Minimum Mean Squared Error Estimation of the Noise in Unobserved Component Models

Agustín Maravall

(*Journal of Business and Economic Statistics*, 5, pp. 115-120)

Abstract

In model-based estimation of unobserved components, the minimum mean squared error estimator of the noise component is different from white noise. In this article, some of the differences are analysed. It is seen how the variance of the component is always underestimated, and the smaller the noise variance, the larger the underestimation. Estimators of small-variance noise components will also have large autocorrelations. Finally, in the context of an application, the sample autocorrelation function of the estimated noise is seen to perform well as a diagnostic tool, even when the variance is small and the series is of relatively short length.

Keywords: Seasonal adjustment; Signal extraction; Time series; ARIMA models.

1. MINIMUM MEAN SQUARED ERROR ESTIMATOR OF THE NOISE COMPONENT

Let an observable series z_t be the sum of several orthogonal components, one of them white noise, as in

$$z_t = \sum_i z_{it} + u_t \;, \tag{1.1}$$

where z_{it} denotes an unobservable component and u_t is normally identically independently distributed (niid) $(0, \sigma_u^2)$. Particular cases of (1.1) are the trend-seasonal-(white noise) irregular and the signal-plus-noise decompositions of a time series. The model-based approach to unobserved components estimation involves the use of models for the components of the type

$$z_{it} = \psi_i(B)\, a_{it} \;, \tag{1.2}$$

where $\psi_i(B)$ represents a rational function of the backward shift operator B and the a_{it}'s are orthogonal white noises, each one with variance σ_i^2. Let the overall model for the observed series, consistent with (1.1) and (1.2), be given by

$$z_t = \psi_z(B)\, a_t \;, \tag{1.3}$$

where $\psi_z(B)$ is also a rational function in B, which can be expressed as

$$\psi_z(B) = \theta(B) / \phi(B) \;,$$

where $\theta(B)$ and $\phi(B)$ are polynomials in B of degree q and p, respectively. Unit roots may be present in the autoregressive polynomial $\phi(B)$.

The previous characterization of the model-based decomposition of a time series can be applied to the so-called “reduced form” approach, in which the overall model (1.3) is assumed known and, from this, the models for the components are derived (see, e.g., Burman 1980; Hillmer and Tiao 1982). It can also be applied to the “structural form” approach, in which the models for the components [i.e., Eq. (1.2) for all i] are directly specified (see, e.g., Engle 1978; Harvey and Todd 1983).

The minimum mean squared error (MMSE) estimator of the ith component is given by

$$\hat{z}_{it} = k_i \left[\psi_i(B)\, \psi_i(F) / \psi_z(B)\, \psi_z(F) \right] z_t \;,$$

where $k_i = \sigma_i^2 / \sigma_a^2$ and $F = B^{-1}$ denotes the forward shift operator (see Cleveland and Tiao 1976; Bell 1984). Similarly for u_t, which will be referred to as the noise or irregular component, the estimator becomes

$$\hat{u}_t = k_u \left[1/\psi_z(B)\psi_z(F)\right] z_t , \tag{1.4}$$

where $k_u = \sigma_u^2 / \sigma_a^2$. [Notice that (1.4) will be valid for all admissible decompositions of z_t into (1.1); the differences among these decompositions will simply imply different values of σ_u^2]. As a consequence, the estimator $\hat{u}_t$ given by (1.4), will follow a model different from the theoretical model for u_t, which is white noise. This is a well-known result (see, e.g., Bell and Hillmer 1984) and, although cause for some concern, it has not been the subject of much analysis.

The concern originates from the consideration that if the aim is to remove white noise variation from a series, intuitively it would seem desirable to remove something that is somewhat close to white noise. Equation (1.4), however, guarantees that the estimator $\hat{u}_t$ will not be white noise. How important in practice can this departure from white noise behavior be expected to be?

Cleveland and Tiao (1976) derived the theoretical autocorrelation function (ACF) of $\hat{u}_t$, for their model-based interpretation of X11. It was certainly different from that of white noise, yet all autocorrelations were small, none of them exceeding in absolute value $.2$. In fact, for the example they discussed, although the empirical ACF of $\hat{u}_t$ estimated with X11 was similar to the theoretical ACF of the X11 irregular, it was also close, considering the sample size, to that of a white noise variable (see Cleveland and Tiao 1976, fig. D). Therefore, for series obeying models reasonably close to the model version of X11 and for standard sample sizes, the difference between the models for u_t and $\hat{u}_t$ will be small and the departure from white noise behavior in the estimator will be of little practical importance.

For series with a different structure, however, the departure can be substantial. Using (1.3) in (1.4), $\hat{u}_t$ can be expressed as a function of the innovations (a_t) in the observed series

$$\hat{u}_t = k_u \, \psi_z(F)^{-1} \, a_t ,$$

or

$$\theta(F)\hat{u}_t = k_u \, \phi(F) \, a_t . \tag{1.5}$$

Notice that if a series z_t follows the model (1.3), it also follows the model $z_t = \psi_z(F)\, e_t$, where e_t is the backward innovation $e_t = z_t - E(z_t \mid z_{t+1,} z_{t+2}, ...)$, a white noise variable independent of all future z's. Therefore, corresponding to the model for z_t that uses the forward shift operator and the backward innovations, $\hat{u}_t$ could be expressed alternatively in terms of the backward shift operator and the backward innovations as $\hat{u}_t = k_u\ \psi_z(B)^{-1}\ e_t$.

From expression (1.5) several results are immediately available. First, since $\hat{u}_t$ can be expressed as a linear function of present and future innovations, it follows that $E_t\ \hat{u}_{t+j} = 0$ for j > 0. Hence, although autocorrelated, $\hat{u}_t$ cannot be forecast.

Second, since $\sigma_u^2 > 0$ in (1.1) implies a positive minimum in the pseudospectrum of z_t, the series will be invertible and the roots of $\theta(B)$ will lie outside the unit circle. Thus, considering (1.5), $\hat{u}_t$ will always be stationary, with finite variance. Furthermore, setting (without loss of generality) $\sigma_a^2 = 1$, (1.5) implies

$$V(\hat{u}_t)/\sigma_u^2 = \sigma_u^2\ V(z_t^*) \tag{1.6}$$

where V denotes variance and z_t^* is the inverse process $\theta(B)\, z_t^* = \phi(B)\ a_t$. From Hillmer (1976, expression 4.2.7) it is found that

$$\sigma_u^2\ [\,\phi(e^{i\omega})\,\phi(e^{-i\omega})\,/\,\theta(e^{i\omega})\,\theta(e^{-i\omega})\,] \le 1$$

for all ω in the range $(-\pi, \pi)$. Thus $\sigma_u^2\ V(z_t^*) \le 1$ and hence $V(\hat{u}_t) \le \sigma_u^2$, with equality holding only in the trivial case $z_t = u_t$. Therefore, the variance of the estimator $\hat{u}_t$ is smaller than the variance of the theoretical component u_t. In fact, from (1.6) it can be seen that as the series z_t gets closer to noninvertibility (i.e., as the noise component becomes smaller), the ratio $V(\hat{u}_t)/\sigma_u^2$ tends to zero and $\hat{u}_t$ gets closer to nonstationarity. Hence large differences between the theoretical noise component and its MMSE estimator (in terms of variances and autocorrelations) will characterize series for which the noise component is of little importance.

Finally, Equation (1.5) shows that when the observed series is nonstationary (the case of applied interest), the estimator of the noise will be noninvertible. The zeros in the spectrum of $\hat{u}_t$ will reflect the fact that the zeros of $\phi(B)$ are associated with frequencies for which the ratio of the noise variance to the signal variance in the observed series is zero. Consequently, in terms of noise estimation, these frequencies will provide no information and therefore can be ignored.

As an example, consider the signal-plus-noise decomposition of the integrated moving average (IMA) (1,1) model in Box, Hillmer, and Tiao (1978). The canonical decomposition (i.e., the one with maximum noise variance) of

$$\nabla z_t = (1-\theta B) a_t \quad ,$$

where $\nabla = 1-B$, into

$$z_t = z_{1t} + u_t$$

yields u_t white noise with $k_u = (1+\theta)^2 / 4$. From (1.5), $\hat{u}_t$ can be expressed as the stationary autoregressive moving average (ARMA) (1,1) process

$$(1-\theta F) \hat{u}_t = k_u (1- F) a_t \quad , \tag{1.7}$$

so $\rho_1 = -(1-\theta)/2$, $\rho_k = \theta \rho_{k-1}$ $(k > 1)$, and

$$V(\hat{u}_t)/\sigma_u^2 = (1+\theta)/2 \ . \tag{1.8}$$

This ratio always lies in the range (0,1). Both variances will be equal when the observed series is pure noise. On the contrary, as θ approaches -1 and the noise variance becomes smaller, the autocorrelations of $\hat{u}_t$ become larger in absolute value and the ratio (1.8) will tend to zero. When θ = -.9 , for example, $\rho_1 = -.95$, and the variance of the one-step-ahead forecast error of z_t is 40 times larger than the variance of the theoretical noise component, which, in turn, is 20 times larger than that of its estimator. Notice also that the moving average (MA) factor $(1-F)$ in (1.7) implies a zero in the spectrum of $\hat{u}_t$ for the zero frequency, for which the pseudospectrum of z_t displays an infinite peak.

The conclusion is that when the noise component is relatively small, little of it will be captured by MMSE estimation. Still, as shall be seen in the next section, even for relatively short series the information contained in this estimator can be of considerable interest.

2. A DIAGNOSTIC CHECK

One of the advantages of a model-based approach is that it provides the grounds for analysis of results by comparing theoretical models with the obtained estimates. In practice, u_t is not estimated directly using (1.4), but as the residual

$$\hat{u}_t = z_t - \sum_i \hat{z}_{it} \ , \tag{2.1}$$

after the other components have been removed from the series. If equations (1.1) – (1.3) are correct and the components orthogonal, then (2.1) can be rewritten as

$$\begin{aligned} \hat{u}_t &= z_t - \sum_i k_i \frac{\psi_i(B)\psi_i(F)}{\psi_z(B)\psi_z(F)} z_t \\ &= \frac{1}{\psi_z(B)\psi_z(F)} \left[\psi_z(B)\psi_z(F) - \sum_i k_i \ \psi_i(B)\psi_i(F) \right] z_t \quad . \end{aligned}$$

Since the term in brackets is equal to σ_u^2, expression (1.4) is finally obtained; hence $\hat{u}_t$, computed as the residual, should also satisfy (1.5). The comparison of the theoretical autocorrelations of $\hat{u}_t$, obtained from (1.5), with the empirical ones for the estimated noise (as suggested originally by Cleveland and Tiao 1976) provides a natural, easy-to-compute way to evaluate results in a particular model-based decomposition. Discrepancies between the two autocorrelations would reveal inadequacies in the procedure that can be due to an incorrect specification of the model for the observed series.

Although comparison of the theoretical and empirical autocorrelations may in theory provide a check on the results, how can the comparison be expected to perform in practice? This issue will be addressed in the context of an example, which considers seasonal adjustment of the monthly series of insurance operations (IO's) one of the components (small, though not trivial) of the Spanish monetary aggregates. The series starts in January 1979 and ends in October 1985; it consists therefore of 82 observations. Thus, we are analyzing a relatively short monthly series.

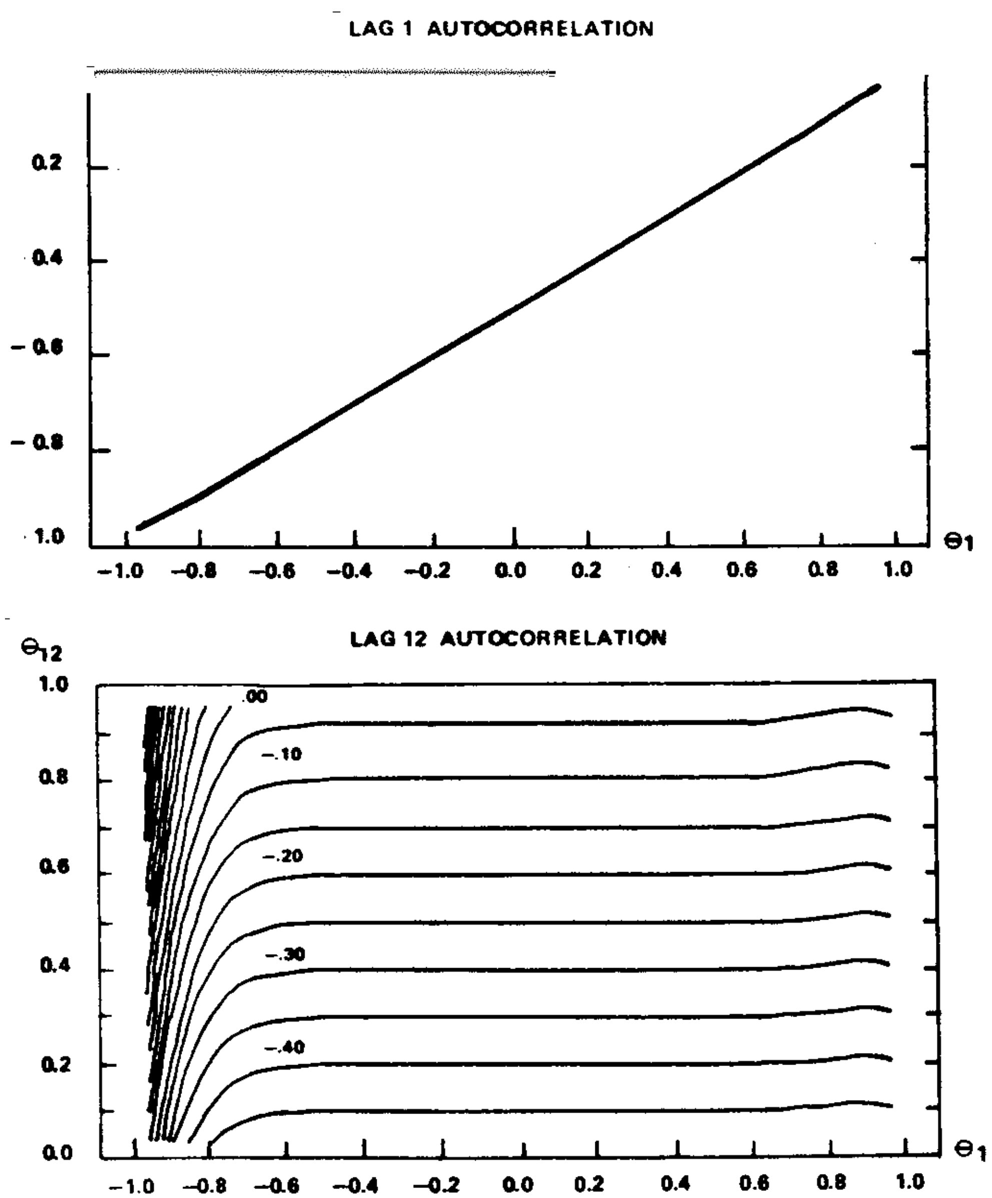

Figure 1: Theoretical Autocorrelations of the Irregular Component Estimator for the Airline Model

When adjusting the series with the program X11 ARIMA (default option), all tests for the presence of seasonality were highly significant. The results failed to pass, however, one of the quality assessment tests, which indicated that there was too much autocorrelation in the irregular. The lag-1 autocorrelation of the estimated irregular was, in fact, $\hat{\rho}_1 = .42$, quite a distance away from $\rho_1 = -.2$, the theoretical value implicit in the MMSE model-based approximation to X11 developed by Cleveland and Tiao (1976). The discrepancy between the two values clearly indicates that X11 cannot be given Cleveland and Tiao's MMSE model-based interpretation in this case, but I am not interested in comparing a non-model-based method with a model-based method (in this respect, see the "consistency with the data" check of Bell and Hillmer 1984). My aim is instead to judge the adequacy of a particular model-based decomposition.

Seasonal adjustment often requires, in practice, treatment of a large number of series, making it unfeasible to perform previous univariate analysis for each of them. Thus there is a need for some standard model that approximates reasonably well a large number of series and can therefore be applied routinely (the "common central model" of Sims 1985). An obvious candidate among ARIMA models is the airline model of Box and Jenkins (1970), given by

$$\nabla\nabla_{12}\, z_t = (1-\theta_1 B)(1-\theta_{12} B^{12})\, a_t \quad , \tag{2.2}$$

which is known to approximate many series encountered in practice. Model-based decomposition of z_t into trend, seasonal, and white noise irregular implies that the MMSE estimator of the irregular, u_t, follows the process

$$(1-\theta_1 F)(1-\theta_{12} F^{12})\, \hat{u}_t = k_u (1-F)(1-F^{12})\, a_t \quad , \tag{2.3}$$

from which the theoretical autocorrelations can be computed.

Figure 1 displays the lag-1 and lag-12 autocorrelations of $\hat{u}_t$ for the different values of the θ parameters within the admissible region ($-1 < \theta_1 < 1,\ 0 < \theta_{12} < 1$; see Hillmer and Tiao 1982). The parameter θ_{12} has practically no effect on ρ_1; and θ_1, except for large negative values, does not affect ρ_{12}. In addition, ρ_1 is always negative and lies in the region (0, -1). As for ρ_{12}, although the range goes from -.5 to .5, unless θ_1 is close to –1, its value will also be negative. In both cases, the smaller the θ parameter, the larger the corresponding autocorrelation of $\hat{u}_t$ will be in absolute value.

To see how close one can expect to get to the theoretical autocorrelations in a particular realization, the values selected for θ_1 and θ_{12} are those of the actual airline model, $\theta_1 = .4$, $\theta_{12} = .6$. [For a series following this model, Cleveland and Tiao (1976) showed that "the census (X11) procedure works reasonably well" (p. 584).] The theoretical standard deviation and lag-1 and lag-12 autocorrelations of $\hat{u}_t$ are given in the first column of Table 1.

Table 1. Airline Model

	Theoretical irregular component	Simulated distribution Mean	Simulated distribution Standard deviation	Estimated irregular component
ρ_1	-.30	-.30	.10	.40
ρ_{12}	-.20	-.24	.10	-.35
Standard deviation*	.42	.43	.02	.22

* In units of σ_a^2

Although the distributions of the autocorrelation estimators are complicated to derive, they can be approximated by simulation. Two hundred fifty independent series of 84 observations (equivalent to seven years of data) were generated for the Airline model (with $\sigma_a^2 = 1$). For each series, a theoretically white noise irregular was estimated using Burman's program (see Burman 1980), and for each irregular component series, the standard deviation and autocorrelations were estimated. In this way, empirical distributions are obtained for these estimators; their means and standard deviations are given in the second and third columns of Table 1. The estimator of ρ_1 appears to be unbiased, although $\hat{\rho}_{12}$ has a small bias. When the length of the series was increased to 11 and then to 14 years, however, the mean of $\hat{\rho}_{12}$ became -.23 and -.22, respectively, approaching the theoretical value. (The same simulation was repeated twice with 250 series, and then four times using 150 series. The differences in the results were minor and none of the values reported in Table 1, for example, changed by more than .01 .)

One issue that has not been addressed is that of revisions. The estimator $\hat{u}_t$ defined by (1.4), is given by a centered and symmetric filter applied to the observed series z_t. This implies that, for periods relatively close to the present, observations needed to complete the filter will not be available yet. Replacing them with forecasts, a preliminary estimator can be computed and, with the passing of time, as forecasts are either updated or replaced with new observations, the estimator of u_t will be revised (see Pierce 1980). As a consequence, the end values of the estimated irregular series will be contaminated by revision error. This contamination will certainly have an effect on the estimators of the moments of $\hat{u}_t$, but Table 1 shows that even for series as short as seven years, the effect on bias is small. As for the precision, the standard deviation of $\hat{\rho}_1$ and of $\hat{\rho}_{12}$ is of the order of .10. (For other lags, the standard deviations of the autocorrelation estimators were larger.)

Table 2. Standard Deviation of the Autocorrelation Estimator

T	Lag 1	Lag 12	$1/\sqrt{T}$
84	.10	.10	.11
132	.09	.09	.09
168	.07	.07	.08

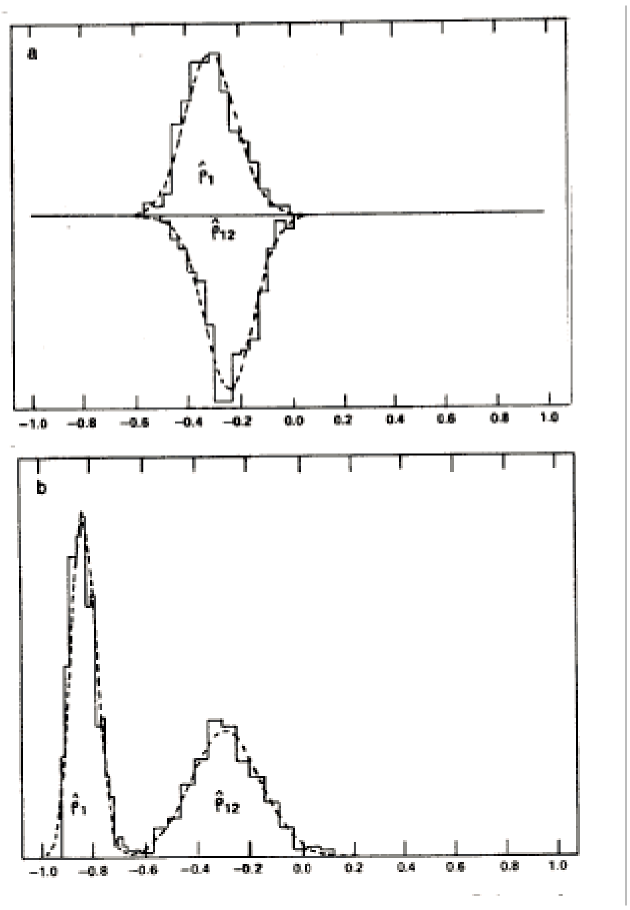

Figure 2: Empirical Distributions of Autocorrelation Estimates for the Irregular Component : (a) Airline Model; (b) Model (2.4)

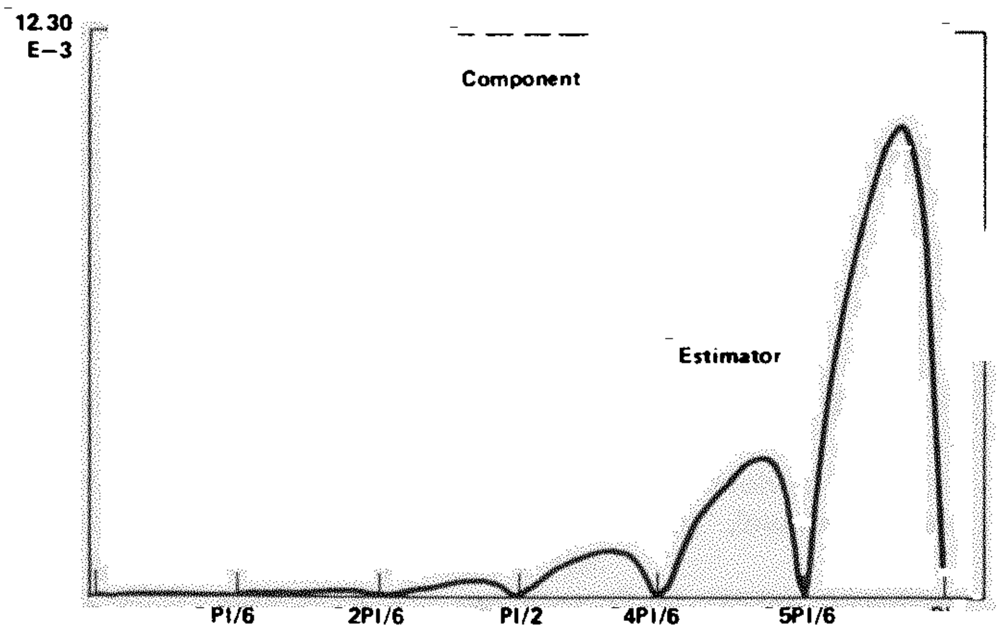

Figure 3: Spectrum of the Irregular Component and Its Estimator
[Model (2.4)]

The series length was increased to 11 and then to 14 years, and again 250 realizations were generated in each case. Table 2 compares the standard deviations of $\hat{\rho}_1$ and $\hat{\rho}_{12}$ for the different sample sizes. It is seen that $1/\sqrt{T}$ seems to be a roughly correct (slightly conservative) approximation to the standard deviations obtained. Finally, the empirical distributions of $\hat{\rho}_1$ and $\hat{\rho}_{12}$ are plotted in Figure 2a, in which they are compared with the Normal approximation with the same mean and variance. The skewness estimates are, respectively, .23 and -.06, and the estimates of kurtosis are 2.83 and 2.80 . It seems reasonable to conclude that for the model considered, even for relatively short series, the check of whether $\hat{\rho}_1$ and $\hat{\rho}_{12}$ fall within the range $\pm 2/\sqrt{T}$ of their theoretical value may provide a useful diagnostic tool.

For the log of the series of IO's, the autocorrelation estimates are given in the last column of Table 1. Although $\hat{\rho}_{12}$ is acceptable (borderline), $\hat{\rho}_1$ (= .40) is unquestionably outside the range (-.10, -.50). (Notice that the estimate is close to the corresponding one obtained for the X11 irregular.) Hence the check agrees with the X11 error message and clearly indicates that model-based decomposition of the series IO using the airline model is not appropriate. The result is confirmed by the estimator of the standard deviation of $\hat{u}_t$ (.22), which also falls outside the acceptance range (.38, .46 .)

In fact, a more adequate model than the airline one for the log of the IO series is

$$\nabla^2 \nabla_{12} \, z_t = (1 - .106\,B - .496\,B^2)(1 - .437\,B^{12})\,a_t \quad , \qquad (2.4)$$

with $\sigma_a^2 = .0234$ (about half the size of σ_a^2 for the airline model). MMSE model-based decomposition of (2.4) yields the following estimator of a white noise irregular:

$$\begin{aligned} &(1 - .106\,F - .496\,F^2)(1 - .437\,F^{12})\,\hat{u}_t \\ &\qquad = k_u \, (1 - F)^2 \, (1 - F^{12})\,a_t \quad , \qquad (2.5) \end{aligned}$$

with $k_u = \sigma_u^2 / \sigma_a^2 = .0123$ when σ_u^2 attains its maximum ("canonical") value. Figure 3 compares the spectrum of the (white noise) u_t with that of its estimator $\hat{u}_t$, and the difference between them is remarkable. The zeros in the spectrum of $\hat{u}_t$ correspond to the unit roots in the autoregressive part of the model for z_t. Comparing the areas under both spectra, it is seen how $V(\hat{u}_t)$ underestimates σ_u^2 [in fact, $V(\hat{u}_t) = \sigma_u^2 / 8$]. Thus we are looking at a series with a very small irregular, which is in turn, strongly underestimated (in accordance with the result in the previous section).

From (2.5), the theoretical variances and autocorrelations of $\hat{u}_t$ can be derived. They are given in the first column of Table 3. A simulation similar to the one for the airline model was done, and the results are presented in the second and third columns of Table 3 and in Figure 2b. As in the airline model case, the estimators of the standard deviation and $\hat{\rho}_1$ and $\hat{\rho}_{12}$ appear to be reasonably unbiased and precise. The fourth column of Table 3 displays the standard deviation and autocorrelations of the irregular component estimated with Burman's program; in this case, the three estimates are comfortably in agreement with their theoretical values.

Table 3. Model (2.4)

	Theoretical irregular component	Simulated distribution: Mean	Simulated distribution: Standard deviation	Estimated irregular component
ρ_1	-.83	-.83	.05	-.82
ρ_{12}	-.27	-.29	.13	-.09
Standard deviation*	.040	.040	.005	.034

* In units of σ_a^2

In summary, a series with an irregular (or noise) component variance as small as 1% of the variance of the one-step-ahead forecast error has been considered. The discussion suggests that, even in that case and for series as short as seven years of monthly data, the comparison between the theoretical and empirical second moment of the irregular estimator performs well as a way of evaluating results in MMSE model-based decomposition of time series.

ACKNOWLEDGMENTS

I thank two anonymous referees for their helpful comments.

[Received July 1985. Revised January 1986.]

REFERENCES

BELL, W.R. (1984), "Signal Extraction for Nonstationary Time Series". *Annals of Statistics* 12, 646-664.

BELL, W.R. and HILLMER, S.C. (1984), "Issues involved with the Seasonal Adjustment of Economic Time Series". *Journal of Business and Economic Statistics* 2, 291-320.

BOX, G.E.P., HILLMER, S.C. and TIAO, G.C. (1978), "Analysis and Modeling of Seasonal Time Series", in Zellner, A. (ed.), *Seasonal Analysis of Economic Time Series*, Washington, D.C.: U.S. Dept. of Commerce-Bureau of the Census, 309-334.

BOX, G.E.P. and JENKINS, G.M. (1970), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.

BURMAN, J.P. (1980), "Seasonal Adjustment by Signal Extraction", *Journal of the Royal Statistical Society* A, 143, 321-337.

CLEVELAND, W.P. and TIAO, G.C. (1976), "Decomposition of Seasonal Time Series: A Model for the X-11 Program", *Journal of the American Statistical Association* 71, 581-587.

ENGLE, R.F. (1978), "Estimating Structural Models of Seasonality", in Zellner, A. (ed.), *Seasonal Analysis of Economic Time Series*, Washington, D.C.: U.S. Dept. of Commerce-Bureau of the Census, 281-297.

HARVEY, A.C. and TODD, P.H.J. (1983), "Forecasting Economic Time Series with Structural and Box-Jenkins Models: A Case Study", *Journal of Business and Economic Statistics* 1, 299-306.

HILLMER, S.C. (1976), "Time Series: Estimation, Smoothing and Seasonal Adjustment", Ph.D. dissertation, University of Wisconsin, Statistics Department.

HILLMER, S.C. and TIAO, G.C. (1982), "An ARIMA-Model Based Approach to Seasonal Adjustment", *Journal of the American Statistical Association* 77, 63-70.

PIERCE, D.A. (1980), "Data Revisions in Moving Average Seasonal Adjustment Procedures", *Journal of Econometrics* 14, 95-114.

SIMS, C.A. (1985), "Comment on 'Issues Involved with the Seasonal Adjustment of Economic Time Series' by W.R. Bell and S.C. Hillmer", *Journal of Business and Economic Statistics* 3, 1, 92-94.