RESEARCH REPORT SERIES
(Statistics #2007-14)

New ARIMA Models for Seasonal Time Series and Their Application to Seasonal Adjustment and Forecasting

John A.D. Aston[1,2]
David F. Findley[1]
Tucker S. McElroy[1]
Kellie C. Wills[1,3]
Donald E.K. Martin[1,4]

U.S. Census Bureau[1]
Washington, D.C. 20233

Institute of Statistical Science[2]
Academia Sinica, Taiwan

Insightful Corporation[3]
Seattle, Washington

North Carolina State University[4]

Report Issued: October 18, 2007

NEW ARIMA MODELS FOR SEASONAL TIME SERIES AND THEIR APPLICATION TO SEASONAL ADJUSTMENT AND FORECASTING

JOHN A. D. ASTON[1,2], DAVID F. FINDLEY[1,*], TUCKER S. MCELROY[1], KELLIE C. WILLS[1,3], AND DONALD E. K. MARTIN[1,4]

Abstract. Focusing on the widely-used Box-Jenkins "airline" model, we show how the class of seasonal ARIMA models with a seasonal moving average factor can be parsimoniously generalized to model time series with heteroskedastic seasonal frequency components. Our frequency-specific models decompose this factor by associating one moving average coefficient with a proper subset of the seasonal frequencies 1, 2, 3, 4, 5 and 6 cycles per year and a second coefficient with the complementary subset. A generalization of Akaike's AIC is presented to determine these subsets. Properties of seasonal adjustment filters and adjustments obtained from the new models are examined as are forecasts. *Keywords:* Airline model; Frequency-Specific Model; Generalized Airline Model; Model selection; AIC; $\mathcal{F}$-AIC

Date: October 18, 2007.

[1]United States Census Bureau, Washington DC, USA.

[2]Institute of Statistical Science, Academia Sinica, Taiwan.

[3]Insightful Corporation, Seattle, USA.

[4]North Carolina State University.

[*]Address for Correspondence:

David F Findley.

Statistical Research Division.

United States Census Bureau.

Washington, DC 20233-9100.

USA.

Tel: (301) 763-4983.

Fax: (301) 457-2299.

Email: david.f.findley@census.gov.

1. Introduction

Box and Jenkins (1976) developed a two-coefficient time series model, now known as the airline model, which is by far the most widely used ARIMA model for monthly and quarterly macroeconomic time series. For example, Fischer and Planas (2000) deem it adequate for 50% of 13,232 Eurostat time series. However, its extensive use may be due in part to the lack of parsimonious alternatives suitable for these kinds of series.

Seasonal adjustment is a major concern for statistical agencies and these models are often used in the production of the adjusted series published as indicators of national and supranational economies. Thus model choice for the seasonal series will often favor properties useful for seasonal adjustment. These include forecasting ability for seasonal adjustment methods based on nonparametric seasonal adjustment filters like the X-11 filters of X-12-ARIMA, see Findley, Monsell, Bell, Otto, and Chen (1998), and the capability to produce sufficiently flexible seasonal adjustment filters in the case of the ARIMA model-based (AMB) seasonal adjustment procedure of Hillmer and Tiao (1982) and Burman (1980). The AMB procedure is implemented in the SEATS software (Gómez and Maravall 1997) and also in a soon-to-be-released extension of the X-12-ARIMA program named X-13A-S that incorporates SEATS; see Monsell, Aston, and Koopman (2003) and Findley (2005b).

In this paper, generalizations of the airline model are proposed that include somewhat common structural features of time series that are not addressed by the airline model or existing alternatives. These models are parsimonious, a necessity given the short series lengths that are typical of the macroeconomic time series that are seasonally adjusted. The models generalize the seasonal moving average factor of the airline model to enable a differentiated treatment of seasonal frequency components in place of the homogeneous treatment of these components by the airline model. We call these models frequency-specific models (FSMs). The same kind of generalization can be applied to the seasonal moving average factors of any seasonal ARIMA model with such factors, see Section 7, but until then, we use this term exclusively for frequency-specific generalizations of the airline model. It will be shown that these models produce smoother seasonal adjustments and have better model diagnostics.

An example motivating the need for an alternative to the airline model is given in the next section. In Section 3, the various types of FSMs are defined. Next, in Section 4, model selection is discussed. For the situation in which entire families of the new models are considered, a limited extension of Akaike's Minimum AIC criterion is developed that takes into account the multiple comparison nature of the model selection undertaken. In Section 5, this criterion is applied to the motivating example and to the other members of an extensive set of U.S. Census Bureau series previously modeled with the airline model. The performance of the new models relative to the airline model is investigated for the series for which the new models are chosen by the new criterion. Subsection 5.5 presents results and issues for noninvertible models, which occur somewhat often and require different treatment. Section 6 considers some estimation properties. The final section provides summary remarks, conclusions and comments regarding extensions. An Appendix discusses estimation of the FSMs and available software with this capability.

2. Motivating Example - U34EVS

U34EVS, *Shipments of Defense Communications Equipment* (January 1992 through September 2001) is a series from the Census Bureau's monthly Manufacturers' Shipments, Inventories and Orders Survey. If the airline model (1) below (with $s = 12$) were chosen for this series, which is likely in the absence of alternatives such as those described in this paper, this would suggest that the seasonal component is adequately modeled by a single parameter. However, the spectrum plot of differenced and logged U34EVS in Figure 1 reveals that the peak around the frequency associated with quarterly movements, 4 cycles per year (4/12 of a

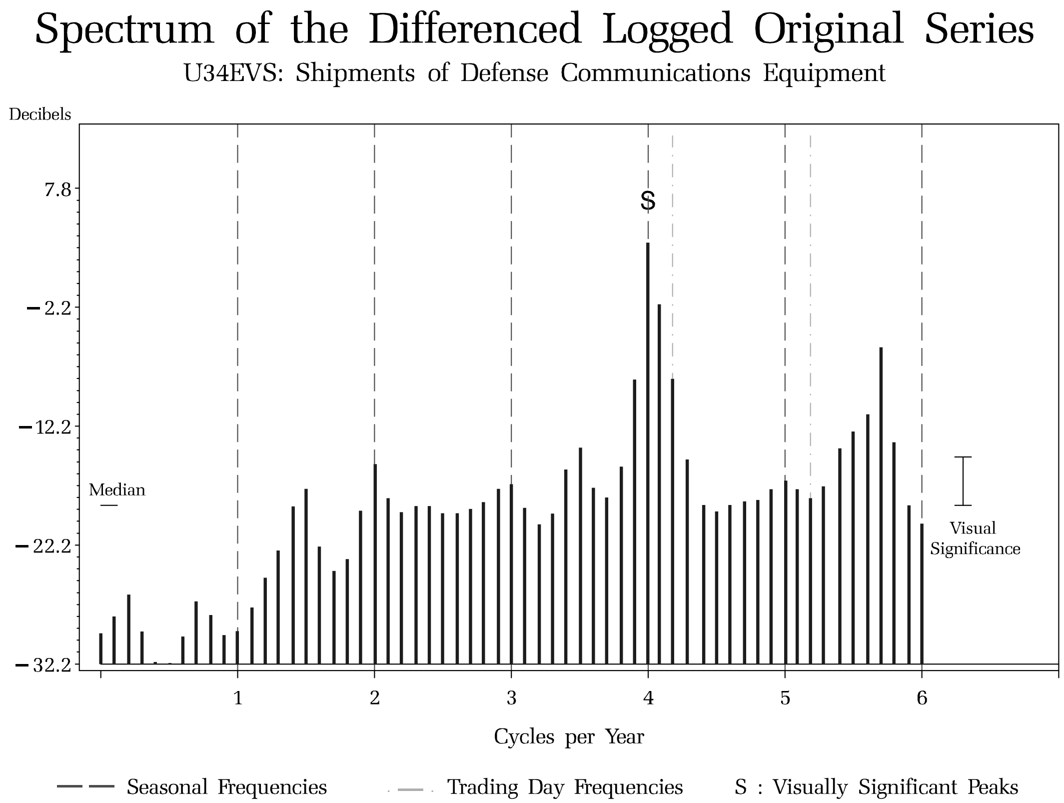


Figure 1. Spectrum in Decibels of First-Differenced Logs of Manufacturers' Shipments of Defense Communications Equipment (U34EVS). There is a dominant peak at the quarterly frequency 4 cycles per year. The S above this seasonal peak indicates that it is classified as visually significant.

cycle per month), is much greater in both magnitude and breadth than the peaks at the other seasonal frequencies. This indicates much greater strength and variability of the quarterly component of the series than of the components associated with other seasonal frequencies. The "visual significance" criterion referred to in the figure is discussed in Soukup and Findley (1999).

Series whose spectrum plots show a dominant seasonal peak (not always as distinct as in Figure 1) are rather common. Other kinds of heterogeneity among the components of variation associated with seasonal frequencies also occur. In the next section, we present some parsimonious generalizations of the airline model that can model seasonal heterogeneity.

The need for such models is especially clear in the context of AMB seasonal adjustment. In Subsection 5.2, it will be seen that the frequency-specific models defined below often have better goodness-of-fit diagnostics than the airline model together with a smoother AMB seasonal adjustment, a very desirable pair of properties. A smoother adjustment is more amenable to interpretation, with enhanced reliability of the interpretation when the adjustment is produced from a well-fitting model (Findley and Martin 2006).

3. The Airline Model and Some Generalizations

The Box-Jenkins airline model for a seasonal time series Z_t, observed at regular intervals $s \geq 2$ times per year, has the form

$$(1-B)(1-B^s)Z_t = (1-\theta B)(1-\Theta B^s)\epsilon_t, \tag{1}$$

where ϵ_t is a zero-mean i.i.d. process with finite variance.

When $\Theta \geq 0$, as is typical for macroeconomic time series, the airline model can be written

$$(1-B)^2(\sum_{j=0}^{s-1} B^j)Z_t = (1-\theta B)(1-\Theta^{\frac{1}{s}}B)(\sum_{j=0}^{s-1}\Theta^{\frac{j}{s}}B^j)\epsilon_t. \tag{2}$$

Findley, Martin, and Wills (2002) substituted a general MA(2) polynomial for $(1-\theta B)(1-\Theta^{\frac{1}{s}}B)$ in (2), yielding the model

$$(1-B)^2(\sum_{j=0}^{s-1} B^j)Z_t = (1-aB-bB^2)(\sum_{j=0}^{s-1} c^j B^j)\epsilon_t. \tag{3}$$

In this model, the seasonal sum polynomial has a third coefficient c distinct from the coefficients associated with the other factors in the model. This model was introduced to decouple the seasonal and nonseasonal dependencies on Θ of (2), where Θ appears in both the MA(2) and the seasonal sum factors.

In the present paper, we investigate various generalizations of (2) and (3) that we call *frequency-specific* models. In these models, the final moving average factor, $\sum_{j=0}^{s-1} c^j B^j$ in (3), is decomposed into several factors with different coefficients. Restricting attention to monthly data for definiteness, i.e. $s = 12$, the model (3) can be generalized by factoring it into terms of frequencies of 1, 2, 3, 4, 5 and 6 cycles per year to obtain a *general frequency-specific model*,

$$(1-B)^2(\sum_{j=0}^{11} B^j)Z_t = (1-aB-bB^2)\left[(1+c_6B)\prod_{j=1}^{5}\left(1-2c_j cos(\frac{2\pi j}{12})B+c_j^2B^2\right)\right]\epsilon_t. \quad (4)$$

If the six c_j's are unconstrained, this model has a different seasonal coefficient for each seasonal frequency and a total of eight coefficients, so it is not a parsimonious model. Moreover, with most of the series to which we fit this model, an estimated value $c_j = 1$ occurred for at least one of these coefficients and was usually a spurious indication of noninvertibility that is problematic for model selection and seasonal adjustment; see Subsections 5.2 and 5.5. Instead of the model (4), we start with the most parsimonious generalizations of (2) of this form, with only two distinct c_j's. That is, the seasonal frequencies are divided into two groups, with all frequencies in a group having the same coefficient and with the factor $1-\Theta^{\frac{1}{12}}B$ replaced by $1-c_1B$. This expands the number of coefficients in the model from two to three. More generally, if the seasonal factor of (3) is generalized in this way, a four-coefficient model results.

Three types of three-coefficient models are considered. The models will be labeled 3-($\lfloor s/2\rfloor - k$)-k, $k = 1, \ldots, \lfloor s/4\rfloor$, for general $s \geq 4$, s even. Again, looking specifically at $s = 12$, in the first model type, designated the 3-5-1 type, five of the frequency factors in brackets in (4) have the same coefficient c_1 and the sixth has its own coefficient c_2. There are six such models, an example being

$$(1-B)^2(\sum_{j=0}^{11} B^j)Z_t = (1-aB)(1-c_1B)\left[(1+c_2B)\prod_{j=1}^{5}\left(1-2c_1 cos(\frac{2\pi j}{12})B+c_1^2B^2\right)\right]\epsilon_t, \quad (5)$$

which we call the 3-5-1(6) model, because the 6/12 cycle per month or 6 cycles per year frequency has the c_2 coefficient.

In the second type, designated 3-4-2, four of the frequency factors in brackets in (4) have the same coefficient c_1 and the remaining two have a different coefficient c_2. There are fifteen such models, an example being the 3-4-2(1,6) model

$$(1-B)^2(\sum_{j=0}^{11} B^j)Z_t = (1-aB)(1-c_1B)\times$$
$$\left[(1+c_2B)(1-2c_2cos(\frac{2\pi}{12})B+c_2^2B^2)\prod_{j=2}^{5}\left(1-2c_1cos(\frac{2\pi j}{12})B+c_1^2B^2\right)\right]\epsilon_t. \tag{6}$$

In the third type of model, designated 3-3-3, three of the frequency factors in brackets in (4) have the same coefficient c_1 and the other three have a different coefficient c_2. There are twenty such models. For example, the 3-3-3(1,2,6) model is

$$(1-B)^2(\sum_{j=0}^{11} B^j)Z_t = (1-aB)(1-c_1B)\times$$
$$\left[(1+c_2B)\prod_{j=1}^{2}\left(1-2c_2cos(\frac{2\pi j}{12})B+c_2^2B^2\right)\prod_{j=3}^{5}\left(1-2c_1cos(\frac{2\pi j}{12})B+c_1^2B^2\right)\right]\epsilon_t. \tag{7}$$

The 3-3-3(3,4,5) looks very similar,

$$(1-B)^2(\sum_{j=0}^{11} B^j)Z_t = (1-aB)(1-c_1B)\times$$
$$\left[(1+c_1B)\prod_{j=1}^{2}\left(1-2c_1cos(\frac{2\pi j}{12})B+c_1^2B^2\right)\prod_{j=3}^{5}\left(1-2c_2cos(\frac{2\pi j}{12})B+c_2^2B^2\right)\right]\epsilon_t, \tag{8}$$

but the coefficient of $1-c_1B$ is associated with the other coefficient group.

We also consider four-coefficient FSMs in which the polynomial $(1-aB)(1-c_1B)$ is replaced by a general nonseasonal MA(2) polynomial $(1-aB-bB^2)$ as in (3). These are defined analogously, with the four-coefficient model types denoted 4-5-1, 4-4-2 and 4-3-3 models when $s=12$. For example, the 4-5-1(6) model is

$$(1-B)^2(\sum_{j=0}^{11} B^j)Z_t = (1-aB-bB^2)\left[(1+c_2B)\prod_{j=1}^{5}\left(1-2c_1cos(\frac{2\pi j}{12})B+c_1^2B^2\right)\right]\epsilon_t. \quad (9)$$

There are six 4-5-1 models and fifteen 4-4-2 models but only ten 4-3-3 models, because the same models now result if c_1 and c_2 are interchanged, e.g. the 4-3-3(1,2,6) and 4-3-3(3,4,5) models are identical.

These FSMs cannot be estimated with standard ARIMA modeling software. For this article, estimation was performed in the object-oriented matrix programming environment Ox (Doornik 2001), using the state space functions in the SSFPack library (Koopman, Shephard, and Doornik 1999). Some details are given in the Appendix.

Before presenting our evaluations of the new models relative to the airline model for seasonal adjustment and forecasting of U.S. Census Bureau series, we consider two model selection criteria for deciding when one of the new models can be preferred over the airline model.

4. AIC-Based Approaches to Model Comparisons with the New Models

4.1. **Comparison of Two Models.** For the airline model, let $\widehat{\vartheta}^A$, dim ϑ^A, and $L(\widehat{\vartheta}^A)$ denote the estimated parameter vector, its dimension, and the associated maximum log-likelihood value respectively. Let $\widehat{\vartheta}^F$, dim ϑ^F, and $L(\widehat{\vartheta}^F)$ denote the corresponding quantities for an FSM. (As usual, log-likelihoods of Gaussian form are used.)

Recall that Akaike's AIC for a given model is defined to be minus two times the model's maximum log-likelihood plus twice the number of its estimated parameters. Akaike's Minimum AIC criterion (MAIC) states that, between two models, the one with the smaller AIC should be preferred; see Akaike (1973) and Findley (1999) for example. This property is determined by the sign of the difference of the AIC values, so we consider the AIC difference

$$AIC(\widehat{\vartheta}^A) - AIC(\widehat{\vartheta}^F) = -2\{\ln L(\widehat{\vartheta}^A) - \ln L(\widehat{\vartheta}^F)\} - 2(\dim \vartheta^F - \dim \vartheta^A). \tag{10}$$

We start with the series U34EVS of Section 2. The spectrum in Figure 1 suggests that a 3-5-1(4) model might be a more appropriate model than the airline model. The airline model has the AIC value of -135.85 while the 3-5-1(4) model has a smaller AIC value, -142.33, so the difference (10) is 6.48. This is highly significant if an asymptotic chi-square distribution with one degree of freedom is assumed for the log-likelihood ratio component of (10); see below. Thus the 3-5-1(4) model is strongly favored by a conventional test as well as by MAIC.

The estimated coefficients for the 3-5-1(4) model are $a = 0.660$, $c_1 = 0.987\,(= \sqrt[12]{0.855})$ and $c_2 = 0.893\,(= \sqrt[12]{0.257})$. The value of a is to be compared to the estimate $\theta = 0.715$ of the airline model. The twelfth powers 0.855 and 0.257 of c_1 and c_2, respectively, are to be compared to $\Theta = 0.780$. Thus the 3-5-1(4) model has a much smaller parameter for the 4/12 frequency than the airline model. In Section 5, this will be seen to result in a smoother seasonal adjustment.

The 4-5-1(4) model could be compared to the 3-5-1(4) model to see if it yields further improvement. However, for this series, it has the estimate $c_1 = 1$, i.e. a seasonal unit root in the moving average polynomial, which presents both practical and theoretical problems that will be discussed in Subsection 5.5.

4.2. **A Modification of Akaike's Minimum AIC Procedure for Multiple Same-Dimension Comparisons.** The airline model with $\Theta \geq 0$ is a special case of each type of FSM, and when such an airline model is correct,

$$-2\{\ln L(\widehat{\vartheta}^A) - \ln L(\widehat{\vartheta}^F)\} \sim \chi^2_{\dim \vartheta^F - \dim \vartheta^A} \tag{11}$$

holds asymptotically under standard assumptions, including the requirement that the airline model be invertible, i.e. without unit roots in the MA factors; see Taniguchi and Kakizawa (2000, p. 61). Under (11), the asymptotic probability that the FSM will have a smaller AIC and thus be incorrectly preferred by Akaike's MAIC criterion is, from (10),

$$P(AIC(\widehat{\vartheta}^A) - AIC(\widehat{\vartheta}^F) > 0) = P(\chi^2_{\dim\, \vartheta^F - \dim\, \vartheta^A} > 2(\dim\, \vartheta^F - \dim\, \vartheta^A)). \qquad (12)$$

Thus, from the r.h.s. of (12), the asymptotic probability of incorrectly rejecting the airline model in favor of an FSM is $p^{(3)} \equiv P(\chi^2_1 > 2) = 0.157$ for a three-coefficient model and $p^{(4)} \equiv P(\chi^2_2 > 4) = 0.135$ for a four-coefficient model.

These type I error probabilities apply when a single FSM is compared to the airline model as above. When a full family of FSMs is compared, e.g. all six 3-5-1s or all fifteen 4-4-2s, a modified procedure is called for that accounts for the multiplicity of comparisons. We now present a generalization of AIC for the situation in which all models of a family $\mathcal{F}$ of FSMs have the same number of coefficients and are being compared to another model. The idea is that, when this other model is the airline model, the asymptotic type I error probability should approximate the value of the r.h.s. of (12) for a single model $F \in \mathcal{F}$. To this end, we define

$$\mathcal{F}\text{-}AIC = \min_{F \in \mathcal{F}} AIC(\widehat{\vartheta}^F) + \Delta^{\mathcal{F}}, \qquad (13)$$

where, when the airline model is correct, $\Delta^{\mathcal{F}} \geq 0$ has the approximation property

$$P(AIC(\widehat{\vartheta}^A) - \mathcal{F}\text{-}AIC > 0) = P(AIC(\widehat{\vartheta}^A) - \min_{F \in \mathcal{F}} AIC(\widehat{\vartheta}^F) > \Delta^{\mathcal{F}}) \doteq p, \qquad (14)$$

where p is the value of the r.h.s. of (12) for a single $F \in \mathcal{F}$. We define $\Delta^{\mathcal{F}} = 0$ when $\mathcal{F}$ has only one model.

Such upper percentiles $\Delta^{\mathcal{F}}$ can be estimated most simply from the empirical probability density of $AIC(\widehat{\vartheta}^A) - \min_{F \in \mathcal{F}} AIC(\hat{\vartheta}^F)$ obtained when the models in $\mathcal{F}$ are fitted to long simulated Gaussian time series generated by an airline model. In this paper, the $\Delta^{\mathcal{F}}$ values

TABLE 1. Thresholds $\Delta^{\mathcal{F}}$ for the Model Families $\mathcal{F}$

$\mathcal{F}$	3-5-1	3-4-2	3-3-3	All	4-5-1	4-4-2	4-3-3	All
$\lvert\mathcal{F}\rvert$	6	15	20	41	6	15	10	31
$\Delta^{\mathcal{F}}$	2.8	3.8	3.9	4.6	2.8	3.7	3.1	4.1

of Table 1 are used. These values were obtained from simulated series for all pairs θ and Θ chosen over a grid of pairs with $0.1 \leq \theta, \Theta \leq 0.9$. (Among the empirical densities of the individual pairs, the upper percentiles of interest varied by less than 5%.) It would not be expected that the limits are dependent on the value of θ, Θ asymptotically, analogously to AIC. Since the distributions used to obtain $\Delta^{\mathcal{F}}$ values are required to approximate those from asymptotic theory, the empirical distributions were determined by using series that are very long ($N = 2001$) in comparison to the usual few hundred observations that are typically available in real macroeconomic data.

The graphs of the empirical densities that yielded Table 1's values for the 3-coefficient models are shown in Figure 2. The smooth lines are the MLE fits from MATLAB® of gamma densities. Gamma densities were suggested by both the shape of the distributions, and as chi-square distributions are special cases of gammas. The gamma scale parameters, which range from 1.34 to 1.60, all have values smaller than the value 2 of a chi-square variate. The gamma means increase from 2.99 to 4.13 as the size $|\mathcal{F}|$ of $\mathcal{F}$ increases from 6 to 41.

We use $\mathcal{F}$-AIC of (13) with the $\Delta^{\mathcal{F}}$ values of Table 1 to compare the minimum AIC models of different model classes $\mathcal{F}$, taking as the preferred model the one whose $\mathcal{F}$-AIC is smallest. (This includes comparisons with a singly chosen ARIMA model, whose $\mathcal{F}$-AIC is AIC.) In analogy with Akaike's MAIC criterion, we call this criterion the $\mathcal{F}$-MAIC criterion.

It can be seen from further analyses (not shown) of the simulations used to calculate the $\Delta^{\mathcal{F}}$ values that the frequency with which a particular FSM is incorrectly chosen over the airline model by AIC is roughly equal across the model family's seasonal frequency groups associated

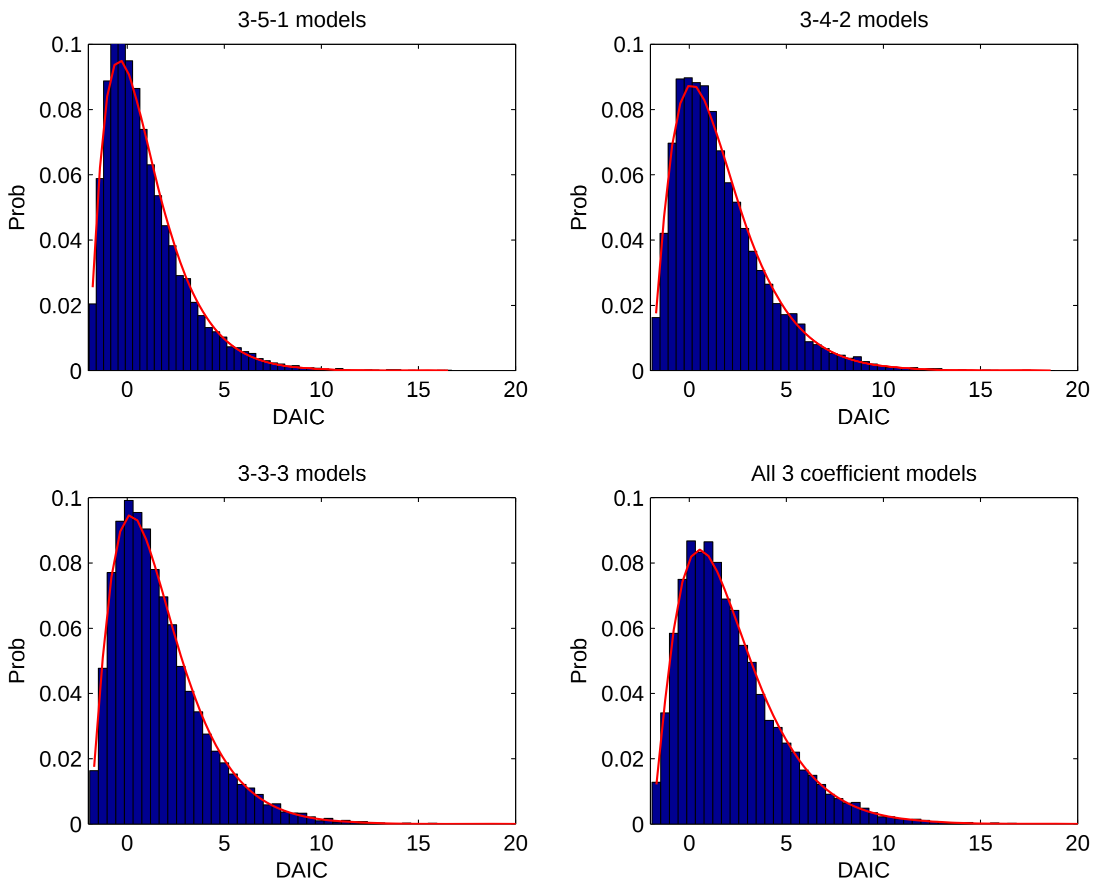


FIGURE 2. Empirical Densities of $DAIC(A, \mathcal{F}) \equiv AIC(\hat{\vartheta}^A) - \min_{F \in \mathcal{F}} AIC(\hat{\vartheta}^F)$ for Each of Model Type and for All Types Together for the Three-Coefficient Models. The densities are generated from 15,000 samples at each point of a grid of parameter values of the airline model with series lengths $N = 2001$ to approximate asymptotic results. Their upper 15.7% quantiles define the first four $\Delta^{\mathcal{F}}$ values of Table 1. The smooth lines indicate the MLE fits of gamma densities to the empirical densities.

with c_2. For example, each of the six 5-1 models has probability approximately 1/6 of being incorrectly chosen.

Finally, returning to MAIC, while the type I error probabilities $p^{(3)}$ and $p^{(4)}$ values from (12) are large relative to subjectively chosen significance levels of tests like 0.05, it should be stressed that these are more fundamental quantities than such empirical choices because of the asymptotic unbiasedness property of $AIC(\widehat{\vartheta}^A) - AIC(\widehat{\vartheta}^F)$ as an estimator of the mean accuracy difference of the two models in the Kullback-Leibler sense; see Akaike (1973), Findley (1999) and Findley and Wei (2002). Given the nonlinearity of the minimization over $F \in \mathcal{F}$, one cannot expect the l.h.s of (13) to have an analogous property. In addition, if other model selection choices such as BIC were preferred, the multiple comparisons nature would still need to be accounted for in a similar way.

5. Empirical Properties of Example Series

We fitted the airline model and each of the six model sets defined by the three- and four-coefficient generalizations to 72 U.S. Census Bureau series (including in all models GLS estimation of trading day and outlier effects when such effects were found when using the airline model). These consist of Value of Shipments series from the monthly Survey of Manufacturers' Shipments, Inventories and Orders, beginning in January 1992 and ending in September 2001 (length $N = 117$), and Foreign Trade series (Imports and Exports), from January 1989 through November 2001 (length $N = 155$). They comprise all Value of Shipments series and Foreign Trade series for which an invertible airline model had originally been chosen over other standard ARIMA models when performing standard seasonal ARIMA model selection of the form used in programs such as TRAMO/SEATS.

The following empirical evaluation of the series is broken down into two different model results, one where the chosen model was invertible and the other where the model was not invertible. Invertibility issues complicate ARIMA modeling in general, and these new models are no exception. Thus it is preferable to deal with these two cases separately. Here,

Table 2. Numbers of Invertible FSMs Preferred over the Airline Model by Type

Model Types	3-5-1	3-4-2	3-3-3	4-5-1	4-4-2	4-3-3	Series
AIC Preferred	36	136	178	25	74	44	51
$\mathcal{F}$-AIC Preferred	12	10	8	8	16	11	18
Global $\mathcal{F}$-MAIC	9	3	2	2	1	1	18

invertibility is considered at a model level, such that if the maximum likelihood estimates from a model are found to be noninvertible for a series, then the model is disregarded when considering invertible models for that series, rather than actually constraining the parameters to be invertible.

5.1. **Preferred Invertible Models for 72 U.S. Census Bureau Series.** Here, only those models with final maximum likelihood estimates that resulted in invertible models are considered. Table 2 gives the breakdowns of MAIC and $\mathcal{F}$-MAIC choices by model type of the invertible FSMs that were preferred over the airline model. For $\mathcal{F}$-MAIC , the $\Delta^{\mathcal{F}}$ values of Table 1 were used.

The first row of Table 2 covers 51 series and the second 18. The individual entries reveal that for some series, more than one FSM is preferred over the airline model. Thus the use of $\mathcal{F}$-AIC in place of AIC reduces the percentage of the 72 series for which an FSM is preferred from 71% to 25%. However, this is still a substantial percentage given that airline models were initially found by a modeler to be adequate and preferable to other standard seasonal ARIMA models for these series. (If the $\mathcal{F}$-MAIC criterion is used with the Δ^{all} value, then 11 series, 15%, have FSMs preferred over the airline model.) Among the 18 series with a preferred FSM, the numbers of $\mathcal{F}$-MAIC models in each family $\mathcal{F}$ are given in the last line of Table 2. Three-coefficient models are by far the most often preferred, being the $\mathcal{F}$-MAIC model for 14 of the 18 series, and therefore for 19% of the 72 series. (We also considered the model (3) for the 72 series, but in every case in which its AIC was less than the airline model's, an FSM had a still smaller $\mathcal{F}$-AIC.)

The spectrum provides an interpretation of the $\mathcal{F}$-MAIC choice for some series, but frequently does not unambiguously indicate the distinctive nature of the frequency or pair or triple of frequencies associated with c_2 as we show below. Alternative spectra could be used, for example that of the twice differenced series or of the differenced and seasonally differenced series, but we found these spectra less informative.

5.2. **Seasonal Adjustment Properties of Two Invertible $\mathcal{F}$-MAIC Three-Coefficient Models.** For data Z_t, $1 \leq t \leq N$, regarded as having an additive seasonal decomposition, most simply $Z_t = S_t + A_t$ with seasonal component S_t and nonseasonal component A_t, the AMB procedure (Hillmer and Tiao 1982) is usually able to derive from the seasonal ARIMA model for Z_t an ARIMA model for S_t and an ARIMA model for A_t. With these models, Gaussian conditional mean calculations are used to obtain linear estimates $\hat{A}_t$ of A_t,

$$\hat{A}_t = \sum_{j=t-N}^{t-1} a_{t,j} Z_{t-j},\ 1 \leq t \leq N,$$

which are the seasonally adjusted values when the Z_t are the data. The seasonally adjusted values are $\exp(\hat{A}_t)$, $1 \leq t \leq N$ when the Z_t are the logs of the data.

For the series U34EVS introduced in Section 2, the estimated coefficients of the 3-5-1(4) model and its AIC comparison with the airline model were given in Subsection 4.1. Figure 3 shows that the 3-5-1(4) model's AMB seasonal adjustment is smoother than that of the airline model.

However, when the full three-coefficient model set was considered using the $\mathcal{F}$-MAIC criterion, there was, unexpectedly, a better model than the 3-5-1(4) model, according to $\mathcal{F}$-MAIC. The 3-4-2(4,6) model, whose c_2 singles out both the quarterly (4/12) and the bimonthly (6/12) frequencies, was preferred. Its AIC value is -145.01. The p-values of the goodness-of-fit Q

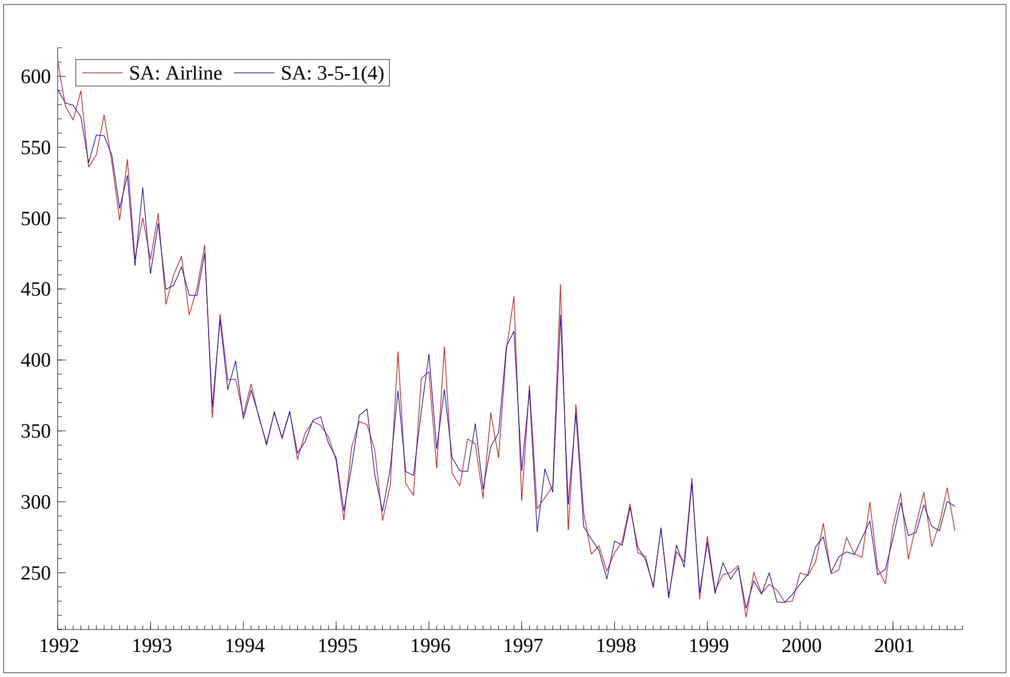


Figure 3. AMB Seasonal Adjustments of U34EVS from the Airline and 3-5-1(4) Models. The latter model's adjustment is usually smoother.

statistics of Ljung and Box (1978) of the models considered for U34EVS are given in Table 3. Those of the 3-4-2(4,6) all exceed by far the customary minimum acceptable value of 0.05, whereas those of the 3-5-1(4) model indicate a poor fit to the data and those of the airline model a still worse fit.

For this series the coefficient estimates of the 3-4-2(4,6) model are $a = 0.604$, $c_1 = 0.990\,(= \sqrt[12]{0.886})$, and $c_2 = 0.870\,(= \sqrt[12]{0.188})$. The value of c_2's twelfth power, 0.188, is smaller than the corresponding value 0.257 for the 3-5-1(4) model, and is very small compared to $\Theta = 0.780$ for the airline model. Figure 4 shows that its AMB seasonal adjustment of U34EVS is smoother than that of the 3-5-1(4) and much smoother than the AMB seasonal adjustment of the airline model shown in Figure 3.

TABLE 3. Q Statistics' p-Values of Three Models for U34EVS

Lag	Airline	3-5-1(4)	3-4-2(4,6)	Lag	Airline	3-5-1(4)	3-4-2(4,6)
3	0.02	–	–	14	0.00	0.02	0.33
4	0.07	0.02	0.57	15	0.00	0.02	0.23
5	0.04	0.02	0.73	16	0.00	0.02	0.22
6	0.00	0.00	0.48	17	0.00	0.01	0.08
7	0.00	0.01	0.62	18	0.00	0.01	0.09
8	0.00	0.00	0.60	19	0.00	0.01	0.09
9	0.00	0.01	0.39	20	0.00	0.01	0.12
10	0.00	0.01	0.41	21	0.00	0.02	0.14
11	0.00	0.01	0.41	22	0.00	0.01	0.11
12	0.00	0.01	0.27	23	0.00	0.01	0.11
13	0.00	0.02	0.33	24	0.00	0.02	0.14

This property, and others of AMB seasonal adjustments, can be inferred or understood from the squared gain functions of the adjustment filters,

$$\left| \sum_{j=t-N}^{t-1} a_{t,j} \exp\left(i2\pi j\lambda\right) \right|^2, \ 1 \leq t \leq N, \tag{15}$$

in which amplitudes less than one are associated with suppression of frequency components; see Findley and Martin (2006) for many examples and technical details, also concerning the fact that smaller values of Θ are associated with greater suppression of variance components around seasonal frequencies, which leads to smoother seasonal adjustments, almost independently of θ.

Figure 5 shows the squared gains of the concurrent seasonal adjustment filter of both the 3-5-1(4) and airline models. These are the filters that provide the value $\hat{A}_N$ for their model's seasonal adjustment of the most recent month, $\exp(\hat{A}_N)$. The 3-5-1(4) model's much broader troughs at the quarterly frequency compared to the airline model's indicate much more suppression of these dominant spectrum components, which is coupled with slightly less suppression than the airline model's around the other seasonal frequencies, which are associated with weaker spectral components. The net effect is a smoother AMB seasonal adjustment from the 3-5-1(4) model. The rapid oscillations in the squared gains around the seasonal frequencies other than 4 cycles per year are essentially due to the high values of Θ and c_1 giving rise to filter coefficients that decay little over the relatively short length of the series; see Findley and Martin (2006).

Figure 6 compares the squared gains of 3-5-1(4) and 3-4-2(4,6) filters. The latter model's smaller value of c_2 results in the squared gains of its filters having wider troughs at the 4 cycles per year frequency than the 3-5-1(4) and airline model filters. Consequently, the 3-4-2(4,6) model's seasonal adjustment is the smoothest, this being visible in the Figure 4. Elsewhere, the squared gains of the filters are similar.

The far superior Ljung-Box Q statistics of the 3-4-2(4,6) make this model much more preferable than the 3-5-1(4) model, and its smoother adjustment makes it still more attractive. The spectrum plot in Figure 1 does not suggest its pairing of the sixth seasonal frequency with the fourth. Its discovery came from the availability of a plausible model selection criterion for automatic searches over several model families. (More often than not, spectrum plots do not unambiguously suggest the frequency groupings of $\mathcal{F}$-MAIC 4-2 or 3-3 models.)

Figures 7–9 present graphs analogous to those above for the Airline and $\mathcal{F}$-MAIC 3-5-1(4) models of the series U36HVS, Value of Shipments of Non-Defense Aircraft Engines and Parts, from the U.S. Census Bureau's monthly Manufacturers' Shipments, Inventories and Orders Survey. The airline model coefficients for this series are $\theta = 0.579$ and $\Theta = 0.501$ and the AIC value for the airline model is 1450.7. The $\mathcal{F}$-MAIC 3-5-1(4) model has an AIC value of 1447.8, with coefficient values $a = 0.534$ and $c_1 = 0.959$ $(= \sqrt[12]{0.605})$. By contrast, $c_2 = 0.895$

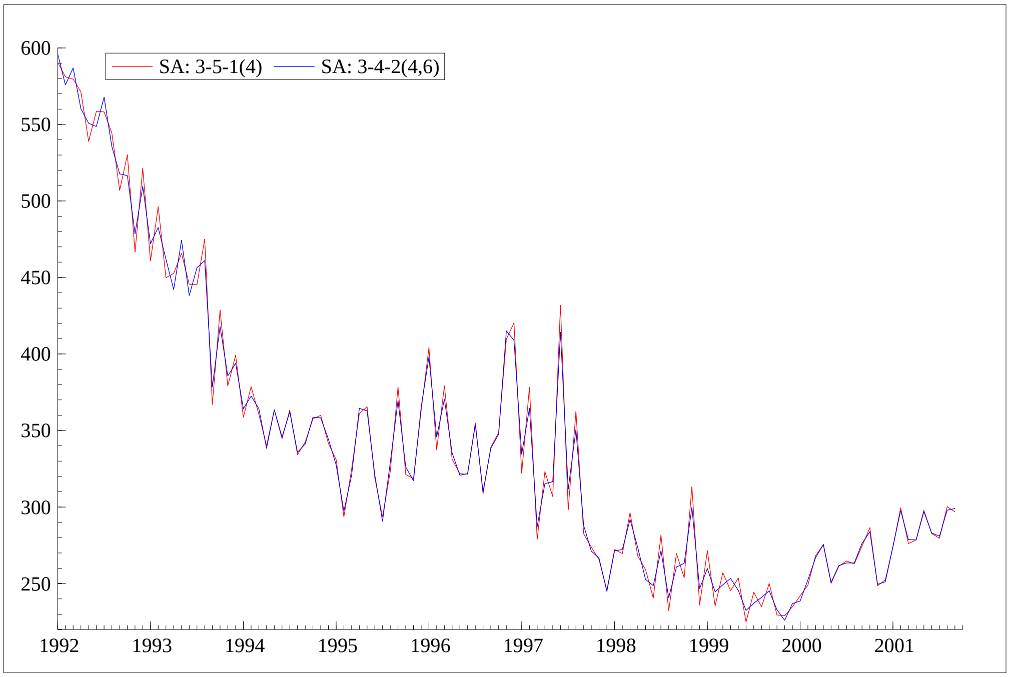


FIGURE 4. AMB Seasonal Adjustments of U34EVS from the 3-5-1(4) Model and the $\mathcal{F}$-MAIC 3-4-2(4,6) Model. The latter model's adjustment is smoother.

$(= \sqrt[12]{0.264})$. This coefficient is associated with the quarterly frequency, which has the largest seasonal peak in the spectrum of the modeled series; see Figure 8. There are smaller peaks at three of c_1's seasonal frequencies and no peaks at two, so the peak at the quarterly frequency is not as distinctive as in the spectrum of U34EVS (Figure 1). However, the 3-5-1(4) model fits the data much better. For the airline model, the p-values of the Ljung-Box Qs (not shown) exceed 0.05 only at lags 2, 3 and 10, whereas those of the 3-5-1(4) exceed 0.05 at all lags except 24, where the p-value is 0.023 compared with 0.0007 for the airline model. The 3-5-1(4) model's AMB seasonal adjustment is usually smoother than the airline model's, see Figure 7.

5.3. **Goodness-of-Fit and Seasonal Adjustment Properties for the 18 Series with Invertible F-MAIC Models.** Overall, among the 18 series for which an invertible FSM

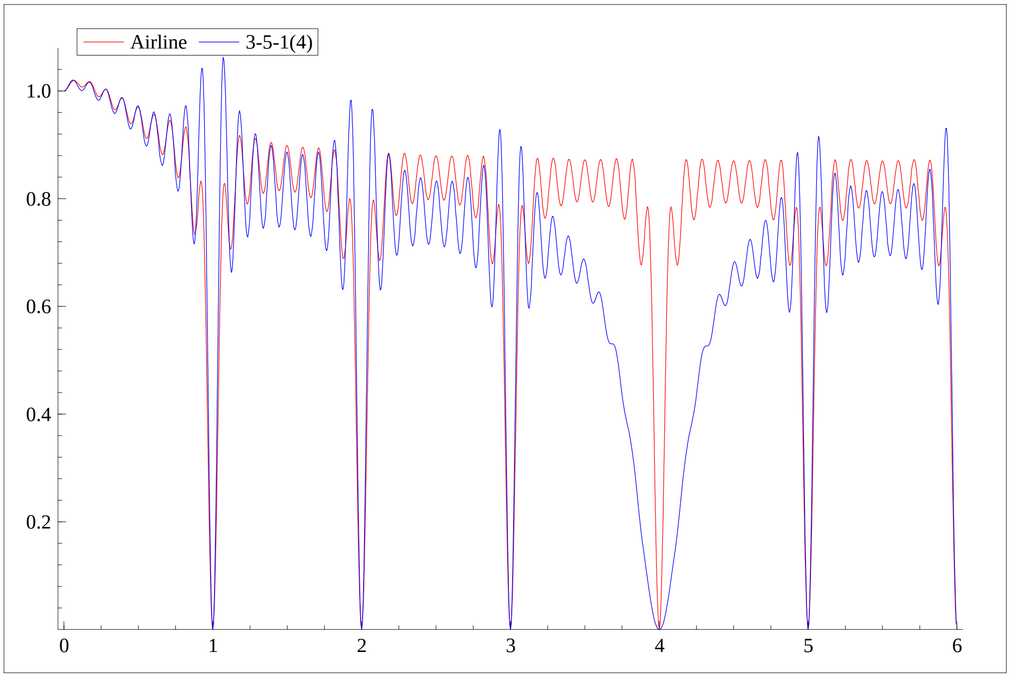


FIGURE 5. Squared Gains of Airline Model and FSM Concurrent AMB Filters for U34EVS. The airline model has $\theta = 0.715$ and $\Theta = 0.780$. The 3-5-1(4) model has $a = 0.660$, $c_1 = 0.987$ $(= \sqrt[12]{0.855})$ and $c_2 = 0.893 (= \sqrt[12]{0.257})$. The last results in greater suppression around 4 cycles per year.

was preferred over the airline model by $\mathcal{F}$-MAIC, the Ljung-Box Q statistics of the $\mathcal{F}$-MAIC FSMs had p-values greater than 0.05 at all lags up through 24 for 16 of the 18 series, and one to two lags with a p-value below 0.05 for the other two series. By contrast, with the airline model, there were only four series for which these statistics had p-values greater than 0.05 at all lags, and there were 10 series for which p-values were below 0.05 at 11 or more lags, indicating a very poor fit to the data.

The FSM's seasonal adjustment was smoother than the airline model's for nine series, smoother more often than not for three, less often than not for three, and smoother as often as not for three. As Findley and Martin (2006) discuss, smoother seasonal adjustments are often associated with greater phase-delays (time-shifts) in the presentation of data characteristics

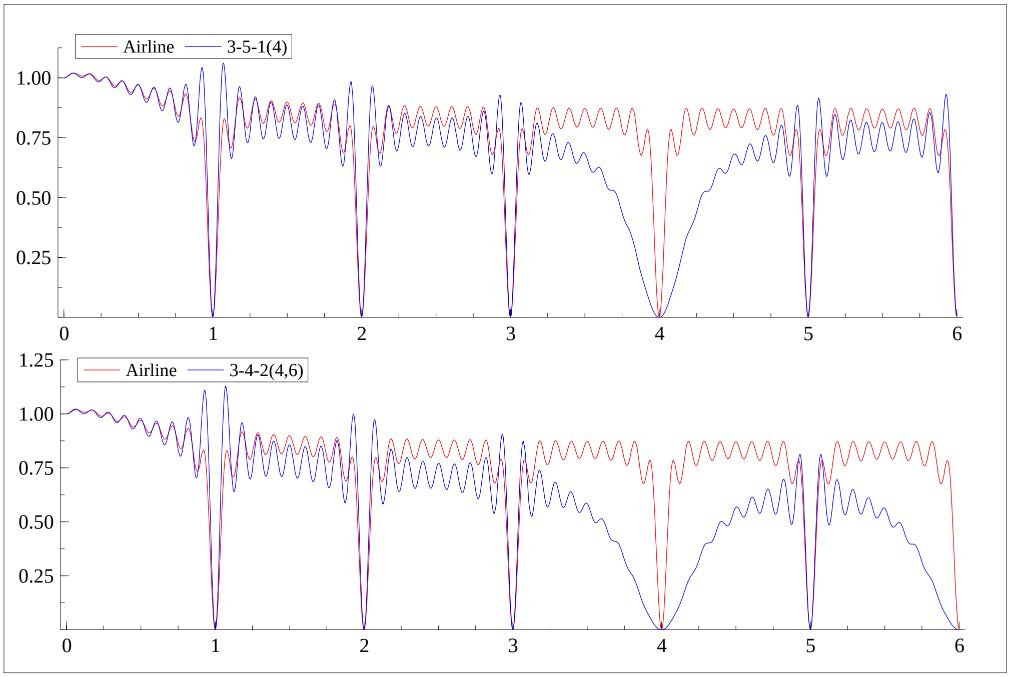


FIGURE 6. Plots of the Squared Gains of the Concurrent AMB Filters for U34EVS from the 3-5-1(4) and the 3-4-2(4,6) Models. The airline model's squared gain is plotted with both for reference. For the 3-4-2(4,6) model, $a = 0.604$, $c_1 = 0.990 (= \sqrt[12]{0.886})$, and $c_2 = 0.870 (= \sqrt[12]{0.188})$. This model's smaller value of c_2 results in greater suppression around 4 as well as around 6 cycles per year.

associated with the business cycle movements near the recent end of the series. However, we do not present concurrent filter phase-delay diagnostic graphs for these series, or for those discussed below in Subsection 5.5, because the phase-delays indicated were of no practical importance: over business cycle frequencies (periods three to 10 years), the phase-delays of concurrent FSM filters, while often larger than those of the airline model, were rarely larger than one month and, in the distinctively worst case, still slightly less than two months, a month more than with the airline model.

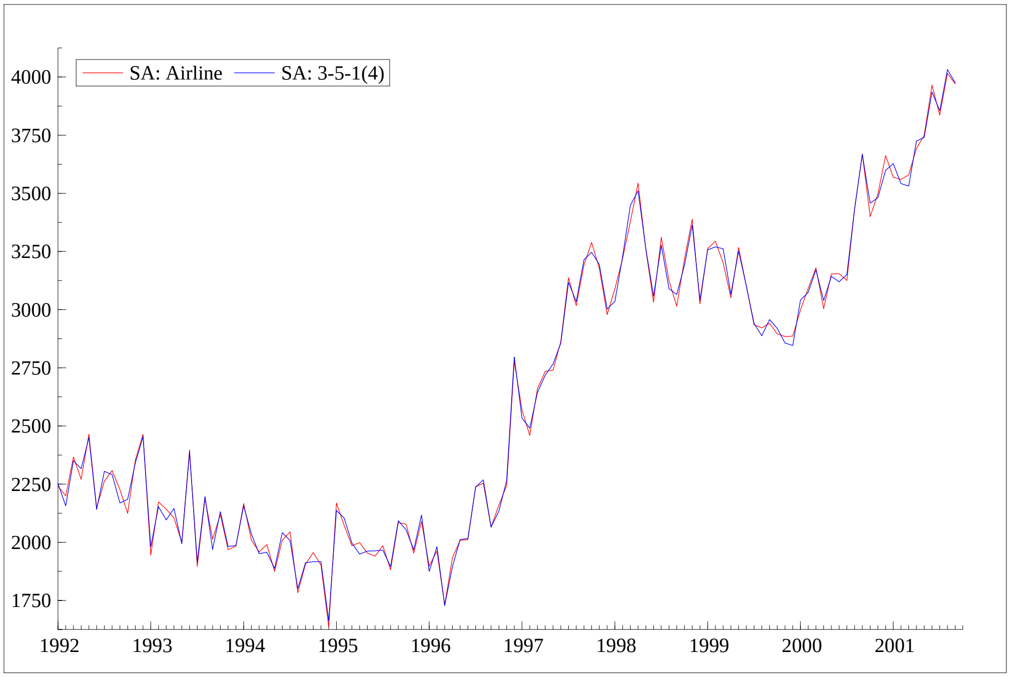


FIGURE 7. AMB Seasonal Adjustments of Manufacturers' Shipments of Non-Defense Aircraft Engines and Parts (U36HVS) from the Airline and $\mathcal{F}$-MAIC 3-5-1(4) Models. The latter model's adjustment is usually smoother.

5.4. **Out-of-Sample Forecasting Performance.** To obtain information about a model's h-step-ahead forecasting performance, some interval of observations at the end of the series can be regarded as future data to be forecasted from a model estimated from preceding data. The resulting out-of-sample forecasts can be compared to the actual series values (or, for series values identified as outliers, to the outlier-adjusted values). The ending date of the data span on which model coefficients are estimated, and from which forecasts are produced, can be increased one observation at a time to obtain a sequence of h-step-ahead forecasts and the sequence of associated forecast errors. Let $e_{A,h,t+h}$ denote the error of an airline model's forecast of Z_{t+h} from a model fitted to $Z_s, 1 \leq s \leq t$ and let $e_{F,h,t+h}$ denote the corresponding error of a specified FSM. Given such errors for $t_0 + h \leq t + h \leq T$, the accumulated differences

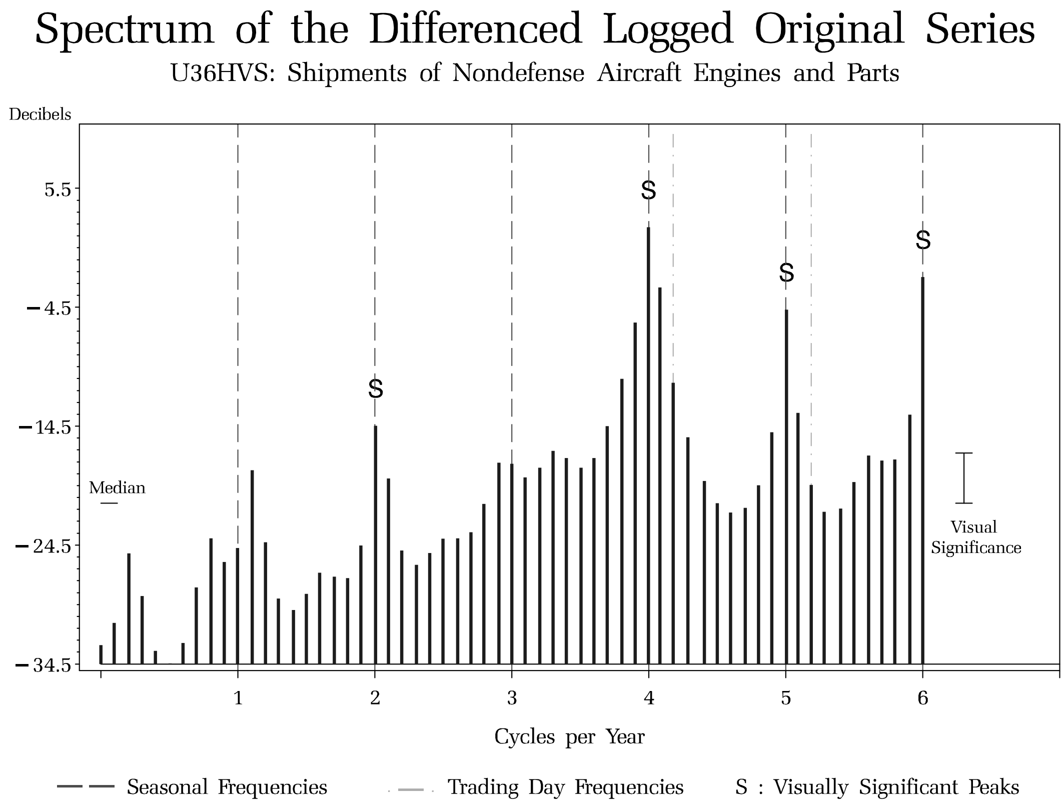


FIGURE 8. Spectrum in Decibels of First-Differenced U36HVS. There is a dominant peak at the quarterly frequency 4 cycles per year.

of squared forecast errors

$$\sum_{s=t_0+h}^{t+h} \left\{ e^2_{F,h,s} - e^2_{A,h,s} \right\}, \quad t_0 + h \leq t + h \leq T, \tag{16}$$

can be graphed as a function of $t+h$. Persistent downward movement in the graph indicates persistently better forecasting by the FSM, whereas persistent upward movement indicates persistently worse forecasting by this model.

Two examples of these diagnostic graphs are given in Figure 10, which shows plots of (16) for $h = 1, 12$. In Figure 10(a), the $\mathcal{F}$-MAIC model is the 3-4-2(4,6) model and the final years of the series U34EVS are being forecasted. In Figure 10(b), the $\mathcal{F}$-MAIC model is the 3-5-1(4) model and the series is U36HVS. The generally descending graphs indicate that the

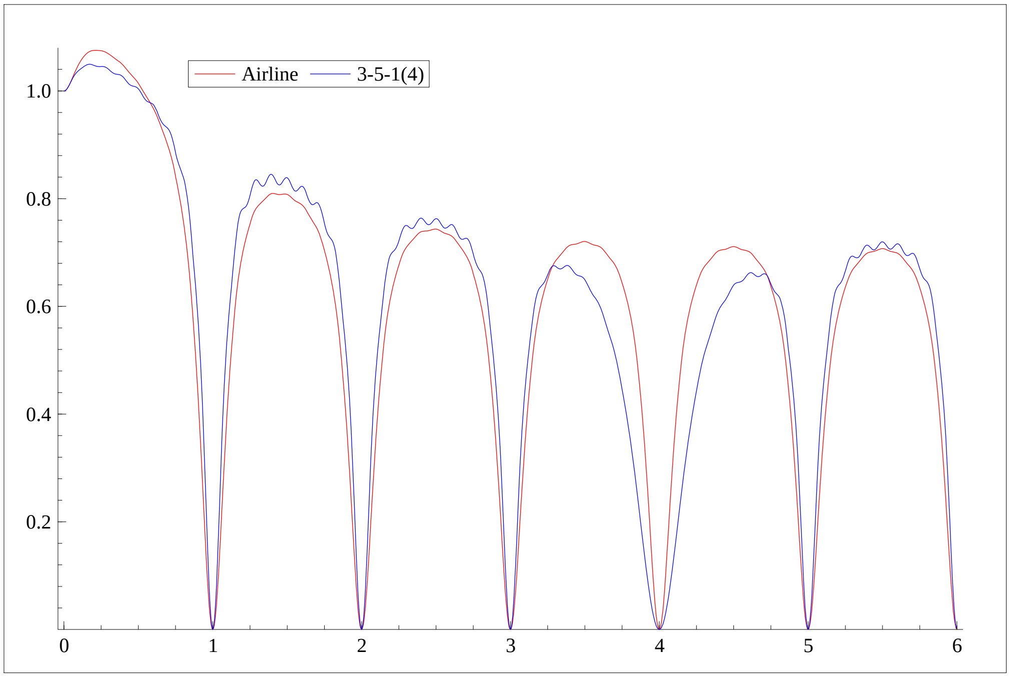
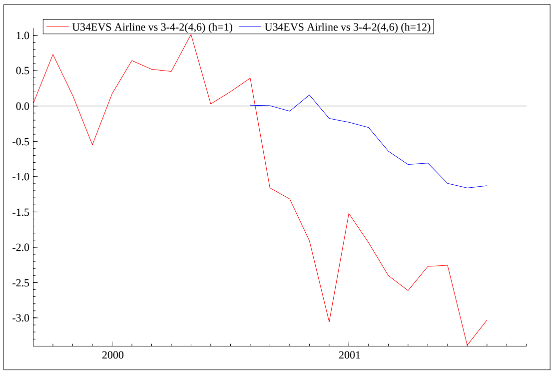


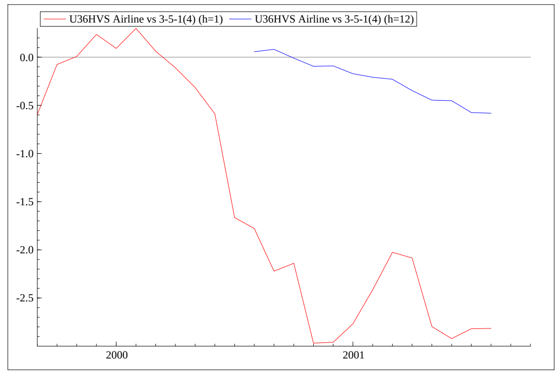
FIGURE 9. Squared Gains of the Concurrent Filters for U36HVS. The airline model has $\theta = 0.579$ and $\Theta = 0.501$. The 3-5-1(4) model has $a = 0.534$ and $c_1 = 0.959$ $(= \sqrt[12]{0.605})$, and $c_2 = 0.895$ $(= \sqrt[12]{0.264})$. The last results in greater suppression around 4 cycles per year.

one-step-ahead forecast performance of both the 3-4-2(4,6) model for U34EVS and 3-5-1(4) model for U36HVS are basically better than that of the airline model. The graphs of 12-step-ahead forecast performance show the almost constantly superior performance of the FSMs. More details and examples of these out-of-sample forecast error diagnostic plots are given in Findley et al. (1998). A supporting theoretical result is provided in Findley (2005a).

Out-of-sample forecasting is computationally expensive, given the need to re-estimate the model for each forecast. Thus it is impractical to determine the out-of-sample forecasting performance for every FSM. Thus we examined this diagnostic only for the 18 FSMs preferred by $\mathcal{F}$-MAIC over the airline model. The results are summarized in Table 4, which shows that $\mathcal{F}$-MAIC preference preponderantly yielded out-of-sample forecasting performance as good as



(a) U34EVS



(b) U36HVS

FIGURE 10. Graphs of the Diagnostic (16) for U34EVS and U36HVS. These show the superior one- and 12-step-ahead out-of-sample forecasting performance of the $\mathcal{F}$-MAIC FSMs compared with the airline model.

TABLE 4. Out-of-Sample Forecast Preferences for the 18 Series

Preferred Model	1-step forecasting	12-step forecasting
$\mathcal{F}$-MAIC FSM	7	7
Airline	1	2
None	10	9

or better than the airline model's. This is further evidence of the practical value of these new models.

5.5. **Preferred Noninvertible Models for 72 U.S. Census Bureau Series.** In addition to the 18 series discussed above, there are four others for which only noninvertible FSMs have $\mathcal{F}$-AIC values smaller than the AIC of the airline model. The $\mathcal{F}$-MAIC models for three of these series are 3-3-3 models with $c_1 = 1$ or $c_2 = 1$, and the $\mathcal{F}$-MAIC model for the fourth is a 4-3-3 model whose MA(2) polynomial has a root equal to one. These series were excluded from Table 2 because invertibility is required for (11) and therefore also for the derivation of the $\Delta^{\mathcal{F}}$-values of Table 1 and of $\mathcal{F}$-AIC. For the same reason, noninvertible $\mathcal{F}$-AIC preferred and noninvertible $\mathcal{F}$-MAIC preferred models were not counted in Table 2. In fact, the initial $\mathcal{F}$-MAIC model for seven of the 18 series was noninvertible. One was a three-coefficient model with $c_2 = 1$. The other six were four-coefficient models, of which three had an MA(2) root of one. But each of these seven series also had one or more invertible $\mathcal{F}$-AIC preferred FSMs, and the last line of Table 2 categorizes the $\mathcal{F}$-MAIC models among such invertible models.

In place of the use of $\mathcal{F}$-AIC with noninvertible models, which is unjustified if noninvertibility is correct for the series, out-of-sample forecast comparisons can be used to justify preference for a noninvertible model. Before presenting such comparisons, we point out two sources of what can be called *spurious noninvertibility*. The first is badly fitting models. For example, with U34EVS, the four estimated 3-5-1 models different from the 3-5-1(4) and 3-5-1(6) models are noninvertible, with $c_2 = 1$. The spectrum estimate of Figure 1 indicates that these other models are poor choices for the series. The second source is the phenomenon that data from an invertible moving average model give rise to maximum Gaussian likelihood parameter estimates which, with positive probability, specify a noninvertible model (with high probability if the moving average polynomial has zeroes close enough to 1 in magnitude); see Chapter 8 of Tanaka (1996) for an extensive treatment of this phenomenon. An indication that noninvertibility is spurious can be that re-estimation of the model from a moderately truncated span of the time series yields an invertible model.

Now we summarize the out-of-sample forecast comparison results for the $\mathcal{F}$-MAIC noninvertible models, first against the airline model and then against the $\mathcal{F}$-MAIC invertible models (from the obvious modification of (16) where the airline model is replaced by the invertible FSM). Among the four series with only noninvertible $\mathcal{F}$-AIC preferred models, compared to its forecasts from the airline model, one series is better forecasted at leads one and 12 by the $\mathcal{F}$-MAIC models, and for two series the $\mathcal{F}$-MAIC model's forecasting is better at lead one and

TABLE 5. Out-of-Sample Forecast Preferences for the 11 Series with $\mathcal{F}$-MAIC Noninvertible FSMs

Preferred Model	1-step forecasting	12-step forecasting
$\mathcal{F}$-MAIC noninvertible FSM	8	6
Airline	1	2
None	2	3

comparable to the airline model's at lead 12. For the remaining series, the $\mathcal{F}$-MAIC model forecasts better at lead one but worse at lead 12. So, taking superior performance at one lead and comparable performance at the other as the criterion for preferring a noninvertible model over another model, the number of series with $\mathcal{F}$-MAIC FSMs preferred over the airline model increases by three to 21, which is 29% of the 72 series.

By the same criterion, six of the other seven $\mathcal{F}$-MAIC noninvertible models excluded from Table 2 are preferable to the airline model, and five of these are also preferable to the $\mathcal{F}$-MAIC invertible models. Table 5 summarizes the results of forecast comparisons to the airline model for the eleven $\mathcal{F}$-MAIC noninvertible models. One of these forecasted worse than the airline model at both leads.

Regarding possibly spurious noninvertibility, for two of the seven series for which the noninvertible model forecasts better than the airline model, deletion of the most recent year or two of data followed by re-estimation yields an invertible model. For both series, the initially estimated noninvertibility is due to an MA(2) root of one. Longer series than those in this study may be required to detect spurious estimates of one for seasonal coefficients in this way.

Noninvertible models with forecasting superiority are immediately useful for the X-11 filter-based adjustment procedure of X-12-ARIMA as this uses only the forecasts of the model. However, for AMB seasonal adjustment, due to current limitations of our software, unit magnitude coefficients must be slightly changed to yield invertible models. That is, an estimate

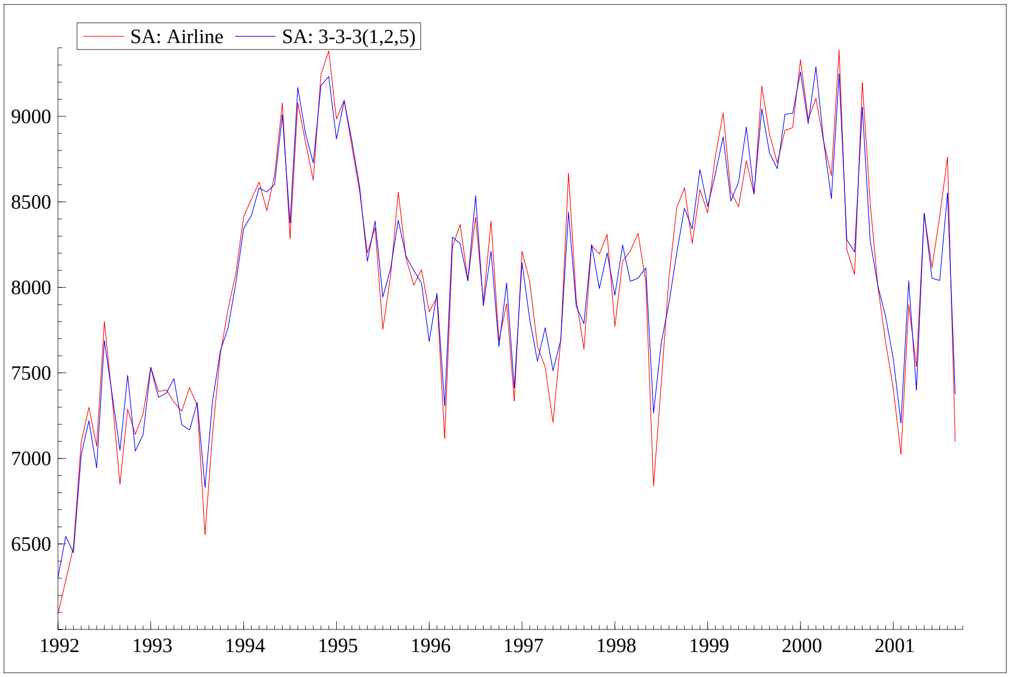


FIGURE 11. AMB Seaonal Adjustments of Manufacturers' Shipments of Automobiles (U36AVS) from the Airline Model and the $\mathcal{F}$-MAIC 3-3-3(1,2,5) Model with the Reset Value $c_1 = 0.999$. The latter model's adjusted series has less extreme movements and is mostly smoother than the airline model's.

$c_1 = 1$ must be changed, e.g. to $c_1 = 0.999 = \sqrt[12]{0.988}$. (Otherwise, in the noninvertible model, the r.h.s. of (4) has a factor $\delta^c(B)$ of degree at least one that coincides with a factor of the differencing operator product on the l.h.s. of (4). Hence, for exact AMB seasonal adjustment calculations, the ARIMA model must be replaced by a reduced model without such common factors and the stationary model that results from applying the remaining differencing operations has a mean function $\mu(t)$ satisfying $\delta^c(B)\mu(t) = 0$ which models some of the seasonality when $\delta^c(B)$ has seasonal factors.)

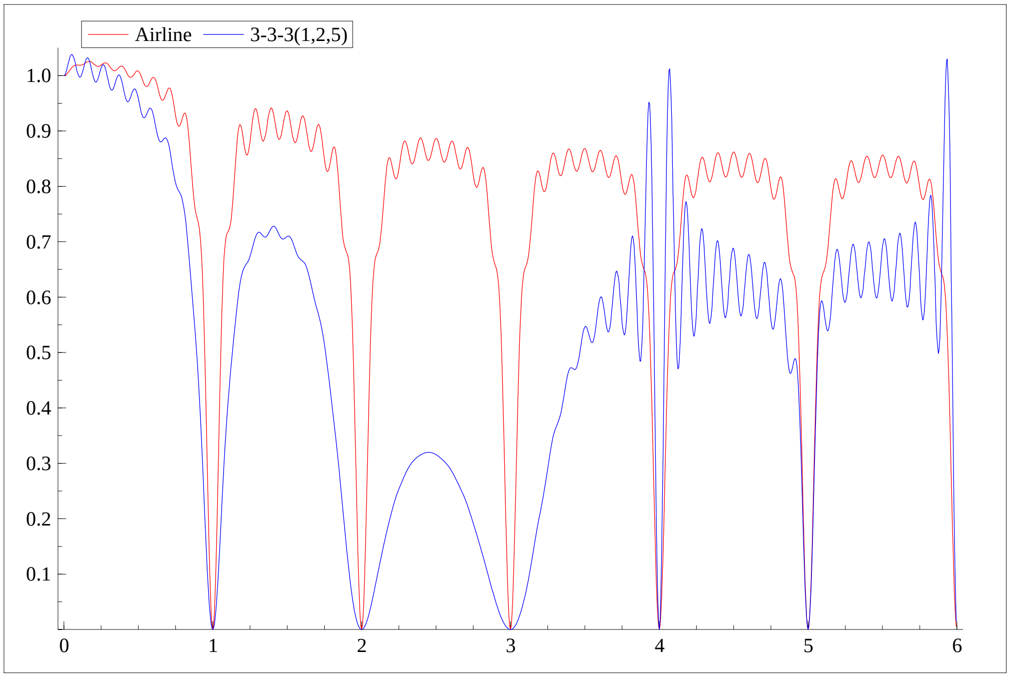


Figure 12. Squared gains of the Concurrent AMB Filters for U36AVS. The airline model has $\theta = 0.479$ and $\Theta = 0.702$. The 3-3-3(1,2,5) model has $a = 0.497$ and $c_2 = 0.945 (= \sqrt[12]{0.507})$. The last results in greater suppression around 1, 2 and 5 cycles per year.

Figures 11 and 12 show the seasonal adjustments and the squared gains of the concurrent filters from the airline model with $\theta = 0.479$ and $\Theta = 0.702$ and from the $\mathcal{F}$-MAIC 3-3-3(1,2,5) model for U36AVS (Value of Shipments of Automobile Manufacturers) with $a = 0.497$, with $c_1 = 0.999$ in place of the value one, and with $c_2 = 0.945 (= \sqrt[12]{0.507})$. The seasonal adjustment of the FSM is usually smoother and, importantly, has notably smaller movements when the movements of the airline model's seasonal adjustment are largest. For each of the other three series having only a noninvertible $\mathcal{F}$-MAIC preferred model, the AMB seasonal adjustment (not shown) is mostly smoother than the airline model's, but the differences are not as conspicuous as with U36AVS.

In summary, noninvertible $\mathcal{F}$-MAIC models often have desirable properties for seasonal adjustment and forecasting, but they require nonstandard diagnostics for comparison with the airline model. Standard inference methods do not apply with noninvertible models because the coefficients of factors with unit magnitude roots have maximum likelihood estimates with nonstandard limiting behavior, including superconvergence (convergence at a rate proportional to the reciprocal of series length); see Chapter 8 of Tanaka (1996). For goodness-of-fit testing in the situation of true noninvertibility, the superconvergence of the MLE estimates of MA coefficients associated with noninvertibility suggests that the degrees of freedom of each Ljung-Box Q statistic should be reduced by the number of such coefficients. Even in this situation, the use of Qs with their p-values as currently calculated, i.e. with no change in the degrees of freedom, might be justifiable as a conservative procedure. However, this is a topic for future research.

5.6. **General Procedure for Model Selection.** The findings above suggest the following general procedure when considering FSM models for seasonal time series.

- Find MLEs for all 72 FSM models and the airline model
- If the $\mathcal{F}$-MAIC model is the airline model, select this model
- If the $\mathcal{F}$-MAIC model is an invertible FSM, select this model
- If the $\mathcal{F}$-MAIC model is a noninvertible FSM, compare the $\mathcal{F}$-MAIC noninvertible FSM model, the best invertible $\mathcal{F}$-MAIC FSM model (if available) and the airline model; choose among these three models on the basis of out-of-sample forecasting performance.

6. Estimation Variability of c_2 for 3-coefficient Models

We consider a final estimation issue. It was observed in simulation results of 3-5-1 and 3-4-2 models (not shown) that the variability of estimates of c_2 in the frequency-specific models is

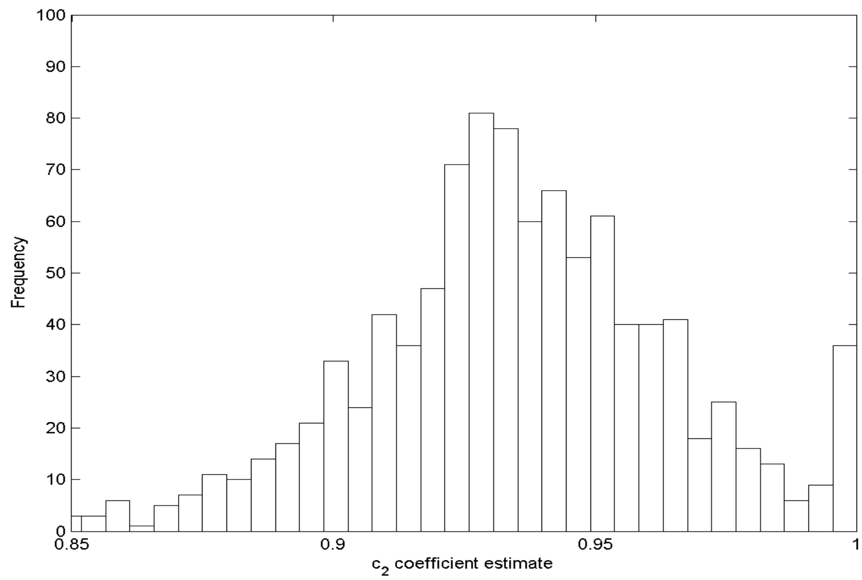


FIGURE 13. Distribution of c_2 estimates for 1000 realizations of the 3-5-1 models with coefficients $a = 0.5$, $c_1 = 0.96$, and $c_2 = 0.93$. 34 estimates were one. Eight were less than 0.85.

substantially greater than that of c_1. Intuitively, this suggests c_1 gains stability by estimating more frequency components than c_2. Here, using histograms of the estimates of c_2 for 3-5-1 and 3-4-2 models, we demonstrate that estimates of c_2 become more stable, and also that fewer spurious estimates with $c_2 = 1$ occur, when this coefficient applies to more frequency components.

1000 realizations of length 150 were generated from 3-5-1 and 3-4-2 models with coefficient values $a = 0.50$, $c_1 = 0.96$, and $c_2 = 0.93$. (These are average values of the coefficients of a set of 21 MAIC preferred 3-5-1 models from data from Section 5.) The histogram of the c_2 estimates of the 3-5-1 model is given in Figure 13. For 3.4% of the realizations, $c_2 = 1$. Figure 14 shows the histogram of c_2 estimates from the 3-4-2 model. Only 1% of the estimates are one and the tails of the histogram are thinner than in Figure 13, indicating less variability.

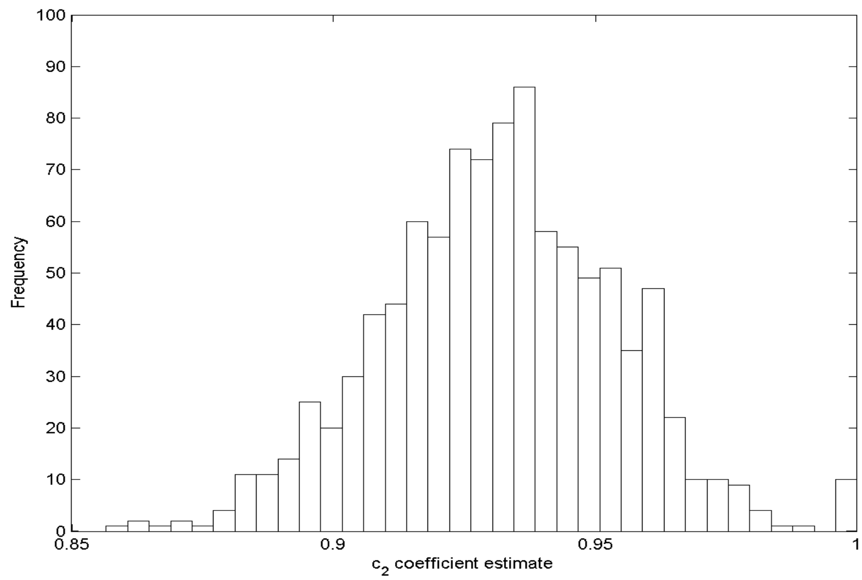


Figure 14. Distribution of c_2 estimates for 1000 realizations of the 3-4-2 model with coefficients $a = 0.5$, $c_1 = 0.96$, and $c_2 = 0.93$. Ten estimates were one. Two estimates were less than 0.85. The tails of the histogram are thinner than in Figure 13, indicating less variability.

7. Concluding Remarks

Spectrum estimates like those of Figures 1 and 8 demonstrate the unsurprising fact that seasonal economic series do not always have frequency components with similar strengths at all seasonal frequencies. However, the current ARIMA model-based seasonal adjustment filters of SEATS and X-13A-S treat variance components around every seasonal frequency in a similar way (as do the nonparametric filters of X-12-ARIMA; see Findley et al. (1998) and Findley and Martin (2006)). This is a consequence of the fact that the seasonal moving average factors of standard seasonal ARIMA models have one coefficient. We have introduced a reparameterization of the seasonal moving average factor that partitions the seasonal frequencies into two groups, each with its own coefficient. The resulting generalizations of the widely used Box-Jenkins airline model were chosen over the airline model for 21 of 72 U.S. Census Bureau economic indicator series that had previously been modeled with an invertible airline model. The choice of the new models for the subset was made with a generalization of Akaike's MAIC criterion that was introduced to deal with the multiplicity of candidate models arising from the partitionings of the six monthly seasonal frequencies 1, 2,...,6 cycles per year into two subsets. The new models usually had better goodness-of-fit diagnostics, smoother AMB seasonal adjustments, and better out-of-sample forecast performance. This last property was used to confirm model selection when the new model chosen was noninvertible, a rather frequently observed phenomenon whose sources and implications were discussed at length.

The three-coefficient FSMs were preferred by $\mathcal{F}$-MAIC significantly more often than the four-coefficient FSMs and were less susceptible to being noninvertible. Thus they are more convenient to use.

Even if attention is restricted to three-coefficient models, there are still 41 such models, or 21 if the 3-3-3 models are not considered. When these numbers are impractically large, as they could be for automatically modeling a large set of series, the six 3-5-1 models by themselves offer substantial benefits, both because of their simpler connection with spectrum diagnostics and because they have been shown to be the most frequently chosen models for the data considered in this study. Yet it should be kept in mind that the use of models from the other families can lead to lower $\mathcal{F}$-AIC values, better goodness-of-fit diagnostics and better forecasting performance. Thus these too should be considered whenever possible.

Although our focus has been on generalizations of the airline model, we have pointed out that any ARIMA model with a seasonal moving average factor like the airline model's can be generalized in the same way. It remains to be determined whether $\Delta^{\mathcal{F}}$s different from those of Table 1 are needed for model selection.

Finally, Figure 2 indicates that the asymptotic distributions of the differences of airline model AIC from FSM $\mathcal{F}$-AIC values closely resemble gamma distributions. This suggests that there could be general asymptotic theoretical principles underlying these empirical distributions that would apply in more general situations in which a set of models of fixed dimension is being compared to a common submodel.

Authors' Notes

The Ox software used to estimate these models is available from the authors. Also, a seasonal adjustment package in C# that includes the models and AIC modification methods of this paper is available at `http://www.nbb.be/app/dqrd/index.htm`. The authors are grateful to Jean Palate of the National Bank of Belgium for taking the initiative to produce this implementation.

The authors are grateful to Kathleen McDonald-Johnson for helpful comments on drafts of this paper and to Demetra Lytras for enhancing the Census Bureau's X-12-Graph program to produce the displayed versions of Figures 1 and 8.

Appendix A. Estimation of the Frequency-Specific Models.

The FSMs are defined in terms of products of moving average factors of degrees one or two rather than in terms of the full MA polynomial of degree $s + 1$. The latter is needed for the state space representation used to calculate the likelihood function and seasonal adjustments; see Durbin and Koopman (2001) for more details on such calculations. The full MA polynomial can be obtained from the factors by writing a routine that carries out polynomial multiplication. However, Fast Fourier Transform functions are already available in Ox and similar software, and these can be used to transform a product of polynomial factors into the coefficient sequence of the product polynomial. Once the full MA polynomial is available, there are routines to produce the ARIMA model's state space representation and implement filtering and smoothing algorithms to obtain maximum Gaussian likelihood values and AMB seasonal adjustments.

References

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In *2nd International Symposium on Information Theory*, pp. 267–281. Akademiai Kiado, Budapest, Hungary.

Box, G. E. P. and G. M. Jenkins (1976). *Time Series Analysis: Forecasting and Control* (2nd ed.). San Francisco, CA: Holden-Day.

Burman, J. P. (1980). Seasonal Adjustment by Signal Extraction. *Journal of Royal Statistical Society A 143*, 321–37.

Doornik, J. A. (2001). *Object-Oriented Matrix Programming Using Ox 3.0*. London: Timberlake Consultants Press.

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

Findley, D. F. (1999). AIC II. In S. Kotz, C. R. Read, and D. L. Banks (Eds.), *Encyclopedia of Statistical Science, Update Volume 3*, pp. 2–6. New York: Wiley.

Findley, D. F. (2005a). Asymptotic Second-Moment Properties of Out-of-Sample Forecast Errors of Misspecified RegARIMA Models and the Optimality of GLS. *Statistica Sinica 15*, 447–476.

Findley, D. F. (2005b). Some Recent Developments and Directions in Seasonal Adjustment. *Journal of Official Statistics 21*, 343–365.

Findley, D. F. and D. E. K. Martin (2006). Frequency Domain Analysis of SEATS and X-11/X-12-ARIMA Seasonal Adjustment Filters for Short and Moderate Length Time Series. *Journal of Official Statistics 22*, 1–34.

Findley, D. F., D. E. K. Martin, and K. C. Wills (2002). Generalizations of the Box-Jenkins Airline Model. In *Proceedings of the American Statistical Association, Section on Business and Economic Statistics [CD-ROM]*, Alexandria.

Findley, D. F., B. C. Monsell, W. R. Bell, M. C. Otto, and B. C. Chen (1998). New Capabilities of the X-12-ARIMA Seasonal Adjustment Program (with discussion). *Journal of Business and Economic Statistics 16*, 127–77. http://www.census.gov/ts/papers/jbes98.pdf.

Findley, D. F. and C.-Z. Wei (2002). AIC, Overfitting Principles, and the Boundedness of Moments of Inverse Matrices for Vector Autoregressions and Related Models. *Journal of Multivariate Analysis 83*, 415–450.

Fischer, B. and C. Planas (2000). Large Scale Fitting of Regression Models with ARIMA Errors. *Journal of Official Statistics 16*, 173–184.

Gómez, V. and A. Maravall (1997). Programs TRAMO and SEATS: Instructions for the User (beta version: June 1997). Working Paper 97001, Ministerio de Economía y Hacienda, Dirrectión General de Análisis y Programación Presupuestaria, Madrid.

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 (1999). Statistical Algorithms for Models in State Space Form Using SsfPack 2.2. *Econometrics Journal 2*, 113–66. http://www.ssfpack.com/.

Ljung, G. M. and G. E. P. Box (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika 65*, 297–304.

Monsell, B. C., J. A. D. Aston, and S. J. Koopman (2003). Towards X-13? In *Proceedings of the American Statistical Association, Section on Business and Economic Statistics [CD-ROM]*, Alexandria. http://www.census.gov/ts/papers/jsm2003bcm.pdf.

Soukup, R. and D. Findley (1999). On the Spectrum Diagnostics Used by X-12-ARIMA to Indicate the Presence of Trading Day Effects After Modeling or Adjustment. In *Proceedings of the American Statistical Association, Section on Business and Economic Statistics*, Alexandria, pp. 144–149.

Tanaka, K. (1996). *Time Series Analysis*. New York: Wiley.

Taniguchi, M. and Y. Kakizawa (2000). *Asymptotic Theory of Statistical Inference for Time Series*. New York: Springer-Verlag.