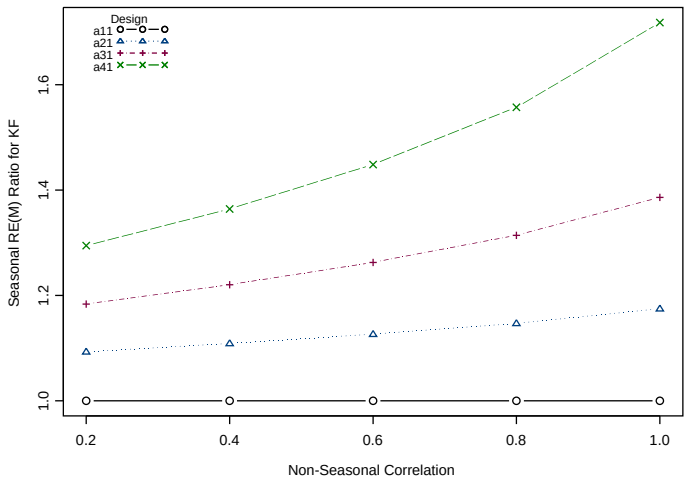

Figure 2: Results of $RE_t(M)$ for sub-series design '$b$' with correlation settings *A1* ($\rho_\omega = 0.1$, and $\rho_\eta = \rho_\varepsilon = 0.2$).

Table 3: Results of $RE_{40}(M)$ for sub-series design '$a$' and '$b$' with correlation settings *A1* ($\rho_\omega = 0.1$, and $\rho_\eta = \rho_\varepsilon = 0.2$).

| | | $c_\eta$ and $c_\varepsilon$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 5 | 10 | 20 | 0.2 | 0.1 | 0.05 |
| $c_\omega$ | 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 |




(a)




(b)

Figure 3: (a) $RE_{40}(M)$ versus seasonal correlation for designs *a11*, *a12*, *a13*, *a14* with *A1-A5*. (b) $RE_{40}(M)$ versus non-seasonal correlation for designs *a11*, *a21*, *a31*, *a41* for *A1-E1*.

Figure 3. The plot also shows the results for design *a11*. It can be seen that the impact of the increasing non-seasonal correlation is dependent upon the relative values of the variance parameters. There seems to be an interaction between the magnitude of $c_\omega$ and the non-seasonal correlation, since the gradient of the curve increases as both $c_\omega$ and $\rho_\eta, \rho_\varepsilon$ increase.

## 7 Other aggregate series

The empirical results presented so far in this paper are for a particular aggregate series with known parameters (5.3). The total series parameters have remained constant, whilst the sub-series parameters were varied. There is an infinite set of total series for which this study could be repeated. In this section, it is shown that similar results are obtained when other aggregate series are used.

To obtain an idea of what might occur if the total series had different parameters to those already chosen, some results are produced for two other series. For ease of reference, the total series previously considered is now referred to as *Series 1*, and consequently, the others are referred to as *Series 2*, and *Series 3*. The model and associated parameters for each will be described in the following subsections.

**Series 1**: As a reminder, *Series 1* is modelled by a local level seasonal BSM (given by (3.2), (3.3) with (3.4)), which has a level component without a slope, and a dummy seasonal component for quarterly data. The parameters are the variances of the disturbance terms, specified as:

$$\textit{Series 1}: \qquad \sigma^2_{tot,\eta} = 0.01, \quad \sigma^2_{tot,\omega} = 1, \quad \sigma^2_{tot,\varepsilon} = 1.$$

The relationship between these univariate parameters may be described by the seasonal to non-seasonal ratio for the series:

$$\frac{\sigma^2_{tot,\omega}}{\sigma^2_{tot,\eta} + \sigma^2_{tot,\varepsilon}} = \frac{1.0}{0.01 + 1.0} = 0.99.$$

**Series 2**: For *Series 2*, the local level seasonal BSM model is retained but the level parameter is increased to 0.5 (as compared to 0.01 for *Series 1*). The resulting univariate parameters are:

$$\textit{Series 2}: \qquad \sigma^2_{tot,\eta} = 0.5, \quad \sigma^2_{tot,\omega} = 1, \quad \sigma^2_{tot,\varepsilon} = 1.$$

The seasonal to non-seasonal ratio for *Series 2* is given by:

$$\frac{\sigma^2_{tot,\omega}}{\sigma^2_{tot,\eta} + \sigma^2_{tot,\varepsilon}} = \frac{1.0}{0.5 + 1.0} = 0.67.$$

**Series 3**: The BSM model including a slope term was fitted to the well known data set, 'airline passengers' described in Harvey (1989, page 93). The monthly airline passengers data was aggregated to produce a quarterly time series and logarithms were taken. Harvey (1989, page 94) specifies the BSM parameters to be

$$\begin{aligned} \textit{Series 3}: \qquad \sigma^2_{tot,\eta} &= 5.32 \times 10^{-4}, \quad \sigma^2_{tot,\omega} = 1.32 \times 10^{-4}, \\ \text{and} \qquad \sigma^2_{tot,\zeta} &= 1.08 \times 10^{-6}, \quad \sigma^2_{tot,\varepsilon} = 0. \end{aligned} \tag{7.1}$$

Since the measurement error variance is zero, $\varepsilon_{U,\,t}$ is excluded from the model equations. In the model for *Series 1* and *Series 2*, the correlated measurement error terms for the sub-series are put into the state vector due to the restriction in the software package. This is not necessary here. Thus, rather than the local level seasonal model, *Series 3* is generated from a BSM with level, slope and dummy seasonal components, and has no measurement error term. The parameters in (7.1), as well as some initial values, are used to generate data for *Series 3* from the model equations.

The seasonal to non-seasonal ratio for *Series 3* is given by:

$$\frac{\sigma^2_{tot,\omega}}{\sigma^2_{tot,\eta}+\sigma^2_{tot,\zeta}} = \frac{1.32\times10^{-4}}{(5.32\times10^{-4})+(1.08\times10^{-6})} = 0.25.$$

*Series 1, 2* and *3* have different parameters and as a result have different seasonal to non-seasonal ratios. *Series 2* and *Series 3* are both disaggregated into two sub-series, as carried out for *Series 1*, described in Section 5.

### 7.1 Results for other aggregate series

Results for *Series 2* and *Series 3* are presented here, and correspond to those for *Series 1* given in Section 6. The relative efficiency, $RE_t(M)$, has been calculated for each combination of the c-ratios given in Table 1 for *Series 2*. Only one correlation design has been considered here, namely *A1*, for which $\rho_\omega = 0.1$ and $\rho_\eta, \rho_\varepsilon = 0.2$. For *Series 3*, the c-ratio for the measurement error, $c_\varepsilon$, does not exist. It is replaced by $c_\zeta$ which is the analogous ratio of variances for the slope component between the series, thus

$$c_\zeta = \frac{\mathrm{Var}(R_{1t})}{\mathrm{Var}(R_{2t})} = \frac{\sigma^2_\zeta + \sigma^2_{1\zeta^*}}{\sigma^2_\zeta + \sigma^2_{2\zeta^*}}. \tag{7.2}$$

The results for $RE_{40}(M)$ for design '*a*' are found in Table 4. For comparison purposes, the results for *Series 2* and *Series 3* are shown together with the previous results for *Series 1*. Looking at the results for the three series, it is clear that the parameters of the total series do not greatly affect the relative efficiency. There are some small differences, notably for designs *a13, a14*, and *a41*. For designs *a13* and *a14*, the relative efficiency increases from *Series 1* to *Series 3*. However, for some designs which have a high $c_\omega$ and low non-seasonal c-ratios, such as design *a41*, the relative efficiency decreases; in this case it decreases from 1.29 to 1.26. Note that design *a41* has the highest relative efficiency for each of the total series.

Table 5 shows the results of $RE_{40}(M)$ for design '*b*'. The results of relative efficiency for *Series 2* and *3* show similar patterns as those for *Series 1*. Design *b44* has the highest relative efficiency for each series, with the largest ($RE_{40}(M) = 2.6$) being for *Series 3*, the airline series. For design *b24*, the relative efficiency increases from 1.62 to 1.77 from *Series 1* to *Series 3*.

The relative efficiency has also been plotted over time for *Series 2* and *Series 3*, as done previously for *Series 1*. Figure 4 plots (a) and (b) show the results of $RE_t(M)$ for design '*a*' for *Series 2* and *Series 3* respectively. Each plot has the same vertical scale and the same legend for comparison with Figure 1. Comparing the three plots, it is noted that the rankings of the designs are

Table 4: *Series 1, 2*, and *3*: results of $RE_{40}(M)$ for design *A1a*.

| | Design '$a$' | | | $c_\eta$ and $c_\varepsilon$ (or $c_\zeta$) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 1 | | 5 | | 10 | | 20 | |
| $c_\omega$ | 1 | Series 1 | *a11* | 1.0000 | *a12* | 1.0674 | *a13* | 1.1158 | *a14* | 1.1588 |
| | | Series 2 | | 1.0000 | | 1.0732 | | 1.1299 | | 1.1840 |
| | | Series 3 | | 1.0000 | | 1.0769 | | 1.1416 | | 1.2092 |
| | 5 | Series 1 | *a21* | 1.0929 | *a22* | 1.0005 | *a23* | 1.0045 | *a24* | 1.0178 |
| | | Series 2 | | 1.0949 | | 1.0005 | | 1.0048 | | 1.0199 |
| | | Series 3 | | 1.0923 | | 1.0005 | | 1.0050 | | 1.0216 |
| | 10 | Series 1 | *a31* | 1.1837 | *a32* | 1.0152 | *a33* | 1.0008 | *a34* | 1.0021 |
| | | Series 2 | | 1.1827 | | 1.0156 | | 1.0008 | | 1.0022 |
| | | Series 3 | | 1.1719 | | 1.0151 | | 1.0008 | | 1.0023 |
| | 20 | Series 1 | *a41* | 1.2945 | *a42* | 1.0454 | *a43* | 1.0124 | *a44* | 1.0010 |
| | | Series 2 | | 1.2849 | | 1.0450 | | 1.0125 | | 1.0010 |
| | | Series 3 | | 1.2581 | | 1.0422 | | 1.0121 | | 1.0010 |

Table 5: *Series 1, 2*, and *3*: results of $RE_{40}(M)$ for design *A1b*.

| | Design '$b$' | | $c_\eta$ and $c_\varepsilon$ (or $c_\zeta$) | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 0.2 | | 0.1 | | 0.05 | |
| $c_\omega$ | 5 | Series 1 | *b22* | 1.3728 | *b23* | 1.5063 | *b24* | 1.6158 |
| | | Series 2 | | 1.3946 | | 1.5532 | | 1.6963 |
| | | Series 3 | | 1.4000 | | 1.5839 | | 1.7705 |
| | 10 | Series 1 | *b32* | 1.6060 | *b33* | 1.8108 | *b34* | 1.9818 |
| | | Series 2 | | 1.6246 | | 1.8636 | | 2.0849 |
| | | Series 3 | | 1.6092 | | 1.8769 | | 2.1576 |
| | 20 | Series 1 | *b42* | 1.9014 | *b43* | 2.2125 | *b44* | 2.4820 |
| | | Series 2 | | 1.9004 | | 2.2529 | | 2.5939 |
| | | Series 3 | | 1.8340 | | 2.2030 | | 2.6072 |

almost identical, with the main differences being seen for designs *a14* and *a31*. For *Series 1*, the time plots intersect at $t = 8$ with *a31* eventually having the greater gain. For *Series 2*, the gain for design *a14* climbs more quickly but then eventually becomes almost the same as for *a31*. For *Series 3*, the gain for design *a14* is always larger than for *a31*.

For some designs, convergence rates also differ across the plots. A marked difference between *Series 1* and *Series 2* is that the level parameter increases from 0.01 to 0.5, thereby decreasing the seasonal to non-seasonal ratio from 0.99 to 0.67. *Series 2* seems to have a slower convergence rate than that for *Series 1*. For *Series 3*, which has a seasonal to non-seasonal ratio of only 0.25, the plot shows slower convergence than for the other two series.

Figure 4 plots (c) and (d) show the results for $RE_t(M)$ for design '$b$' for *Series 2* and *Series 3* respectively. Again, each plot has the same vertical scale and legend for comparison with Figure 2. The rankings of the nine designs are again similar for each of the three series. The designs which differ here the most are *b43* and *b34*. Convergence rates are again slower for *Series 3* than for the other two series.

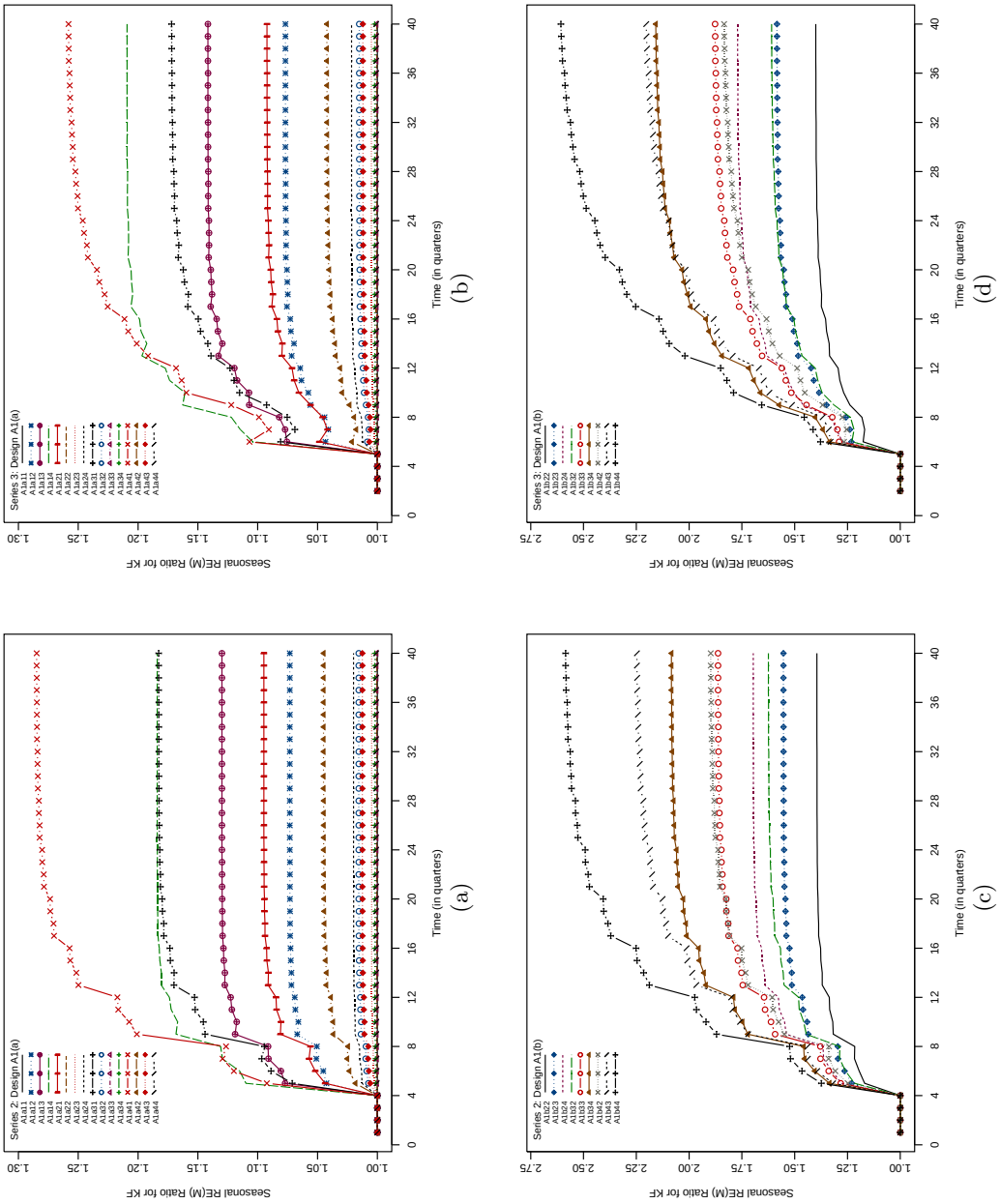



Figure 4: (*a*) $RE_t(M)$ for designs *A1a11* to *A1a44* given by *Series 2* , and in (*b*) for *Series 3*. (*c*) $RE_t(M)$ for designs *A1b22* to *A1b44* given by *Series 2*, and in (*d*) for *Series 3*.

## 8 Conclusion

In this empirical study, the relative efficiency of the seasonally adjusted aggregate series has been investigated by using a multivariate structural time series model applied to the non-stationary sub-series. It focuses on one particular local level seasonal aggregate series and utilises a selection of designs for two sub-series. Keeping the aggregated parameters fixed, 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), then there is no difference between the multivariate and the univariate methods. This is due to the design being close to the homogeneous system.

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 components, 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 study shows that even when the correlations between the series are low, gains are achievable with a multivariate model with only two sub-series.

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.

Fourthly, 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 correlations are increased. Thus, better gains are achieved when the seasonal and non-seasonal correlations are at opposite ends of the (positive) correlation scale.

Lastly, this study also examines the evolution of relative efficiency over time. For the first four time points, the multivariate method and univariate method yield exactly the same MSEs for the filtered estimates. 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 of the relative efficiency reaches a steady state. Those with higher c-ratios for the seasonal component tend to be slowest to converge.

This paper reports the results of modelling two disaggregate series of a particular aggregate time series with known parameters. Two other aggregate series with different seasonal to non-seasonal ratios are also studied and they both produce similar results. The method proposed here may be extended to aggregated series with more than two sub-series (Birrell, 2008).

# A Appendix

## A.1 State space form for the univariate model

The state space form of the univariate LLS model as described in Section 3.1 is given by:

$$Y_t = \mathbf{Z}\alpha_t + \varepsilon_{U,\,t}, \tag{A.1}$$
$$\alpha_{t+1} = \mathbf{T}\alpha_t + \mathbf{G}\gamma_t, \tag{A.2}$$

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

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

$$\mathbf{Z} = \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix}, \qquad \varepsilon_{U,\,t} \sim N(0, \mathrm{H}),$$
$$\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} \sigma^2_{U,\,\eta} & 0 \\ 0 & \sigma^2_{U,\,\omega} \end{pmatrix},$$

$$\mathrm{Var}\,(\mathbf{G}\gamma_t) = \mathbf{G}\mathbf{Q}\mathbf{G}' = \begin{pmatrix} \sigma^2_{U,\,\eta} & 0 & 0 & 0 \\ 0 & \sigma^2_{U,\,\omega} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathrm{H} = \sigma^2_{U,\,\varepsilon}. \tag{A.3}$$

The standard set of filtering equations which make up the Kalman filter may be found in Chapter 4 of Durbin and Koopman (2001). These equations are applied to the aggregate series data and yield the estimates of the components for each time point $\mathbf{a}_{t|t}$ and their MSEs, $\mathbf{P}_{t|t}$.

The variance matrix, $\mathbf{P}_1$, of the initial state vector $\alpha_1$, is assumed to have the form:

$$\mathbf{P}_1 = \kappa\mathbf{P}_{\infty,1} + \mathbf{P}_{*,1}, \tag{A.4}$$

where $\kappa$ is a large scalar value, $\mathbf{P}_{*,1}$ is the covariance matrix of the stationary components in $\alpha_1$ and $\mathbf{P}_{\infty,1}$ is the covariance matrix of the non-stationary components in $\alpha_1$ (Zivot and Wang, 2006).

Non-stationary components in the state vector require diffuse initialisation. A state is called diffuse if its covariance matrix is arbitrarily large. The problem of dealing with diffuse initial states in the Kalman filter was first solved by de Jong (1991). An alternative approach of dealing with diffuse initial conditions is to apply the exact initial Kalman filter as described in detail in Koopman and Durbin (2000). The exact approach is computationally more efficient in the process of parameter estimation when compared to other initialisation strategies such as that of de Jong (1991) and Koopman and Durbin (2000, p293).

In particular, for the univariate local level seasonal BSM with a dummy seasonal component model, $\mathbf{a}_1 = E(\alpha_1)$ is a $4 \times 1$ zero vector, $\mathbf{P}_{\infty,1}$ is a $4 \times 4$ identity matrix and $\mathbf{P}_{*,1}$ is a $4 \times 4$ zero matrix. The term 'zero vector' is a vector in which each element is zero, and 'zero matrix' is used to describe a matrix in

which each element is zero. The exact initial Kalman filter can be applied using the `S+FinMetrics` software, in the *SsfPack* set of functions (Koopman *et al.*, 1999).

### A.2 State space form for the multivariate model

The multivariate model in Section 3.2 is transformed as described earlier in Section 4. As a result of the transformation on two sub-series, the two series in the system are the total series, $Y_{tot,\,t}$, and series 1, denoted by $Y_{1t}$. The transformation permits estimation of the seasonal component of the total series, $S_{tot,\,t}$, within a multivariate framework.

Due to the common disturbance term, $\varepsilon_t$, in (3.5), 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. Their suggested approach has been implemented in this paper.

The state space model for the transformed system, $\mathbf{Y}_{(M),\,t}$ (4.1), with $K = 2$ may be specified as follows, given that the measurement errors are placed within the state vector:

$$\begin{aligned} \mathbf{Y}_{(M),\,t} &= (\mathbf{Z}_{(m)} \otimes \mathbf{I}_2)\alpha_{(M),\,t}, \\ \alpha_{(M),\,t+1} &= (\mathbf{T}_{(m)} \otimes \mathbf{I}_2)\alpha_{(M),\,t} + (\mathbf{G}_{(m)} \otimes \mathbf{I}_2)\gamma_{(M),\,t}, \end{aligned} \tag{A.5}$$

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

$$\begin{aligned} \alpha_{(M),\,t} &= [L_{tot,\,t}, L_{1t}, S_{tot,\,t}, S_{1t}, S_{tot,\,t-1}, S_{1,\,t-1}, S_{tot,\,t-2}, S_{1,\,t-2}, \\ &\quad \varepsilon_{tot,\,t}, (\varepsilon_t + \varepsilon^*_{1t})]', \\ \gamma_{(M),\,t} &= \big[\, \eta_{tot,\,t}, (\eta_t + \eta^*_{1t}),\ \omega_{tot,\,t}, (\omega_t + \omega^*_{1t}), \varepsilon_{tot,\,t+1},\ (\varepsilon_{t+1} + \varepsilon^*_{1,\,t+1})\big]' \end{aligned} \tag{A.6}$$

with $\alpha_{(M),\,1} \sim N(\mathbf{a}_{(M),\,1}, \mathbf{P}_{(M),\,1})$. The system matrices are given by:

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

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

$$\mathrm{Var}\left((\mathbf{G}_{(m)} \otimes \mathbf{I}_2)\gamma_{(M),\,t}\right) = \begin{pmatrix} \boldsymbol{\Sigma}_{(M),\,\eta} & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & \boldsymbol{\Sigma}_{(M),\,\omega} & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \boldsymbol{\Sigma}_{(M),\,\varepsilon} \end{pmatrix}, \tag{A.7}$$

where $\mathbf{0}_2$ is a $2 \times 2$ zero matrix. The transformed covariance matrix for the level component (refer to (3.8)) is denoted by $\boldsymbol{\Sigma}_{(M),\,\eta}$:

$$\boldsymbol{\Sigma}_{(M),\,\eta} = \begin{pmatrix} \sigma^2_{tot,\eta} & 2\sigma^2_\eta + \sigma^2_{1\eta^*} \\ 2\sigma^2_\eta + \sigma^2_{1\eta^*} & \sigma^2_\eta + \sigma^2_{1\eta^*} \end{pmatrix}, \tag{A.8}$$

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

To compensate for the restructuring of the state vector, the set up of the exact initial conditions matrices described in Durbin and Koopman (2001, Section 5.2) is amended. The variance matrix, $\mathbf{P}_{(M),\,1}$, of the initial state vector, $\alpha_{(M),\,1}$, is given by

$$\mathbf{P}_{(M),1} = \kappa \mathbf{P}_{(M)\infty,1} + \mathbf{P}_{(M)*,1}, \tag{A.9}$$

The $\mathbf{P}_{(M)*,1}$ matrix in (A.9) holds the variance of the stationary part of $\alpha_{(M),\,1}$. It is a $10\times10$ zero matrix with the lower right $2\times2$ block diagonal replaced by the $\mathbf{\Sigma}_{(M),\,\varepsilon}$ covariance matrix. The $\mathbf{P}_{(M)\infty,1}$ matrix (also of dimension $10 \times 10$ here since $K = 2$) is an identity matrix but with the lower right $2 \times 2$ block diagonal replaced by a $2 \times 2$ zero matrix. For further details of the exact initialisation of the filter refer to Koopman and Durbin (2000).

## References


Birrell, C. L. (2008) *Efficiency gains for seasonal adjustment by joint modelling of disaggregated series*. Ph.D. thesis, University of Wollongong.

Burman, J. P. (1980) Seasonal adjustment by signal extraction. *Journal of the Royal Statistical Society, Series A*, **143**, 321–337.

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

Dagum, E. (1988) The X-11-ARIMA/88 seasonal adjustment method, foundations and user's manual. *Tech. rep.*, Time Series Research and Analysis Division, Statistics Canada.

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

Engle, R. F. (1978) Estimating structural models of seasonality. In *Seasonal Analysis of Economic Time Series* (ed. A. Zellner), 281–295. Washington DC: US Department of Commerce, Bureau of Census.

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

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

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

Gomez, V. and Maravall, A. (1996) Programs TRAMO and SEATS. Instructions for the user (with some updates). *Working Paper 9628*, Servicio de Estudios, Banco de Espana.

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

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

Harvey, A. and Todd, P. (1983) Forecasting economic time series with structural and Box-Jenkins models: A case study. In *Modelling Seasonality* (ed. S. Hylleberg), 341–358. Oxford University Press.

Hillmer, S. and Tiao, G. (1982) An ARIMA model-based approach to seasonal adjustment. In *Modelling Seasonality* (ed. S. Hylleberg), 321–339. Oxford University Press.

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

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

de Jong, P. (1991) The diffuse Kalman filter. *The Annals of Statistics*, **19**, 1073–1083.

Kalman, R. (1960) A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions, ASMA, Series D*, **82**, 35–45.

Koopman, S. (1997) Exact initial Kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association*, **92**, 1630–1638.

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

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

Ladiray, D. and Mazzi, G. (2003) Seasonal adjustment of European aggregates: direct versus indirect approach. In *Seasonal Adjustment* (eds. M. Manna and R. Peronaci), 37–65. Frankfurt am Main, Germany: European Central Bank.

Maravall, A. (1995) Unobserved components in economic time series. In *The Handbook of Applied Econometrics* (eds. M. Pesaran and M. Wickens), 12–72. Basil Blackwell, Oxford.

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

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

Shiskin, J., Young, A. and Musgrave, J. (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.

Taylor, J. B. (1978) Comments on 'The temporal and sectorial aggregation of seasonally adjusted time series' by John Geweke. In *Seasonal Analysis of Economic Time Series* (ed. A. Zellner), 431–432. U.S. Government Printing Office.

Whittle, P. (1963) *Prediction and Regulation using Least Squares Methods*. London: English Universities Press.

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