TI 2010-018/4
Tinbergen Institute Discussion Paper

Modeling Trigonometric Seasonal Components for Monthly Economic Time Series

Irma Hindrayanto[1,2]
John A.D. Aston[3]
Siem Jan Koopman[1,2]
Marius Ooms[1]

[1] *VU University Amsterdam;*
[2] *Tinbergen Institute;*
[3] *University of Warwick.*

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Modeling trigonometric seasonal components for monthly economic time series

Irma Hindrayanto*, John A D Aston†, Siem Jan Koopman*,‡, Marius Ooms*

January 23, 2010

Abstract

The basic structural time series model has been designed for the modelling and forecasting of seasonal economic time series. In this paper we explore a generalisation of the basic structural time series model in which the time-varying trigonometric terms associated with different seasonal frequencies have different variances for their disturbances. The contribution of the paper is two-fold. The first aim is to investigate the dynamic properties of this frequency specific basic structural model. The second aim is to relate the model to a comparable generalised version of the Airline model developed at the U.S. Census Bureau. By adopting a quadratic distance metric based on the restricted reduced form moving-average representation of the models, we conclude that the generalised models have properties that are close to each other compared to their default counterparts. In some settings, the distance between the models is almost zero so that the models can be regarded as observationally equivalent. An extensive empirical study on disaggregated monthly shipment and foreign trade series illustrates the improvements of the frequency-specific extension and investigates the relations between the two classes of models.

Key words: frequency-specific model, Kalman filter, model-based seasonal adjustment, unobserved components time series model.

*VU University Amsterdam and Tinbergen Institute Amsterdam, The Netherlands

†Department of Statistics, University of Warwick, United Kingdom, and Institute of Statistical Science, Academia Sinica, Taiwan

‡Corresponding author; phone: +31 20 598 60 10, e-mail: s.j.koopman@feweb.vu.nl

1 Introduction

The Airline model popularised by Box and Jenkins (1970) and the basic structural model (BSM) popularised by Harvey (1989) are amongst the most widely used models for seasonal adjustment. Their popularity can be attributed to their simplicity and accuracy for a wide range of seasonal economic time series. However, the simplicity of both models inevitably implies that there will be a substantial number of practical cases where either model is inadequate. In this paper, we consider a similar generalisation for both seasonal specifications. We aim to investigate the dynamic properties of the frequency specific basic structural model (FS-BSM) and to relate the model to a frequency specific version of the Airline model developed recently at the U.S. Census Bureau, as in Aston, Findley, McElroy, Wills, and Martin (2007).

The BSM belongs to the class of unobserved component time series models that decompose time series into trend, seasonal and irregular components. Here we focus on the seasonal component that is studied in detail by Proietti (2000,2004). In particular, we consider the trigonometric representation of the seasonal component. To address the criticism that the BSM can be too restrictive to fit seasonal time series adequately, we modify the BSM to be less restrictive in the specification of the seasonal component. Instead of having a single seasonal variance for all frequencies, we let the time-varying trigonometric terms associated with different seasonal frequencies have different variances. Therefore we develop the FS-BSMs that are more flexible than the standard BSM while still capable of producing component estimates. The extended set of parameters can be estimated using maximum likelihood procedures based on the Kalman Filter.

To illustrate that less restrictive models can be needed to fit seasonal time series, we focus on two particular time series from a database of 75 monthly seasonal time series provided by the U.S. Census Bureau. The two time series are presented in Figure 1. The first series has code X41140 and is one of the Foreign Trade series which corresponds to the *Final Export of Musical Instruments* from January 1989 through November 2001. The second series has code U37AVS and corresponds to *Shipments of Household Furniture and Kitchen Cabinet* from January 1992 through September 2001. Both time series y_t are differenced by $\Delta\Delta_{12}y_t = y_t - y_{t-1} - y_{t-12} + y_{t-13}$ after taking the natural logarithm. The autocorrelation functions (ACF) of the differenced series are also presented in Figure 1. We observe that for X41140, the only significant correlations are at lags 1, 11, 12 and 13, while those for U37AVS are at lags 1, 3, 5 and 12. Comparing the characteristics of these two time series, we learn that a basic model may suffice for X41140 while a more elaborate model is needed for U37AVS. Details of estimation and testing results for different models applied to the two time series are discussed in Section 4.

The procedure of ARIMA model based seasonal adjustment dates back to the early 1980s, see Burman (1980) and Hillmer and Tiao (1982), but the automatic implementation in seasonal adjustment software was carried out more than a decade later, see the documentation of SEATS from Gómez and Maravall (1996). Using this software, the Airline model is frequently chosen to identify seasonal time series. In the search for a useful alternative for the Airline model, Aston, Findley, McElroy, Wills, and Martin (2007) introduce extensions to the standard Airline

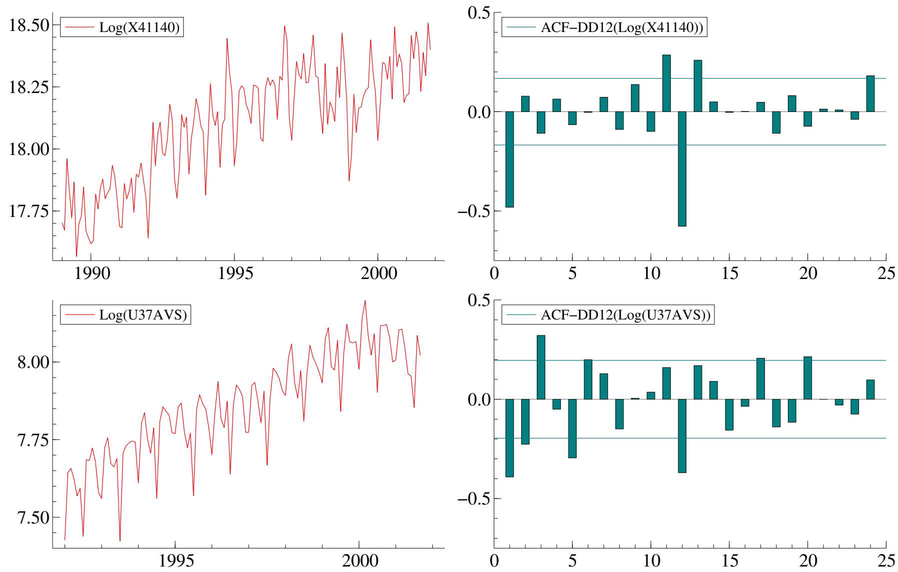

Figure 1: Motivating example. *Top:* Final Export of Musical Instruments series (X41140) and *bottom:* Manufacturers' Shipment on Household Furniture and Kitchen Cabinet series (U37AVS). The first column shows the series in natural logarithm and the second column depicts the ACF of the differenced series. X41140 does not require an FS model while U37AVS does require an FS model as the ACF shows more periodicities.

model by means of decomposing and re-parametrizing the seasonal moving average (MA) factor and partitioning the factors of different seasonal frequencies into two groups, each with its own coefficient. Because of the dependency of its parameters on the seasonal frequencies, the new model is called the frequency specific Airline model (FS-AM). In the empirical part of their research, they show that the FS models are preferred above the standard Airline model for 22 series out of the 75 selected US Census Bureau economic indicator series that we also examine here. The Airline model was known to be adequate for these 75 time series compared to other SARIMA models. The comparison between FS and non-FS Airline model is based on the Minimum Akaike's Information Criterion (MAIC) and a modification of it, the so-called $\mathcal{F}$-MAIC, see also Aston et al (2004, 2007).

The remainder of the paper is organised as follows. Section 2 discusses the FS-AM, introduces the general FS-BSM and provides an alternative representation for the FS-BSM. Section 3 introduces restricted specifications of the FS-BSM and FS-AM that we consider in this paper. It also discusses testing procedures for deterministic seasonal components and defines a distance metric to measure the difference between FS-BSM and FS-AM specifications. In Section 4 we investigate empirically whether FS models lead to increases in fit and to more similar decompositions compared to non-FS models. Here we consider a U.S. Census Bureau database of seasonal time series that has been analysed previously with the FS-AMs of Aston, Findley, McElroy, Wills, and Martin (2007). A concluding review is presented in Section 5.

2 Frequency specific time series models

In this section we discuss two classes of frequency specific seasonal time series models and develop a common stationary representation for both classes. We adopt the following notation. We define L as the lag operator with $L^p y_t = y_{t-p}$ for any $p = 0, 1, \ldots$, Δ as the difference operator with $\Delta y_t = (1-L)y_t = y_t - y_{t-1}$, Δ_s as the seasonal difference operator with $\Delta_s y_t = y_t - y_{t-s}$ and $S(L)$ as the seasonal sum operator with $S(L)y_t = \sum_{i=0}^{s-1} L^i y_t = y_t + y_{t-1} + \ldots + y_{t-s+1}$ for any seasonal length $s = 2, 3, \ldots$. The moving average model of order q is denoted by MA(q) and is given by

$$y_t = \theta(L)\varepsilon_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}, \qquad \varepsilon_t \sim \text{NID}(0, \sigma^2),$$

with lag polynomial $\theta(L) = 1 + \theta_1 L + \ldots + \theta_q L^q$, for any $q = 1, 2, \ldots$, and where coefficients θ_i are fixed, for $i = 1, \ldots, q$, and the disturbance ε_t is normally and independently distributed with mean zero and variance σ^2.

2.1 Frequency specific Airline model

The frequency-specific Airline model (FS-AM) is developed by Aston, Findley, McElroy, Wills, and Martin (2007). For a seasonal time series y_t, the Airline model is given by

$$(1-L)(1-L^s)y_t = (1-\theta L)(1-\Theta L^s)\varepsilon_t, \tag{2.1}$$

where $s \geq 2$. When $\Theta > 0$, the Airline model (2.1) can be written as

$$\begin{aligned}(1-L)(1-L^s)y_t &= (1-\theta L)(1-\Theta^{1/s}L)\left(\textstyle\sum_{j=0}^{s-1} \Theta^{j/s} L^j\right)\varepsilon_t, \\ &= \left(1 - (\theta + \Theta^{1/s})L + \theta\Theta^{1/s}L^2\right)\left(\textstyle\sum_{j=0}^{s-1} \Theta^{j/s} L^j\right)\varepsilon_t, \end{aligned} \tag{2.2}$$

where the non-seasonal polynomial $1 - (\theta + \Theta^{1/s})L + \theta\Theta^{1/s}L^2$ and the seasonal polynomial $\sum_{j=0}^{s-1} \Theta^{j/s} L^j$ both contain $\Theta^{1/s}$. Therefore coefficient Θ both affects the trend and seasonal dynamic properties. Aston, Findley, Wills, and Martin (2004) replace the restrictive polynomials in (2.2) with two coefficients by less restrictive polynomials with three coefficients and obtain

$$(1-L)(1-L^s)y_t = \left(1 - aL + bL^2\right)\left(\textstyle\sum_{j=0}^{s-1} c^j L^j\right)\varepsilon_t, \tag{2.3}$$

where the seasonal sum polynomial relies on coefficient c that is distinct from the non-seasonal coefficients a and b. Aston, Findley, McElroy, Wills, and Martin (2007) decompose the seasonal factor $\sum_{j=0}^{s-1} c^j L^j$ further into several factors with different coefficients for different frequencies. By expanding the seasonal factor

$$\sum_{j=0}^{s-1} c^j L^j = (1 + cL)\prod_{i=1}^{s/2-1}\left(1 - 2c\cos(2\pi i/s)L + c^2L^2\right), \tag{2.4}$$

a generalisation of the right-hand-side with different coefficients c_i at different seasonal frequencies $\frac{2\pi \cdot i}{s}$ is obtained. For monthly series, the generalisation becomes

$$(1 + c_6 L)\prod_{i=1}^{5}\left(1 - 2c_i\cos(2\pi i/12)L + c_i^2 L^2\right).$$

Hence the frequency specific Airline model for $s = 12$ is then given by

$$(1-L)(1-L^{12})y_t = (1-aL-bL^2)\left[(1+c_6L)\prod_{i=1}^{5}\left(1-2c_i\cos(2\pi i/12)L+c_i^2L^2\right)\right]\varepsilon_t, \quad (2.5)$$

which is a less parsimonious model representation with the nine coefficients a, b, $c_1, \ldots, c_6$, and σ^2. For many typical monthly macroeconomic time series of moderate length (say 8 to 13 years), the estimated coefficients often imply that several roots of the seasonal MA polynomial are close to the unit circle. Aston, Findley, McElroy, Wills, and Martin (2007) alleviate the unit root problem by proposing more parsimonious formulations that consist of two or three seasonal MA coefficients.

2.2 Frequency specific basic structural model

In this section we give details of FS-BSM and its variations, starting from the most general form of FS-BSM to the grouped FS-BSMs. Let the time series observation y_t at time t be modelled as the sum of the trend μ_t, the seasonal term γ_t and the irregular disturbance ϵ_t. The FS-BSM is then given by,

$$y_t = \mu_t + \gamma_t + \epsilon_t, \qquad \epsilon_t \sim \text{NID}(0, \sigma_\epsilon^2), \quad (2.6)$$

for $t = 1, \ldots, n$, where the trend μ_t is specified as

$$\mu_{t+1} = \mu_t + \beta_t + \eta_t, \qquad \eta_t \sim \text{NID}(0, \sigma_\eta^2), \quad (2.7)$$

$$\beta_{t+1} = \beta_t + \zeta_t, \qquad \zeta_t \sim \text{NID}(0, \sigma_\zeta^2), \quad (2.8)$$

and the seasonal component γ_t follows the trigonometric specification as given by

$$\gamma_t = \sum_{j=1}^{s/2} \gamma_{j,t}, \quad (2.9)$$

$$\begin{pmatrix} \gamma_{j,t+1} \\ \gamma_{j,t+1}^* \end{pmatrix} = \begin{pmatrix} \cos\lambda_j & \sin\lambda_j \\ -\sin\lambda_j & \cos\lambda_j \end{pmatrix} \begin{pmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{pmatrix} + \begin{pmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{pmatrix}, \qquad \begin{pmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{pmatrix} \sim \text{NID}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_{\omega,j}^2 I_2\right], \quad (2.10)$$

with $\lambda_j = 2\pi j/s$ as the j-th seasonal frequency, $j = 1, \ldots s/2$, and I_2 is the 2×2 identity matrix. We assume s is even to simplify notation. The extension to odd s does not add new insights. In monthly time series ($s = 12$), equation (2.10) implies a seasonal component with six different variances. Further it is assumed that all disturbances in the model are mutually and serially uncorrelated at all leads and lags. The standard BSM is obtained when all seasonal variances $\sigma_{\omega,j}^2$ are equal, that is $\sigma_{\omega,j}^2 = \sigma_\omega^2$ for all $j = 1, \ldots, s/2$. More details and properties of the BSM are given by Harvey (1989). Note that we have $s = 12$ in all applications in this paper.

2.3 Stationary form of FS-BSM

The stationary formulation of the seasonal component γ_t in (2.9) and the time series y_t are required for determining the ACF of $S(L)\gamma_t$ and $\Delta\Delta_s y_t$. An alternative specification of the seasonal component γ_t is given by

$$\gamma_t = \sum_{j=1}^{s/2} \gamma_{j,t} = \sum_{j=1}^{s/2-1} \frac{(1-\cos\lambda_j L)\omega_{j,t-1} + (\sin\lambda_j L)\omega_{j,t-1}^*}{1 - 2\cos\lambda_j L + L^2} + \frac{\omega_{s/2,t-1}}{(1+L)}, \tag{2.11}$$

for s is even, see Harvey (1989) for more details. Bell (1993) has shown that the trigonometric specification (2.11) is a sum of ARIMA components. The numerator of (2.11) is an MA(1) process for each seasonal frequency so that we can rewrite the above equation as

$$\gamma_t = \sum_{j=1}^{s/2} \frac{(1-\alpha_j L)w_{j,t}}{\delta_j(L)}, \tag{2.12}$$

where α_j is a fixed constant for a given s, $w_{j,t} \sim \text{NID}(0, \sigma_j^2)$ with σ_j^2 a function of $\sigma_{\omega,j}^2$ in (2.10) and $\delta_j(L) = 1 - 2\cos\lambda_j L + L^2$ for each seasonal frequency $j = 1, \ldots, s/2$. In particular, $\sigma_j^2 = 2\sigma_{\omega,j}^2/(1+\alpha_j^2)$ for $j = 1, \ldots, s/2-1$, and $\sigma_{s/2}^2 = \sigma_{\omega,s/2}^2$ for $j = s/2$. The specifications for $\delta_j(L)$, α_j and σ_j^2 for $s = 12$ are given in Table 1 as taken from Bell (2004). Since $\prod_{j=1}^{s/2} \delta_j(L) = S(L)$, we have

$$S(L)\gamma_t = \sum_{j=1}^{s/2} S(L)\gamma_{j,t} = \sum_{j=1}^{s/2} \frac{S(L)}{\delta_j(L)}(1-\alpha_j L)w_{j,t} = \sum_{j=1}^{s/2}\prod_{i\neq j} \delta_i(L)(1-\alpha_j L)w_{j,t}, \tag{2.13}$$

where the right hand side is the sum of $s/2$ independent MA$(s-2)$ processes since there is no power in the polynomial $\prod_{i\neq j}\delta_i(L)(1-\alpha_j L)$ that is higher than $s-2$ for each seasonal frequency j, see Harvey (1989, §2.4.3) for more details.

The FS-BSMs and FS-AMs can be most conveniently related by expressing both models in their stationary MA formulations. For the FS-BSM, we specify the trend μ_t as in (2.7)-(2.8) by

$$\mu_t = \frac{\eta_{t-1}}{\Delta} + \frac{\zeta_{t-2}}{\Delta^2}, \tag{2.14}$$

with $\Delta^m = (1-L)^m$. The stationary form of the seasonal component is given by (2.13). Substituting equations (2.13) and (2.14) into the FS-BSM (2.6) yields

$$y_t = \frac{\eta_{t-1}}{\Delta} + \frac{\zeta_{t-2}}{\Delta^2} + \frac{\sum_{j=1}^{s/2}\prod_{i\neq j}\delta_i(L)(1-\alpha_j L)w_{j,t}}{S(L)} + \epsilon_t. \tag{2.15}$$

The minimum order of differencing for y_t is given by $\Delta^2 S(L)$ and the stationary MA formulation of y_t in the FS-BSM becomes

$$\Delta\Delta_s y_t = \Delta_s\eta_{t-1} + S(L)\zeta_{t-2} + \Delta^2 \sum_{j=1}^{s/2}\prod_{i\neq j}\delta_i(L)(1-\alpha_j L)w_{j,t} + \Delta\Delta_s\epsilon_t. \tag{2.16}$$

The highest polynomial order is $s+1$. The stationary time series $\Delta\Delta_s y_t$ in (2.16) can therefore be represented by an MA$(s+1)$ process with at most $3 + s/2$ coefficients.

Table 1: ARIMA representation (2.12) for the individual trigonometric seasonal components with $s = 12$

j	λ_j	$\delta_j(L)\gamma_{j,t}$	$=$	$(1-\alpha_j L)w_{j,t}$	σ_j^2
1	$\pi/6$	$(1-\sqrt{3}L+L^2)\gamma_{1,t}$	$=$	$(1-\frac{1}{3}\sqrt{3}L)w_{1,t}$	$1.5\sigma^2_{\omega,1}$
2	$\pi/3$	$(1-L+L^2)\gamma_{2,t}$	$=$	$(1-(2-\sqrt{3})L)w_{2,t}$	$(1+\frac{1}{2}\sqrt{3})\sigma^2_{\omega,2}$
3	$\pi/2$	$(1+L^2)\gamma_{3,t}$	$=$	$w_{3,t}$	$2\sigma^2_{\omega,3}$
4	$2\pi/3$	$(1+L+L^2)\gamma_{4,t}$	$=$	$(1+(2-\sqrt{3})L)w_{4,t}$	$(1+\frac{1}{2}\sqrt{3})\sigma^2_{\omega,4}$
5	$5\pi/6$	$(1+\sqrt{3}L+L^2)\gamma_{5,t}$	$=$	$(1+\frac{1}{3}\sqrt{3}L)w_{5,t}$	$1.5\sigma^2_{\omega,5}$
6	π	$(1+L)\gamma_{6,t}$	$=$	$w_{6,t}$	$\sigma^2_{\omega,6}$

3 Design of empirical study

To investigate the frequency specific models in more detail, we carry out an extensive empirical study in the next section. Given the flexible nature of FS models and the need for parsimony, we introduce a set of restricted FS models in section § 3.1. Only these model classes will be considered in our empirical study. The study also relies on tests for deterministic seasonal components and tests for the null hypothesis of a non-FS model. Section § 3.2 discusses some details of these tests. To relate the dynamic properties of the FS-BSM with the FS-AM we propose a distance measure in section § 3.3. The behaviour of the distance measure is investigated in a limited Monte Carlo experiment in section § 3.4.

3.1 Classes of restricted FS-AM and FS-BSM specifications

The unrestricted form of the FS-AM for a monthly time series is given by (2.5). It shows that the monthly FS-AM allows an MA(13) representation for $\Delta\Delta_{12}y_t$ in a similar way that the monthly FS-BSM in (2.6) has the MA(13) representation (2.16) for$\Delta\Delta_{12}y_t$. Both unrestricted models have nine parameters. To obtain more parsimonious specifications, we consider the following restrictive specifications.

Different restrictions can be imposed on the FS-BSM class of models. In our study we consider the following set of FS-BSM restrictions. We reduce the number of parameters to five. The variances of the trend and irregular (non-seasonal) components σ^2_η, σ^2_ζ and σ^2_ϵ are unrestricted. The number of seasonal variances are reduced to two and are denoted by $\sigma^2_{\omega,\mathrm{I}}$ and $\sigma^2_{\omega,\mathrm{II}}$. The following seasonal restrictions are considered.

FS-BSM($\{i\}/s$) The seasonal variance corresponding to a single frequency $\{i\}/s$ is set differently from those for the remaining $(s/2)-1$ frequencies with $i = 1,\ldots,s/2$. For example, the FS-BSM($\{3\}/12$) imposes the restrictions

$$\sigma^2_{\omega,3} = \sigma^2_{\omega,\mathrm{I}} \qquad \text{and} \qquad \sigma^2_{\omega,1} = \sigma^2_{\omega,2} = \sigma^2_{\omega,4} = \sigma^2_{\omega,5} = \sigma^2_{\omega,6} = \sigma^2_{\omega,\mathrm{II}}.$$

FS-BSM($\{i, j\}/s$) A pair of variances associated with the seasonal frequencies i/s and j/s are set equal to $\sigma^2_{\omega,\mathrm{I}}$ while the seasonal variances for the remaining frequencies are set equal to $\sigma^2_{\omega,\mathrm{II}}$ where $i < j$ and $i, j = 1, \ldots, s/2$. The FS-BSM($\{1, 2\}/12$) for example has the restrictions

$$\sigma^2_{\omega,1} = \sigma^2_{\omega,2} = \sigma^2_{\omega,\mathrm{I}} \qquad \text{and} \qquad \sigma^2_{\omega,3} = \cdots = \sigma^2_{\omega,6} = \sigma^2_{\omega,\mathrm{II}}.$$

FS-BSM($\{i, j, k\}/s$) The variances associated with three seasonal frequencies $i/s, j/s$ and k/s are set equal to $\sigma^2_{\omega,\mathrm{I}}$ while the seasonal variance for the remaining three frequencies are set equal to $\sigma^2_{\omega,\mathrm{II}}$ where $i < j < k$ and $i, j, k = 1, \ldots, s/2$. An example is the FS-BSM($\{1, 2, 3\}/12$) in which we impose the restrictions

$$\sigma^2_{\omega,1} = \sigma^2_{\omega,2} = \sigma^2_{\omega,3} = \sigma^2_{\omega,\mathrm{I}} \qquad \text{and} \qquad \sigma^2_{\omega,4} = \sigma^2_{\omega,5} = \sigma^2_{\omega,6} = \sigma^2_{\omega,\mathrm{II}}.$$

In case of the FS-AM in (2.5), we consider the same restrictions but with respect to $c_1, \ldots, c_6$ for $s = 12$. We denote the three classes of FS-AM models by **FS-AM**($\{i\}/s$), **FS-AM**($\{i, j\}/s$) and **FS-AM**($\{i, j, k\}/s$) which correspond to their FS-BSM counterparts. These FS-AM models have a seasonal component that relies on the two coefficients c_I and c_II. For all FS-BSM and FS-AM classes, the models contain five parameters (two coefficients for the trend, one for the noise or irregular and two for the seasonal).

3.2 Testing for deterministic seasonal components in a FS-BSM

In a FS-BSM we may test for a deterministic seasonal component $\gamma_{j,t}$ in (2.10) by means of the hypothesis $H_0 : \sigma^2_{\omega,j} = 0$ for $j = 1, \ldots, 6$ with $s = 12$. For this purpose we consider the Cramer von Mises (CvM) seasonality test of Harvey (2001) and Busetti and Harvey (2003). We effectively need to determine whether the estimated seasonal variance $\sigma^2_{\omega,j}$ is significantly larger than zero. The CvM test statistic can be constructed using standardised one-step ahead prediction errors from the model with parameters that are estimated under the null hypothesis H_0. The resulting test statistic follows a Cramer von Mises distribution with two degrees of freedom for $j = 1, \ldots, 5$ and with one degree of freedom for $j = 6$. In case of the restricted models in FS-BSM($\{i\}/s$), FS-BSM($\{i, j\}/s$) and FS-BSM($\{i, j, k\}/s$), the test hypotheses $H_0 : \sigma^2_{\omega,I} = 0$ and $H_0 : \sigma^2_{\omega,II} = 0$ are effectively joint tests. When we consider a model in FS-BSM($\{i\}/s$), the test for $H_0 : \sigma^2_{\omega,II} = 0$ leads to a CvM test with 9 degrees of freedom for $i = 1, \ldots, 5$ (and 10 for $i = 6$). The critical values of the CvM distribution for the different degrees of freedom are given in Harvey (2001). Finally, in case of the BSM, the test for $H_0 : \sigma^2_\omega = 0$ is CvM distributed with 11 degrees of freedom.

Once it is determined for a FS-BSM that both variances $\sigma^2_{\omega,I}$ and $\sigma^2_{\omega,II}$ are significantly different from zero, we can also test whether the FS-BSM can be reduced to the standard BSM by means of the hypothesis $H_0 : \sigma^2_{\omega,I} = \sigma^2_{\omega,II}$. The likelihood-ratio test statistic based on the maximized likelihoods under the null and the alternative hypotheses can be considered for this purpose. This test is standard and has an asymptotic χ^2 distribution with 1 degree of freedom.

3.3 Distance measure between AM and BSM

The common representation of the AM and BSM (FS and non-FS) is the MA(s+1) representation for $\Delta\Delta_s y_t$. The MA representation for the FS-AM is given by (2.5) and for the FS-BSM by (2.16). We can relate the two models by comparing their corresponding MA coefficients individually. To give an overall measure of closeness between the FS-BSM and FS-AM models with $s = 12$, we define the distance metric as

$$D = \sqrt{\sum_{i=1}^{13}[\theta_i^* - \theta_i]^2}, \qquad (3.1)$$

where θ_i^* and θ_i refer to the ith MA coefficient of the FS-AM and the FS-BSM, respectively, for $i = 1, \ldots, 13$. The 13 coefficients in the MA representation of the FS-BSM are computed numerically from the theoretical autocovariance function using the method of Tunnicliffe-Wilson (1969), see also the discussion in McElroy (2008). When the value of D is small, we regard the two models to be close to each other. Since the AM and BSM are special cases of the FS-AM and FS-BSM, respectively, the distance metric D can also be computed for non-FS models.

3.4 Distribution of the distance measure: a Monte Carlo experiment

By means of a limited Monte Carlo experiment we can obtain an indication of the distribution of the distance metrics (3.1). The simulations are based on 1000 simulated time series with a length of 120 data points. First, we generate time series from a BSM with $\sigma_\eta = 0.02$, $\sigma_\zeta = 0.0001$, $\sigma_\omega = 0.003$ and $\sigma_\epsilon = 0.03$. Each series is used for the estimation of parameters in both the BSM and the Airline model. Second, we generate another 1000 series from a FS-BSM($\{1,4\}/12$) with $\sigma_\eta = 0.02$, $\sigma_\zeta = 0.0001$, $\sigma_{\omega,I} = \sigma_{\omega,1} = \sigma_{\omega,4} = 0.005$, $\sigma_{\omega,II} = \sigma_{\omega,2} = \sigma_{\omega,3} = \sigma_{\omega,5} = \sigma_{\omega,6} = 0.002$ and $\sigma_\epsilon = 0.03$. Each series is used for the estimation of parameters in the FS-BSM($\{1,4\}/12$) and those in the corresponding FS-AM($\{1,4\}/12$). Further we compute the coefficients of the MA(13) representations of the BSM and FS-BSM and we calculate the distance metric D in (3.1) for the respective AM and FS-AM for each time series. These comparisons are based on the model with the estimated parameters.

Figure 2 presents the histogram of the 1000 distances for comparisons BSM versus AM and FA-BSM versus FS-AM. It confirms our expectation that the FS models lie closer to each other than the non-FS models. The median of the distance metric distribution for the FS comparisons lies around 0.1 and for the non-FS comparisons around 0.13. We also learn from the histograms that all distance metrics are computed as non-zero. This implies that the BSM and FS-BSM never display exactly the same fit as the AM and FS-AM, respectively. This confirms the findings reported in Maravall (1985) and Harvey (1989). However, the differences are often small. In many cases the models lead to very similar time series decompositions into seasonal and non-seasonal component estimates.

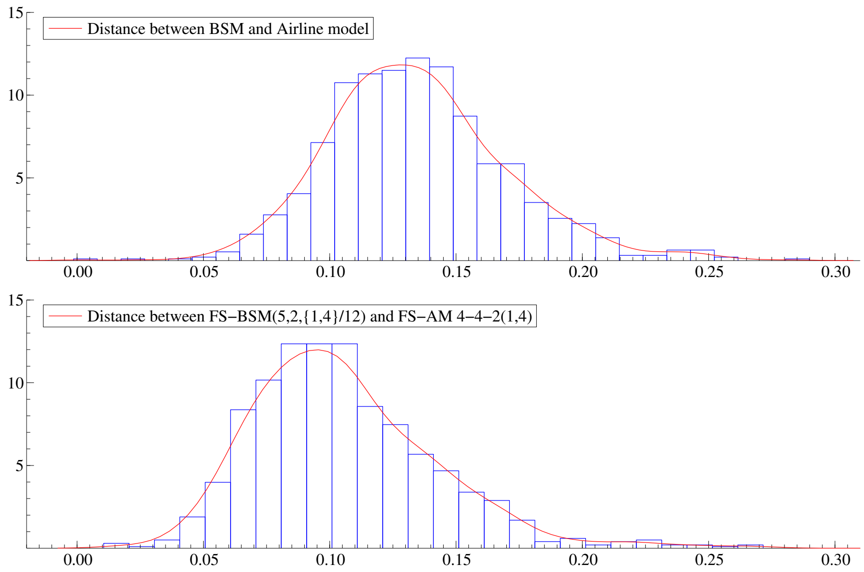

Figure 2: Histogram of distance metric D (defined equation (3.1)) based on simulation study with 1000 replications. *(Top panel)* Distance metric between the BSM and Airline model for a BSM data generating process. *(Bottom panel)* Distance metric between FS-BSM($\{1,4\}/12$) and FS-AM($\{1,4\}/12$) for a FS-BSM data generating process. The dotted line is the estimated density as a smooth function through the histogram.

4 Empirical results for U.S. shipment and foreign trade series

In this section we discuss the results of our empirical study into frequency specific seasonal time series models for 75 U.S. Census Bureau monthly data. Two time series on manufacturing and foreign trade are presented in Figure 1 as examples. Strong seasonal patterns are present in both series. The database of 75 time series consists of 36 monthly Manufacturers' Shipments, Inventories and Orders Survey data from January 1992 until September 2001, and 39 monthly Foreign Trade series (Imports and Exports) from January 1989 through November 2001. The database has been analysed previously with the Airline model (AM) and the FS-AM by Aston, Findley, McElroy, Wills, and Martin (2007). In this paper state space methods are adopted for parameter estimation by maximum likelihood; see Durbin and Koopman (2001). All calculations are carried out by the object-oriented matrix programming environment Ox of Doornik (2006) using the SsfPack library of Koopman, Shephard, and Doornik (2008).

4.1 Illustration 1: U37AVS

We analyse the time series U37AVS (U.S. shipment of household furniture and kitchen cabinets) in more detail. We consider the non-FS BSM and all 31 models in FS-BSM, for each model we add regression variables to account for trading days, Easter effect and outlier effects. The strategies of including regression variables and of taking logs or no-logs are the same as in Aston, Findley, McElroy, Wills, and Martin (2007). When effects are not significant at a significance level of 5%, the regression variables are removed. After this extensive analysis (which consists of estimating the parameters for many different model specifications) we select the model FS-BSM({4}/12) for the time series U37AVS (in logs). The corresponding estimated parameters, together with those of the BSM, are presented in Table 2. The AIC(c) of FS-BSM({4}/12) is smaller than the AIC(c) of the BSM and the average of the estimated seasonal disturbance variances for the FS-BSM appears much larger than the estimated seasonal variance for the BSM.

Table 2 also reports the Cramer von Mises test statistics discussed in section §3.2. We learn that our FS-BSM model for the log(U37AVS) requires a stochastic seasonal component since the null hypothesis of a deterministic seasonal component is rejected at the 5% significance level according to the CvM seasonality test for both the BSM and the FS-BSM({4}/12). The likelihood-ratio test rejects the hypothesis $H_0 : \sigma^2_{\omega,I} = \sigma^2_{\omega,II}$ and therefore we advocate the use of the FS model. The estimated trend variance in FS-BSM({4}/12) is larger than the one in the BSM while the estimated irregular variance is smaller in case of the FS-BSM. Indeed all results indicate that FS-BSM({4}/12) improves the fit for log(U37AVS) compared to the BSM.

The Ljung-Box Q-statistic p-values for the prediction errors are reported in Table 3. It shows that the FS-BSM captures the dynamic features in log(U37AVS) more adequately than the BSM. We do not find significant autocorrelations in the standardized prediction errors from the FS-BSM while the prediction errors from the BSM have clear traces of autocorrelation. We also report the standard asymptotic χ^2 normality test of two degrees of freedom and for both models the normality tests are satisfactory.

Figure 3 presents the seasonally adjusted time series for log(U37AVS). Seasonal adjustment is based on the estimated seasonal component obtained from the Kalman filter and smoother. The FS-BSM produces smoother seasonally adjusted series than the BSM. This finding may be caused by a more adequate specification of the individual seasonal components in the FS-BSM compared to the BSM. Since a smoother seasonally adjusted series is often preferred by the economic policy maker, we regard the FS-BSM as providing a superior model-based seasonal adjustment method for the log(U37AVS) series.

4.2 Illustration 2: X41140

The time series X41140 (export of music instruments) of the U.S. Census database is presented in Figure 1. After the same extensive analysis as for the U37AVS series, we opt for modeling the log(X41140) series and consider the BSM and the 31 FS-BSM models. We find that the maximized likelihood values for the 31 FS-BSM models do not increase in comparison with the maximized likelihood value for the BSM; see the bottom panel of Table 2. Since the likelihood values are very close, the AIC(c) of each FS-BSM is larger than the AIC(c) of the BSM. We therefore prefer the more parsimonious BSM for the log(X41140) time series. Furthermore, the CvM test statistics reveal that the seasonal component is close to being deterministic in both the BSM and FS-BSM cases. The Ljung-Box Q-statistics show no significant residual

Table 2: Estimated parameters of log furniture shipment series, log(U37AVS), using BSM and FS-BSM($\{4\}/12$) (*top panel*) and of log export musical instrument series, log(X41140), using BSM and FS-BSM($\{5,6\}/12$) (*bottom panel*).

	BSM					FS-BSM				
Freq.		Est.par	S.E.	CvM	df		Est.par	S.E.	CvM	df
	σ_η	0.0080	0.0043			σ_η	0.0132	0.0035		
	σ_ζ	0.0011	0.0006			σ_ζ	0.0006	0.0006		
	σ_ϵ	0.0241	0.0031			σ_ϵ	0.0154	0.0045		
$\{i\}/12$	σ_ω	0.0009	0.0005	2.910*	11	$\sigma_{\omega,I}$	0.0008	0.0004	2.513*	9
4/12						$\sigma_{\omega,II}$	0.0054	0.0016	1.032*	2
Log L		196.94					201.72			
AIC		-385.87					-393.44			
AICc		-385.52					-392.90			
	BSM					FS-BSM				
Freq.		Est.par	S.E.	CvM	df		Est.par	S.E.	CvM	df
	σ_η	0.0262	0.0062			σ_η	0.0262	0.0062		
	σ_ζ	5.12e-07	0.0005			σ_ζ	8.64e-08	0.0005		
	σ_ϵ	0.0634	0.0052			σ_ϵ	0.0634	0.0052		
$\{i\}/12$	σ_ω	1.81e-07	0.0009	1.7834	11	$\sigma_{\omega,I}$	3.45e-07	0.0012	1.2939	8
$\{5,6\}/12$						$\sigma_{\omega,II}$	1.96e-07	0.0015	0.4895	3
Log L		145.44					145.44			
AIC		-282.88					-280.88			
AICc		-282.61					-280.48			

Notes: Maximum likelihood estimation of model (2.1)-(2.5) extended with regressors for trading days, outliers, and Easter effect. Regression parameters are not included in AIC and AICc. Standard errors are obtained using a numerical estimate of the Hessian of the log-likelihood and the delta method. Further, * indicates that $H_0 : \sigma_\omega^2 = 0$ (for BSM) or $H_0 : \sigma_{\omega,i}^2 = 0$ (for FS-BSM) is rejected at 5% significance level with $i = I, II$. For the BSM, all seasonal frequencies have the same variance so that the degrees of freedom in the CvM seasonality test is equal to the sum of the individual degrees of freedom. For the grouped FS-BSM, the degrees of freedom for each group equals the sum of individual degrees of freedom within the tested group. Freq = seasonal frequency; Est.par = estimated parameters; S.E. = estimated standard errors; CvM = Cramer von Mises test statistic, df = CvM degrees of freedom.

Table 3: p-values from Ljung-Box Q-Statistics and Normality test of the residuals of log furniture shipment series, log(U37AVS), resulting from BSM and FS-BSM({4}/12).

LB Statistics	BSM	FS-BSM	LB Statistics	BSM	FS-BSM
$Q(5)$	0.0027 ***	-	$Q(15)$	0.0473 *	0.5748
$Q(6)$	0.0011 ***	0.0505	$Q(16)$	0.0453 *	0.5318
$Q(7)$	0.0007 ***	0.1271	$Q(17)$	0.0566	0.4702
$Q(8)$	0.0015 ***	0.0738	$Q(18)$	0.0757	0.5247
$Q(9)$	0.0025 ***	0.1391	$Q(19)$	0.0981	0.5789
$Q(10)$	0.0051 **	0.1934	$Q(20)$	0.1293	0.6445
$Q(11)$	0.0098 **	0.2859	$Q(21)$	0.0572	0.2279
$Q(12)$	0.0147 *	0.3890	$Q(22)$	0.0761	0.2777
$Q(13)$	0.0239 *	0.4947	$Q(23)$	0.0970	0.3336
$Q(14)$	0.0379 *	0.5196	$Q(24)$	0.0985	0.2841
Normality test	0.5494	0.6974			

Here, *,**, and *** symbolise that the p-value is asymptotically significant at $5\%, 1\%$, and 0.5% rejection levels respectively, with H_0 of no autocorrelation in the residuals.

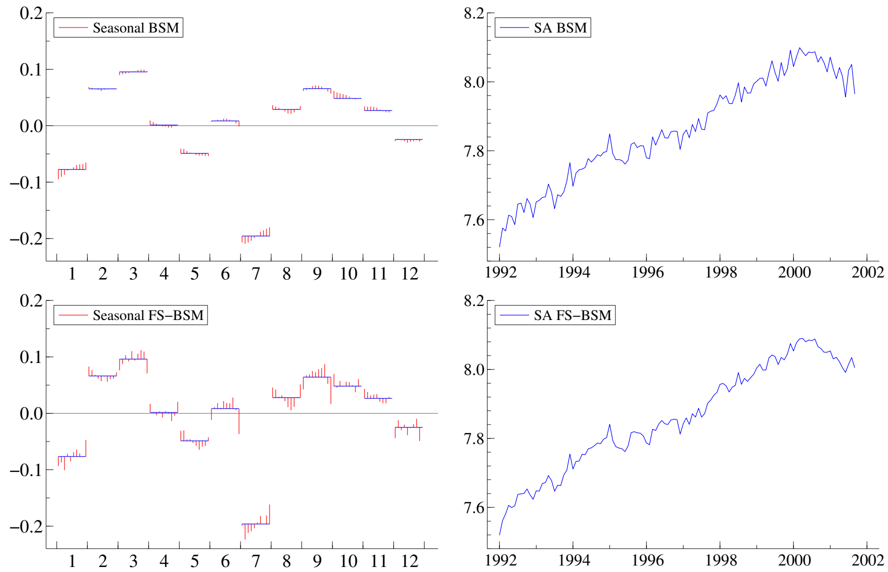

Figure 3: Estimated seasonal component of log furniture shipment series, log(U37AVS), plotted separately for each month and the seasonally adjusted series. BSM's estimates are shown at the top row while the estimates from FS-BSM({4}/12) are presented at the bottom row.

autocorrelation in the one-step ahead prediction errors. The BSM with a nearly fixed seasonal component is therefore a satisfactory model to describe the dynamic features in the log(X41140) series.

4.3 Analysis of all 75 time series

To obtain an overall assessment of the usefulness of the FS-BSM, we analyse all 75 time series from our U.S. Census database in the same way as we have done for the U37AVS and X41140 series. For each time series, we estimate the parameters for the BSM and for all 31 FS-BSM specifications that contain the parameters σ^2_η, σ^2_ζ, $\sigma^2_{\omega,I}$, $\sigma^2_{\omega,II}$ and $\hat{\sigma}^2_\epsilon$. We opt for the model with the lowest corrected Akaike criterion (AICc) function. After completing the analysis for all 75 time series, we have found that in 55 (73%) cases a particular FS-BSM model has the lowest AICc compared to the BSM, while 35 (46%) cases passed the likelihood ratio test with a significance level of 5%. These findings indicate that the FS-BSM leads to improvements in the fit of a time series in many cases.

4.4 Distance measures between FS-BSM and FS-AM

Given the preferred model choice for the 75 time series, we derive the 13 coefficients of the MA representations together with the innovation variance in each case. It enables comparisons with the BSM and AM models. For this purpose we consider the distance metric D defined in (3.1). We note that corresponding (FS-)BSM and (FS-)AM models are specified with the same set of additional regressors (to account for trading days, Easter and outliers).

We first take a closer look at the U37AVS series of section § 4.1. In Table 4 the MA(13) representations of BSM vs AM and FS-BSM({4}/12) vs FS-AM({4}/12) are presented for the log(U37AVS) series. The MA(13) coefficients for the BSM at lag $2, \ldots, 11$ tend to zero while those at lags 1 and $12, 13$ have the same magnitude as those of AM. The corresponding distance metric $D = 0.128$ is also reported in Table 4 and is a typical value in view of the distribution of D as presented in Figure 2. For the FS-BSM({4}/12) and FS-AM({4}/12), the distance metric $D = 0.065$ is clearly smaller.

Next we have computed the distance metric for all series in the database (only for the optimal model specifications in terms of lowest AICc). The 75 distance metrics for both the non-FS and FS models are presented in Figure 4. We note that the smallest distance for each series does not always correspond to the combination of lowest AICc for the FS-BSM and FS-AM. However, the combination of lowest AICc for both types of FS models has produced a distance metric that is smaller than the distance metric resulting from the non-FS models in all cases. Therefore we can conclude that the FS-BSM and FS-AM specifications are closer to each other than the BSM and AM.

Figure 5 indicates whether the FS-AM and FS-BSM are more similar to each other compared

Table 4: Restricted reduced form MA(13) coefficients of the (FS-)AM and (FS-)BSM applied to log furniture shipment series, log(U37AVS).

	Airline model	BSM	FS-AM 4-5-1(4)	FS-BSM({4}/12)
i	$\hat{\theta}_i^*$	$\hat{\theta}_i$	$\hat{\theta}_i^*$	$\hat{\theta}_i$
1	0.609	0.651	0.607	0.634
2	0	-0.045	0.051	0.021
3		-0.010	-0.348	-0.340
4		-0.043	0.269	0.260
5	.	-0.010	0.046	0.046
6	.	-0.040	-0.319	-0.345
7	.	-0.008	0.247	0.269
8		-0.037	0.043	0.018
9		-0.006	-0.293	-0.282
10		-0.032	0.227	0.230
11	0	-0.002	0.039	0.059
12	0.667	0.697	0.440	0.431
13	-0.406	-0.478	-0.222	-0.211
$\log(\hat{\sigma})$	-3.341	-3.355	-3.394	-3.398
Log L	196.162	196.937	201.569	201.722
D	0.128		0.065	

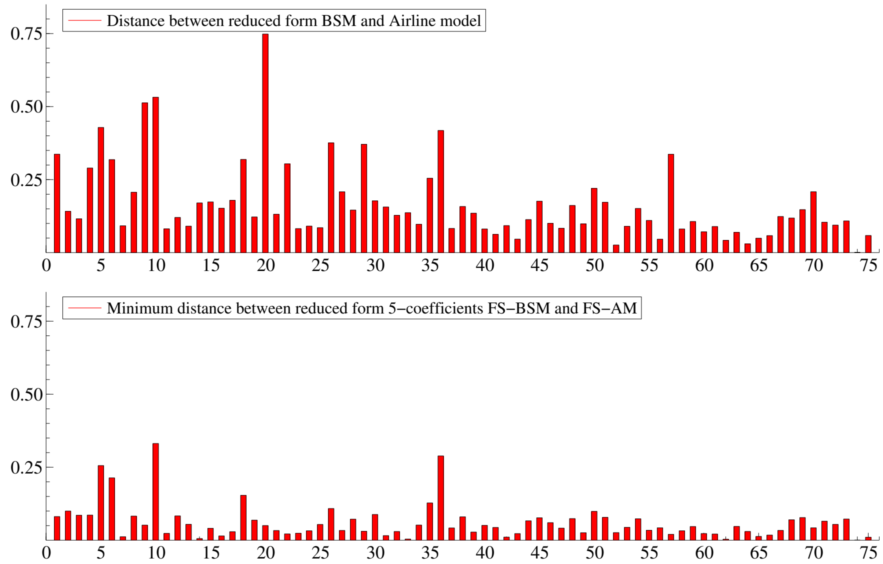

Figure 4: *(Top panel)* Distance metric D in (3.11) between estimated BSM and Airline models applied to 75 manufacturing and export monthly time series from the U.S. Census Bureau database. *(Bottom panel)* The smallest distance between 31 types of grouped FS-BSMs and 4-coefficients FS-AMs. For each series we find a smaller distance metric between the FS models compared to the distance between the non-FS models.

to their non-FS counterparts. For this purpose we introduce the relative distance as given by

$$RD = \frac{D_{\text{nonFS}} - D_{\text{FS}}}{D_{\text{nonFS}}}, \tag{4.1}$$

where D_{nonFS} and D_{FS} are the distance (3.1) between BSM vs AM and FS-BSM vs FS-AM, respectively. In Figure 5 we present the average values of RD for all 75 time series in our database. A positive (negative) RD indicates by how much the similarity between the FS-AM and FS-BSM increases (decreases) compared to the similarity between their non-FS counterparts. The x-axis of Figure 5 represents a specific FS-BSM as discussed in section § 3.1 while the y-axis represents the corresponding FS-AM. The model class FS-BSM($\{i\}/s$) are indexed from 1 to 6 on the x-axis, FS-BSM($\{i,j\}/s$) are indexed from 7 to 21 (15 models) and FS-BSM($\{i,j,k\}/s$) are indexed as 22-31 (10 models). The indexing is similar for the FS-AM models. In total we consider 31 different FS-BSM and FS-AM models. For all series in the database, we obtained positive RD values which indicate that the distance for the FS models is always smaller than the distance for the non-FS models. The most positive RD values are obtained on the leading diagonal which shows that similar restrictions applied to FS-BSM and FS-AM lead to more similar models compared to non-FS models.

All results in this subsection are for FS models with two seasonal coefficients and a total of five coefficients. We have also considered the estimation of models with six different variances for the six different seasonal frequencies in the FS-BSM. When we compare the resulting distances from their FS-AM counterparts, the empirical models are even more similar. However, we do not advocate such over-parametrized specifications as the estimation results can be spurious with many seasonal variances frequently estimated as zero in the FS-BSM and many seasonal unit roots in the moving average part of the FS-AM.

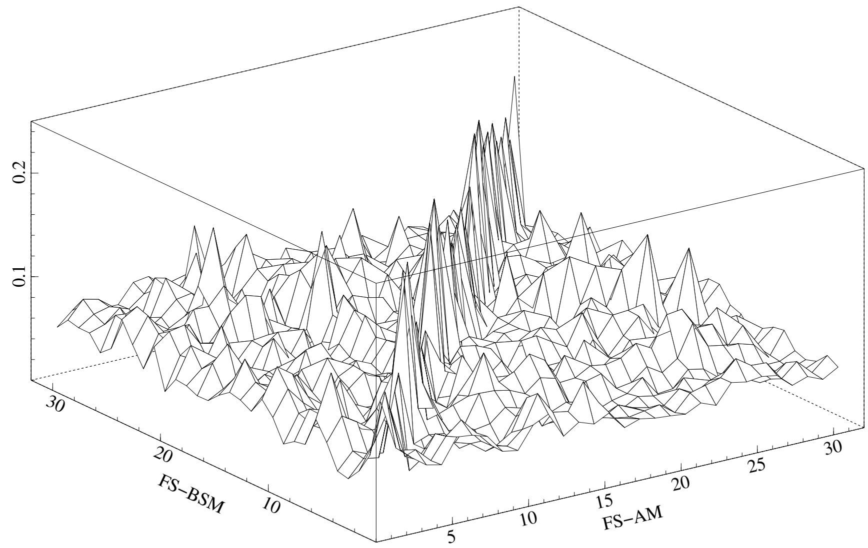

Figure 5: Average of relative distance (RD) of non-FS models versus FS models, RD is defined in (4.1) and the average is for the RD values obtained for the 75 time series in our database. The RD values are computed for the 31 FS-BSM and FS-AM specifications (with five parameters).

5 Conclusion

We have investigated frequency specific (FS) time series models for seasonal time series. In particular, we focussed on the FS versions of the basic structural time series model and compared it to the FS versions of the well-known Airline model. The dynamic properties of the FS-BSM are investigated in terms of its MA representation that also allows its comparison with the FS-AM. The relations between the parameters of the FS-BSM and FS-AM are highly non-linear. We therefore rely on numerical comparisons between the two classes of models. For this purpose we propose a distance measure based on the MA coefficients of both models. In general, BSM models can be quite different from AM models when they are fitted to the same time series. However, we show in our simulation and empirical studies that FS versions of these models are more similar to each other. Furthermore, we have shown that, in many cases, the FS version of a model offers a significantly better description of the dynamic properties of a time series.

References

Aston, J. A. D., D. F. Findley, T. S. McElroy, K. C. Wills, and D. E. K. Martin (2007). New ARIMA models for seasonal time series and their application to seasonal adjustment and forecasting. Research Report Series RRS2007/14, U.S. Census Bureau.

Aston, J. A. D., D. F. Findley, K. C. Wills, and D. E. K. Martin (2004). Generalizations of the Box-Jenkins' Airline Model with frequency-specific seasonal coefficients and a generalization of Akaike's MAIC. In *Proceedings of the 2004 NBER/NSF Time Series Conference*, Washington D.C. and Census Bureau, http://www.census.gov/ts/papers/findleynber2004.pdf. NBER.

Bell, W. R. (1993). Empirical comparisons of seasonal ARIMA and ARIMA component (structural) time series models. In *Proceedings of the Business and Economic Statistics Section*, pp. 226–231. Alexandria, VA: American Statistical Association.

Bell, W. R. (2004). On RegComponent time series models and their applications. In A. Harvey, S. J. Koopman, and N. Shephard (Eds.), *State Space and Unobserved Component Models*, pp. 248–283. Cambridge, U.K.: Cambridge University Press.

Box, G. E. P. and G. M. Jenkins (1970). *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco, CA, USA.

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

Busetti, F. and A. Harvey (2003). Seasonality tests. *Journal of Business & Economic Statistics 21*, 420–436.

Doornik, J. A. (2006). *Ox: An Object-Oriented Matrix Language*. London, U. K.: Timberlake Consultants Press.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford, UK: Oxford University Press.

Gómez, V. and A. Maravall (1996). Programs TRAMO and SEATS: Instructions for the user. Technical Report Working paper 9628, Servicio de Estudios, Banco de España, Madrid.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman Filter*. Cambridge, UK: Cambridge University Press.

Harvey, A. C. (2001). Testing in unobserved components models. *Journal of Forecasting 20*, 1–19.

Hillmer, S. C. and G. C. Tiao (1982). An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association 77*, 63–70.

Koopman, S. J., N. Shephard, and J. A. Doornik (2008). *Statistical Algorithms for Models in State Space Form - SsfPack 3.0*. Timberlake Consultants Ltd, London, UK.

Maravall, A. (1985). On structural time series models and the characterization of components. *Journal of Business & Economic Statistics 3*, 350–355.

McElroy, T. (2008). Matrix formulas for nonstationary ARIMA signal extraction. *Working paper*.

Proietti, T. (2000). Comparing seasonal components for structural time series models. *International Journal of Forecasting 16*(2), 247–260.

Proietti, T. (2004). Seasonal specific structural time series. *Studies in Nonlinear Dynamics & Econometrics 8*(2), Article 16, http://www.bepress.com/snde/vol8/iss2/art16.

Tunnicliffe-Wilson, G. (1969). Factorization of the covariance generating function of a pure moving average process. *SIAM Journal of Numerical Analysis 6*, 1–7.