Modifications of SEATS' Diagnostic for Detecting Over- or Underestimation of Seasonal Adjustment Decomposition Components

David F. Findley, Tucker S. McElroy and Kellie C. Wills
U.S. Census Bureau

KEY WORDS Seasonal adjustment (model-based), Residual seasonality, Spectrum

Abstract

We consider two modifications of SEATS' diagnostics for determining whether, for an estimated seasonal decomposition component, there is underestimation or overestimation, meaning inadequate or excessive suppression of the other components. The new diagnostics are calculated from time-varying variances associated with the finite-length filters actually used. They thereby avoid SEATS' diagnostics' strong bias toward identifying underestimation. Tests for the statistical significance of any indicated misestimation are presented and analyzed.

1. Introduction

Let $Y_t, 1 \leq t \leq N$ be a series modeled by a seasonal ARIMA model that has a seasonal-trend-irregular decomposition

$$Y_t = S_t + T_t + I_t,\ 1 \leq t \leq N. \tag{1}$$

The component models lead to estimates $\hat{S}_t$, $\hat{T}_t$, and $\hat{I}_t$ that are Gaussian conditional means, e.g.

$$\hat{I}_t = E\left(I_t | Y_s, 1 \leq s \leq N\right) = \sum_{j=t-N}^{t-1} c_{j,t}^I\left(N\right) Y_{t-j}. \tag{2}$$

Specifically, the $c_{j,t}^I\left(N\right)$ are chosen to minimize $E\left(I_t - \sum_{j=t-N}^{t-1} c_{j,t}^I Y_{t-j}\right)^2$ under assumptions that make it possible to evaluate this expectation by treating the values of model parameters as if they were correct. (For simplicity, we shall usually suppress the dependence of the estimates on N, the models, and the decomposition.) Here and throughout, E denotes the mean *calculated according to the model specified for the data*, whether or not this model is correct. When the model is incorrect, we use E^{true} to denote the mean calculated using the true distribution of the data (which is known for simulated time series). Model inadequacy can lead to inadequacies in this decomposition. The most fundamental inadequacy is the presence of an easily detectable seasonal component in the adjusted series, $\hat{A}_t = Y_t - \hat{S}_t$, or in the detrended seasonally adjusted series $\hat{I}_t = \hat{A}_t - \hat{T}_t$, i.e., the estimated irregular component. Therefore seasonal adjustment programs need a diagnostic (or several) to detect residual seasonality. Spectrum estimates are the most developed and widely used diagnostics for the detection of essentially periodic components such as seasonality and trading day effects and are part of the automatic output of X-12-ARIMA, see Findley et al. (1998).

Instead of a spectrum estimate, SEATS (Maravall and Gomez, 1997) has a diagnostic for "underestimation" and "overestimation." Maravall (2003) defines underestimation of the seasonal component heuristically to mean that its estimate fails to capture all of the seasonal variation, and overestimation means too much variation has been assigned to this component. To have a heuristic definition that applies to all components, we interpret underestimation of the component of interest to mean that the signal extraction filters used do not adequately suppress the other component or components. This will be the case for the seasonal adjustment filters, for example, if the dips in their squared gain functions at the seasonal frequencies are too narrow. Overestimation means too much suppression, as happens if these dips are too wide. The irregular filters provide suppression both around the seasonal frequencies and also around near-zero frequencies associated with trend, so misestimation of I_t could also result from inappropriate suppression of low-frequency (long-period) components. Some examples of squared gains of irregular component extraction filters illustrating overestimation and underestimation are given in Fig.1.

2. The Basic Diagnostic

We focus on detecting misestimation of the irregular component because the model for this component is stationary and usually fully specified by a constant variance, $\sigma_I^2 = EI_t^2$. As a consequence, the signal extraction filter formulas are simpler than for the other components, and so are the formulas of the over/underestimation diagnostics.

Suppose the (estimated or fixed) ARIMA model for the Y_t is written in the usual backshift operator polynomial notation as

$$\delta(B)\,\phi(B)\,Y_t = \eta(B)\,a_t. \tag{3}$$

Thus $\delta(B) = 1 - \delta_1 B - \cdots - \delta_d B^d$ denotes the differencing operator which transforms Y_t to stationarity, e.g. $\delta(B) = (1-B)^2\left(1 + B + \cdots + B^{s-1}\right)$, a_t is the one-step-ahead prediction-error process (with variance parameter σ_a^2), etc. Here s is the number of observations per year: ($s = 12$ in our analyses), $\phi(0) = \eta(0) = 1$, and the zeros of $\phi(z)$ have magnitudes exceeding one. For simplicity, the same will be assumed of $\eta(z)$.

We begin with our basic diagnostic and describe how it differs from SEATS' diagnostic. Both start with

$$\overline{\hat{I}^2} = \frac{1}{N}\sum_{t=1}^{N} \hat{I}_t^2, \tag{4}$$

which SEATS calls the "variance of the estimate". Set $\sigma_t^2 = E\hat{I}_t^2$. We reformulate SEATS' diagnostic to be that overestimation is indicated when $\overline{\hat{I}^2} > E\overline{\hat{I}^2} (= N^{-1}\sum_{t=1}^{N}\sigma_t^2)$, underestimation when $\overline{\hat{I}^2} < E\overline{\hat{I}^2}$. The variances σ_t^2 depend on t. In their place, SEATS' actual diagnostic uses the variance $\sigma_{WK,I}^2$ of the Wiener-Kolmogorov "estimator" of I_t from bi-infinite data, $I_{WK,t} = E\left(I_t | Y_s, -\infty < s < \infty\right)$, which does not depend on t but requires infinitely many Y_s values when the degree of $\eta(B)$ is positive as in (9) below. In this case, $\sigma_t^2 < \sigma_{WK,I}^2$ holds for all t, because of $E(I_t - \hat{I}_t)^2 > E(I_t - I_{WK,t})^2$ and

$$\begin{aligned} EI_t^2 &= EI_t^2 + E\left(I_t - \hat{I}_t\right)^2 \\ &= \sigma_{WK,I}^2 + E\left(I_t - I_{WK,t}\right)^2, \end{aligned}$$

see Findley et al. (2003). Consequently, $E\overline{\hat{I}^2} < \sigma_{WK,I}^2$, with the result that SEATS' diagnostic is biased toward indicating underestimation. (Some quantitative bias results are given in Section 3.3).

In practice, the calculations that produce the component models in SEATS yield variances such as σ_I^2, σ_t^2 and $\sigma_{WK,I}^2$ calculated as though the innovation variance of (3) were equal to one. We shall denote these unscaled variances by σ_I^2/σ_a^2 , σ_t^2/σ_a^2 and $\sigma_{WK,I}^2/\sigma_a^2$. Thus, they must be scaled by multiplication with some estimate $\hat{\sigma}_a^2$. Let $\sigma_{a,mle,N}^2$ denote the maximum likelihood estimate of σ_a^2 given by (17) in the Appendix. Following SEATS, we shall use the bias-corrected estimate of Ansley and Newbold (1981),

$$\hat{\sigma}_a^2 = \left[\left(N - n_{\delta,\phi}\right) / \left(N - n_{\delta,\phi} - n_{coeffs}\right)\right] \sigma_{a,mle,N}^2, \tag{5}$$

where $n_{\delta,\phi} = d$ (or d plus the degree of $\phi(B)$ if SEATS' conditional estimate of $\phi(B)$ is used), and n_{coeffs} is the number of estimated ARMA coefficients in the model. When ARMA parameters are fixed, as in some of our simulations below, then $n_{coeffs} = 0$ and $\hat{\sigma}_a^2 = \sigma_{a,mle,N}^2$. SEATS' criteria are $\overline{\hat{I}^2} > \hat{\sigma}_a^2\left(\sigma_{WK,I}^2/\sigma_a^2\right)$ for overestimation, and $\overline{\hat{I}^2} < \hat{\sigma}_a^2\left(\sigma_{WK,I}^2/\sigma_a^2\right)$ for underestimation. Instead we use

$$\overline{\hat{I}^2} > \frac{\hat{\sigma}_a^2}{N}\sum_{t=1}^{N}\frac{\sigma_t^2}{\sigma_a^2} \tag{6}$$

for overestimation and

$$\overline{\hat{I}^2} < \frac{\hat{\sigma}_a^2}{N}\sum_{t=1}^{N}\frac{\sigma_t^2}{\sigma_a^2} \tag{7}$$

for underestimation. A formula for $N^{-1}\sum_{t=1}^{N}\sigma_t^2/\sigma_a^2$ is given in Subsection 6.2. One-sided tests of (6) and (7) will be described in Section 4.

3. Investigating and Modifying the Diagnostic

3.1 Mean performance of (6) and (7)

In the numerical results we present, Y_t follows the Box-Jenkins airline model,

$$(1-B)(1-B^{12})Y_t = \left(1 - \tilde{\theta}B\right)\left(1 - \tilde{\Theta}B^{12}\right)\tilde{a}_t, \tag{8}$$

with $\sigma_{\tilde{a}}^2$ denoting the variance of $\tilde{a}_t$. To obtain a "proof of concept" of our modification (6) and (7) of SEATS's diagnostic, we compare the means of their left- and right-hand sides for zero mean Gaussian data Y_t satisfying (8) with $\tilde{\theta} = 0.6 = \tilde{\Theta}$ and with SEATS seasonal adjustments from specified models

$$(1-B)(1-B^{12})Y_t = (1-\theta B)\left(1 - \Theta B^{12}\right)a_t, \tag{9}$$

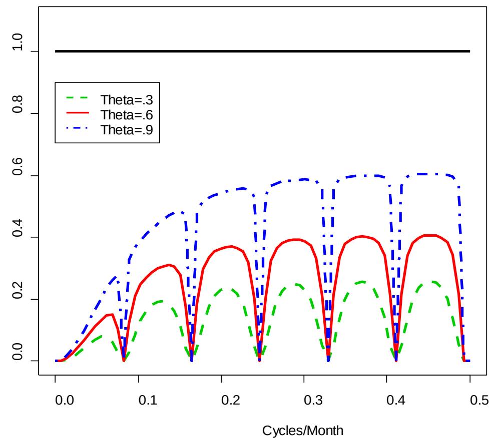

Figure 1: Squared gain function of the midpoint irregular filters of length 144 for $\theta = 0.6$ and $\Theta = 0.3, 0.6, 0.9$.

with $\theta = \tilde{\theta} = 0.6$ and $\Theta = 0.3, 0.4, \ldots, 0.9$. Similar results were obtained with other values of $\tilde{\theta}$ and $\tilde{\Theta}$. Graphs of the squared gain functions $\left|\sum_{j=t-N}^{t-1} c_{j,t}^{I}(N)\, e^{i2\pi j\lambda}\right|^2, -0.5 \leq \lambda \leq 0.5$ from SEATS' decompositions of (9) show that overestimation occurs when $\Theta < \tilde{\Theta}$ and underestimation occurs when $\Theta > \tilde{\Theta}$; see Fig.1 for the case $t = [N/2] + 1$ (midpoint) and $\tilde{\Theta} = 0.6$.

The mean $E^{true}\overline{\hat{I}^2}$ of the left-hand-sides of (6) and (7) is given by (22) in the Appendix. With $\sigma^2_{a,mle,N}(\theta, \Theta)$ denoting the value of $\sigma^2_{a,mle,N}$ (see (17)) specified by (9), the r.h.s. mean is the product

$$\left\{\frac{1}{N}\sum_{t=1}^{N}\frac{\sigma_t^2}{\sigma_a^2}\right\} E^{true}\sigma^2_{a,mle,N}(\theta, \Theta), \qquad (10)$$

which can be calculated from (18) and (23) of the Appendix. Values of the ratios

$$\left\{\frac{1}{N}\sum_{t=1}^{N}\frac{\sigma_t^2}{\sigma_a^2}\right\}\frac{E^{true}\sigma^2_{a,mle,N}(\theta, \Theta)}{E^{true}\overline{\hat{I}^2}}$$

are given in Table 1.

Table 1 shows that $\Theta < 0.6$ leads to ratio values greater than one and therefore to (6) on average, whereas $\Theta > 0.6$ leads to ratio values less than one, hence to (7) on average. Thus, on average, (6) and (7) provide the correct diagnoses.

The ratios are closer to 1.0 in the case of underestimation, suggesting this will be more difficult to detect than overestimation. Moreover, the ratios suggest that the most extreme underestimation case, $\Theta = 0.9$, is more difficult to detect with this approach than overestimation from using $\Theta = 0.7$ or 0.8.

Table 1. True Means of $\overline{\hat{I}^2}$ from Series with $\tilde{\theta} = \tilde{\Theta} = 0.6$ Compared to (10) for Various Θ

Θ	$E^{true}\overline{\hat{I}^2}$	(10)	Ratio
0.3	0.1634	0.1264	1.2927
0.4	0.1879	0.1607	1.1692
0.5	0.2148	0.2003	1.0724
0.7	0.2834	0.2966	0.9555
0.8	0.3356	0.3534	0.9496
0.9	0.4036	0.4135	0.9761

3.2 Local properties and a further modification

The properties revealed by Table 1 for the average of the I_t^2 make it natural to ask if the average inherits these properties from analogous local properties. That is, for every t, is $E^{true}\hat{I}_t^2$ less than

$$\bar{\sigma}_t^2 = \left\{\sigma_t^2/\sigma_a^2\right\} E^{true}\sigma^2_{a,mle,N}(\theta, \Theta) \qquad (11)$$

when there is underestimation and greater than $\bar{\sigma}_t^2$ when there is overestimation? We found for the models of Table 1 that such a local property holds with overestimation. It does not hold, for certain t, with underestimation: the values of $E^{true}\hat{I}_t^2$ in this case are consistently *larger* than $\bar{\sigma}_t^2$ near the ends of the series. Figs.2–4 present graphs of $E^{true}\hat{I}_t^2$ and $\bar{\sigma}_t^2$ that illustrate these findings.

The time intervals at the ends of the series over which $E^{true}\hat{I}_t^2 > \bar{\sigma}_t^2$ are substantially wider in Fig.4 for the case $\Theta = 0.9$ than in Fig.3 for $\Theta = 0.8$. For sample sizes not too much smaller than $N = 144$, these results suggest that the sum in (4) and the sum

$$\frac{1}{N}\sum_{t=1}^{N}\frac{\sigma_t^2}{\sigma_a^2}, \qquad (12)$$

should be restricted to run from 13 to $N - 12$, to obtain a second modification of SEATS' diagnostic that might be able to identify underestimation more reliably. With

$$\tau_N^{(1)} = \overline{\hat{I}^2} - \frac{\hat{\sigma}_a^2}{N}\sum_{t=1}^{N}\frac{\sigma_t^2}{\sigma_a^2} \qquad (13)$$

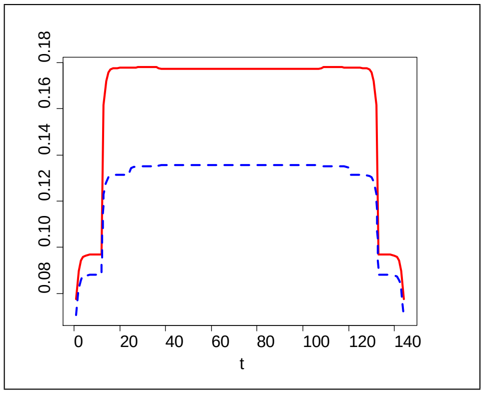

Figure 2: $E^{true}\hat{I}_t^2$ (solid) and $\bar{\sigma}_t^2$ (dots) from $\Theta = 0.3$

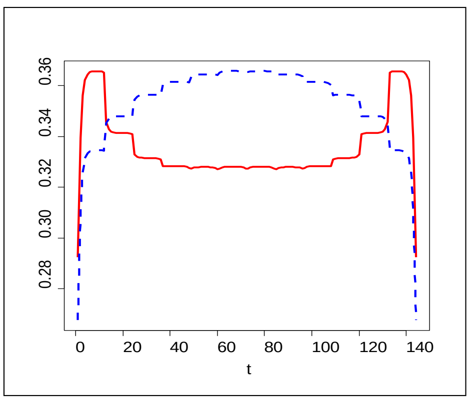

Figure 3: $E^{true}\hat{I}_t^2$ (solid) and $\bar{\sigma}_t^2$ (dots) from $\Theta = 0.8$

being the statistic based on (4) and (12), we define $\tau_N^{(2)}$ to be analogous statistic obtained by restricting the sums as indicated.

3.3 Basic performance of the modified diagnostics

We begin our exploration of $\tau_N^{(1)}$ and $\tau_N^{(2)}$ by investigating how often, in a simulation experiment, their signs, and the sign of the corresponding SEATS diagnostic,

$$\tau_N^{SEATS} = \overline{\hat{I}^2} - \hat{\sigma}_a^2 \left(\sigma_{WK,I}^2/\sigma_a^2\right) \qquad (14)$$

are correct. 5000 independent realizations of (8) with $\tilde{\theta} = \tilde{\Theta} = 0.6$. of length 144 were obtained from pseudo-$IN\,(0,1)$ innovations $\tilde{a}_t$. Always using $\theta = 0.6$, overestimated irregulars were obtained from SEATS by specifying adjustment via a model (9) with $\Theta < 0.6$, underestimated irregulars by specifying $\Theta >$

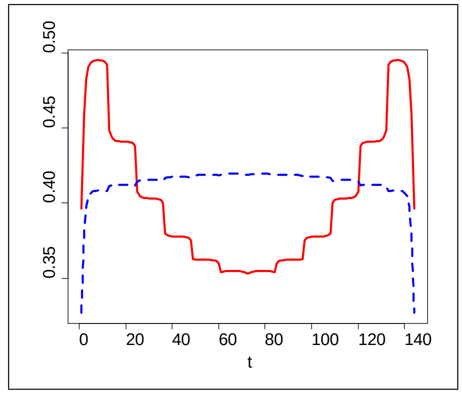

Figure 4: $E^{true}\hat{I}_t^2$ (solid) and $\bar{\sigma}_t^2$ (dots) from $\Theta = 0.9$

0.6. Table 2 lists the underestimation percents. The SEATS diagnostic has a strong bias toward incorrectly indicating underestimation when $0.4 \leq \Theta \leq 0.6$, so it is not a reliable diagnostic for residual seasonality. $\tau_N^{(1)}$ and $\tau_N^{(2)}$ show no such strong bias. $\tau_N^{(2)}$ has a conspicuous advantage over $\tau_N^{(1)}$ only when $\Theta = 0.9$, which is the most difficult case for correct detection. Underestimation is more difficult to detect than overestimation in the situations considered.

Table 2. Percents of Simulated Airline Model Series with $\tilde{\theta} = \tilde{\Theta} = 0.6$ for Which Underestimation is Indicated for Irregulars Produced from Estimated θ and Θ or from Incorrect Θ

Θ	τ_N^{SEATS}	$\tau_N^{(1)}$	$\tau_N^{(2)}$
0.3	12.1	1.4	2.1
0.4	32.2	6.9	8.6
0.5	62.7	22.0	24.4
estimated θ, Θ	100.0	47.4	48.2
0.7	96.6	75.0	73.3
0.8	99.1	84.1	84.0
0.9	98.4	66.7	81.4

4. Tests for the Significance of Over- or Underestimation

For significance testing, we interpret the value of $\tau_N^{(1)}$ by reference to an estimate $\hat{\sigma}_N(\tau_N^{(1)})$ of its standard deviation given by the r.h.s. of (25) in the Appendix. An analogous $\hat{\sigma}_N(\tau_N^{(2)})$ is used with $\tau_N^{(2)}$. For a given size (significance level) α, and with z denoting an $N\,(0,1)$ variate, let $z_{1-\alpha}$ denote the value for which $P\{z > z_{1-\alpha}\} = P\{z < -z_{1-\alpha}\} = 1-\alpha$. We performed simulation experiments to determine, for various α and Θ, the proportion of simulated series for which

$$\tau_N^{(i)} > z_{1-\alpha}\hat{\sigma}_N(\tau_N^{(i)}), \qquad (15)$$

occurs, which is interpreted to indicate overestimation (at the α level of significance), or for which

$$\tau_N^{(i)} < -z_{1-\alpha}\hat{\sigma}_N(\tau_N^{(i)}) \qquad (16)$$

occurs, indicating underestimation, for $i = 1, 2$ for 1000 series of length $N = 144$ simulated from (8) with $\tilde{\theta} = \tilde{\Theta} = 0.6$.

Table 3. Specified vs. Observed Sizes of (15) and (16): $\theta = 0.6, \Theta = 0.4$

	proport. of (15)		proport. of (16)	
$\alpha \diagdown \tau$	$\tau_{144}^{(1)}$	$\tau_{144}^{(2)}$	$\tau_{144}^{(1)}$	$\tau_{144}^{(2)}$
.05	.432	.394	0	.001
.10	.559	.533	0	.003
.15	.662	.641	.003	.009
.20	.733	.713	.006	.013
.25	.795	.760	.010	.022

Table 4. Specified vs. Observed Sizes of (15) and (16): $\theta = 0.6, \Theta = 0.9$

	proport. of (15)		proport. of (16)	
$\alpha \diagdown \tau$	$\tau_{144}^{(1)}$	$\tau_{144}^{(2)}$	$\tau_{144}^{(1)}$	$\tau_{144}^{(2)}$
.05	.044	.004	.113	.242
.10	.025	.018	.178	.362
.15	.051	.030	.241	.442
.20	080	.052	.314	.523
.25	.119	.065	.390	.586

A minimal "power" requirement for a usable test would seem to be that the proportion of detections of the correct kind of misestimation should exceed 0.5. Our experiments showed that (15) easily satisfies this minimal power requirement for detecting moderately strong overestimation, $\Theta \leq 0.4$, for $i = 1, 2$ when $\alpha \geq 0.10$, see Table 3. However, for (16), and $\Theta \geq 0.8$, $\alpha \geq 0.20$ must be used, and for $\Theta = 0.9$, $\tau_N^{(2)}$ is required, see Table 4.

Our simulation results suggest that testing with $\alpha = .20$ will provide adequate to excellent power against a broad range of alternatives and that $\tau_N^{(2)}$ need only be used when Θ is rather close to one. We also produced results (not shown) for the sample size $N = 72$. For this sample size, testing with $\alpha = .25$, there is reasonable power for detecting overestimation with $0.3 \leq \Theta \leq 0.4$ but not adequate power for detecting underestimation.

5. Empirical Results

We now apply $\tau_N^{(1)}$ and $\tau_N^{(2)}$ to model-based adjustments of two series. The first is the series of dollar values of U.S. Exports of Other Agricultural Materials (Manufactured) from January, 1989-December, 2001 ($N = 156$), which we abbreviate as *Export*. Because of its heterogenous nature, its seasonal pattern is not very well defined and it is not seasonally adjusted by the Census Bureau. SEATS' seasonal adjustment has left some seasonality, as indicated by the one seasonal peak in the spectrum of the irregular component, at the highest seasonal frequency (0.5 cycles/month), see Fig.5. This is the only frequency at which the spectrum of the (differenced, logged) original series has a seasonal peak (not shown). The peak in Fig.5 is visually significant according to the criterion used by X-12-ARIMA; see Soukup and Findley (1999).

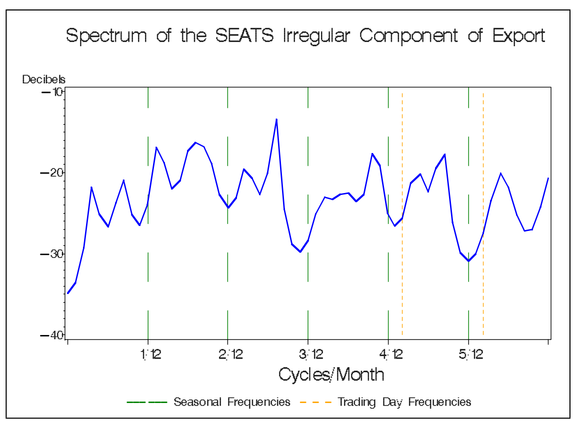

Figure 5: Spectrum of the Irregulars of *Export*. Note the seasonal peak at 6/12 cycles/month.

For $i = 1, 2$, let $p_N^{(i)}$ denote the probability that a standard normal variate z has a value at least as extreme as $\bar{\tau}_N^{(i)} = \tau_N^{(i)}/\hat{\sigma}_N(\tau_N^{(i)})$, i.e. $p_N^{(i)} = pr\left\{z \leq \bar{\tau}_N^{(i)}\right\}$ if $\bar{\tau}_N^{(i)} < 0$ and $p_N^{(i)} = pr\left\{z \geq \bar{\tau}_N^{(i)}\right\}$ if $\bar{\tau}_N^{(i)} > 0$. The first row of Table 5 provides the $\bar{\tau}_N^{(i)}$ and $p_N^{(i)}$ values for

Export. Both $\bar{\tau}_N^{(i)}$ values are negative, commensurate with underestimation, but the $p_N^{(i)}$ values show that only the test based on $\tau_N^{(2)}$ indicates significant underestimation. The fitted model is an airline model with both coefficients equal to 0.80 to this level of precision. The Box-Ljung statistics of the model residuals are poor (p-values below 0.05) at lags 3–6 but acceptable at higher lags (p-value of 0.289 at lag 24). There are no indications of skewness or kurtosis in the model residuals. This example confirms the utility of $\tau_N^{(2)}$.

SEATS' analogue of $\tau_N^{(1)}$, i.e. τ_N^{SEATS} defined in (14), has the value -0.033. SEATS provides no estimate of the standard error of τ_N^{SEATS} (Maravall, 2003 provides an approach to an estimate.)

Table 5. Values of $\bar{\tau}_N^{(i)}$ and $p_N^{(i)}$, $i = 1, 2$

	$\bar{\tau}_N^{(1)}$; $p_N^{(1)}$	$\bar{\tau}_N^{(2)}$; $p_N^{(2)}$
Export	-0.546; 0.292	-2.177; 0.015
Vol	0.299; 0.383	0.142; 0.443

Whereas the spectrum can call attention to the possibility of underestimation, the main source of evidence for oversmoothing, which we interpret to mean overestimation, has been the opinions of data experts. We denote by *Vol* the sales volume series for large department stores (Grands Magasins) from January, 1990 through March, 2004 ($N = 171$) produced by the Chamber of Commerce and of Industry of Paris (CCIP). Mr. J. Anas of CCIP communicated the concern of CCIP that the model-based adjustment of SEATS produced with TRAMO's automatically chosen (0 1 2)(0 1 0) model, and hence with $\Theta = 0$, might be oversmoothing. We used the outlier treatment preferred by CCIP, modeling a temporary change outlier for October, 1995 but not two other outliers indicated by the automatic outlier identification procedure. We were not able to obtain enough information about the holiday adjustment method used by CCIP for this series to able to replicate it, so we used instead the Easter and trading day effect modeling options of X-12-ARIMA, thereby ignoring several French holidays. In spite of this compromise, our model diagnostics were mostly good: the Box-Ljung Q statistics for the model residuals had p-values greater than 0.11 at all lags, with a value slightly greater than 0.90 at lag 24. There were no indications of skewness or kurtosis in the model residuals.

The values in the second row of Table 5 show support, albeit weak, for a diagnosis of overadjustment. The Table D 9.A diagnostics of the X-12-ARIMA adjustment of this series suggest that the seasonal adjustment of five and perhaps six of the calendar months should be done with a standard length filter (a 3×5 seasonal filter, which yields an adjustment similar to a SEATS adjustment with $\Theta = 0.6$, see Findley and Martin, 2003), whereas a shorter filter should be used for the remaining months, quite short for some of the months. (A SEATS filter from $\Theta = 0$ has length about 37 months, slightly shorter than the filter obtained by using the shortest (i.e., the 3×1) seasonal filter in X-12-ARIMA.) Thus, the indications of overestimation may be weak because overestimation is a problem only for half or less of the months. For this series, from the three digit rounded values of SEATS' output, $\tau_N^{SEATS} = 0.000$, giving no indication of overestimation.

5.1 A Simulation-Based Alternative to Table 5

The $p_N^{(i)}$ values presented in Table 5 were obtained by assuming that the $\bar{\tau}_N^{(i)}$ have a standard normal distribution. We also obtained simulation-based alternatives to these $p_N^{(i)}$ values to confirm the conclusions obtained from the $\bar{\tau}_N^{(i)}$ values of Table 5. We simulated 5000 Gaussian series of the appropriate length from each series' estimated model, of length $N = 156$ from *Export*'s estimated airline model and of length $N = 171$ from *Vol*'s estimated (0 1 2)(0 1 0) model, reestimating model parameters for each simulated series. From the irregular component obtained from the reestimated model for a given simulated series, we obtain an analogue of $\bar{\tau}_N^{(i)}$ which we denote by $\bar{\tau}_N^{*(i)}$. The four histograms (not shown) of 5000 $\bar{\tau}_N^{*(i)}$'s obtained for each model and for $i = 1, 2$ have means very close to zero, but are skewed for $i = 1$, and, for *Export*'s model, also for $i = 2$. For each histogram, let $\bar{p}_N^{(i)}$ denote the proportion of the 5000 $\bar{\tau}_N^{*(i)}$'s that are at least as extreme as the corresponding $\bar{\tau}_N^{(i)}$ value of Table 5. The $\bar{p}_N^{(i)}$ value associated with each $\bar{\tau}_N^{(i)}$ is shown in Table 6.

The simulation-based p-values of Table 6 have the advantages over those of Table 5 that they make allowance both for uncertainty arising from parameter estimation and for deviations from Gaussianity of the $\bar{\tau}_N$. The extent to which they support the conclusions drawn from Table 5 is quite reassuring. Both series represent challenging cases for the detection of misestimation.

Table 6. Values of $\bar{\tau}_N^{(i)}$ and $\bar{p}_N^{(i)}$, $i = 1, 2$

	$\bar{\tau}_N^{(1)}$; $\bar{p}_N^{(1)}$	$\bar{\tau}_N^{(2)}$; $\bar{p}_N^{(2)}$
Export	−0.546; 0.317	−2.177; 0.013
Vol	0.299; 0.297	0.142; 0.375

6. Appendix

6.1 Formulas for $\sigma^2_{a,mle,N}(\theta, \Theta)$ and $E^{true}\sigma^2_{a,mle,N}\left(\tilde{\theta}, \Theta\right)$

A time series with spectral density $\tilde{g}(\lambda)$ has autocovariances

$$\gamma_j(\tilde{g}) = \int_{-\pi}^{\pi} \cos j\lambda \tilde{g}(\lambda) d\lambda, j = 0, \pm 1, \ldots.$$

For any positive integer $n \geq 1$, define the autocovariance matrix $\Sigma_n(\tilde{g}) = \left[\gamma_{j-k}(\tilde{g})\right]_{0 \leq j,k \leq n-1}$. For the model (3), the spectral density of $y_t = \delta(B) Y_t$ is $g(\lambda) = (\sigma_a^2/2\pi) \left|\eta\left(e^{i\lambda}\right)\right|^2 \left|\phi\left(e^{i\lambda}\right)\right|^{-2}$. If we define $\mathbf{y}_{d+1:N} = \left[\begin{array}{ccc} y_{d+1} & \cdots & y_N \end{array}\right]'$ and $g_1(\lambda) = (1/2\pi) \left|\eta\left(e^{i\lambda}\right)\right|^2 \left|\phi\left(e^{i\lambda}\right)\right|^{-2}$, then it follows from $\Sigma_{N-d}(g) = \sigma_a^2 \Sigma_{N-d}(g_1)$ and (3.2) of Ansley and Newbold (1981) that the m.l.e. of σ_a^2 is

$$\sigma^2_{a,mle,N}(\eta/\phi) = \frac{1}{N-d} \mathbf{y}'_{d+1:N} \Sigma^{-1}_{N-d}(g_1) \mathbf{y}_{d+1:N}, \quad (17)$$

with $\mathbf{y}'_{d+1:N}$ denoting the transpose of $\mathbf{y}_{d+1:N}$, and that, with tr denoting trace of a matrix,

$$\begin{aligned} & E^{true}\sigma^2_{a,mle,N}(\eta/\phi) \\ = & \frac{1}{N-d} tr\left\{\Sigma^{-1}_{N-d}(g_1) \Sigma_{N-d}(\tilde{g})\right\}. \end{aligned} \quad (18)$$

For (9), we denote (17) by $\sigma^2_{a,mle,N}(\theta, \Theta)$.

6.2 Formulas associated with $\hat{I}_t$

To obtain standard errors of test statistics, variances and covariances of $\overline{\hat{I}^2}$ and $\sigma^2_{a,mle,N}(\eta/\phi)$ are required. These can be obtained fairly easily if, as usual, the model coefficients are treated as fixed rather than random. We first need the covariance matrix of the $\hat{I}_t$. In the notation introduced above, and with Δ_N denoting the $(N-d) \times N$ band matrix of the form

$$\Delta_N = \left[\begin{array}{ccccccc} -\delta_d & \cdots & -\delta_1 & 1 & & & \\ & \ddots & \ddots & \ddots & \ddots & & \\ & & -\delta_d & \cdots & -\delta_1 & 1 \end{array}\right], \quad (19)$$

the formula for the vector $\hat{I} = \hat{I}(g, N)$ of estimates $\hat{I}_t, 1 \leq t \leq N$ from (2) is

$$\hat{I} = \frac{\sigma_I^2}{\sigma_a^2} \Delta'_N \Sigma^{-1}_{N-d}(g_1) \mathbf{y}_{d+1:N}, \quad (20)$$

when I_t is white noise with variance σ_I^2, see (4.3) of Bell and Hillmer (1988). Thus the covariance matrix $\Sigma^{true}_{\hat{I}} = \Sigma_{\hat{I}}(\tilde{g}, g, N)$ has the formula

$$\Sigma^{true}_{\hat{I}} = \left(\frac{\sigma_I^2}{\sigma_a^2}\right)^2 \Delta'_N \Sigma^{-1}_{N-d}(g_1) \Sigma_{N-d}(\tilde{g}) \Sigma^{-1}_{N-d}(g_1) \Delta_N. \quad (21)$$

For $1 \leq t \leq N$, $E^{true}\hat{I}_t^2$ is the t-th diagonal entry of $\Sigma^{true}_{\hat{I}}$, and

$$E^{true}\overline{\hat{I}^2} = N^{-1} tr \Sigma^{true}_{\hat{I}}. \quad (22)$$

For $1 \leq t \leq N$, the scaled model-based variance σ_t^2/σ_a^2 is the t-th diagonal entry of

$$\Sigma_{\hat{I}/\sigma_a} = \left(\frac{\sigma_I^2}{\sigma_a^2}\right)^2 \Delta'_N \Sigma^{-1}_{N-d}(g_1) \Delta_N, \quad (23)$$

and (12) is given by $N^{-1} tr \Sigma_{\hat{I}/\sigma_a}$.

6.3 Test statistic standard errors

We now derive formulas for the variances of $\overline{\hat{I}^2}$ and $\sigma^2_{a,mle,N}(\eta/\phi)$ and their covariance under the assumptions that the model spectral density $g(\lambda)$ is *correct*, $\tilde{g}(\lambda) = g(\lambda) = \sigma_a^2 g_1(\lambda)$, and that the y_t have no skewness or kurtosis. Using the special case of the formula of McCullagh (1987, p. 65) for the covariance of two symmetric quadratic forms of zero mean variates X_i whose joint third and fourth cumulants are zero, we obtain

$$\begin{aligned} var\left(\overline{\hat{I}^2}\right) &= \frac{2}{N^2} \sum_{j,k=1}^{N} cov^2\left(\hat{I}_j, \hat{I}_k\right) \\ &= \frac{2\sigma_a^4}{N^2} tr\left\{\Sigma^2_{\hat{I}/\sigma_a}\right\}, \end{aligned}$$

$$var\left(\sigma^2_{a,mle,N}(\eta/\phi)\right) = \frac{2\sigma_a^4}{N-d},$$

and

$$cov\left(\overline{\hat{I}^2}, \sigma^2_{a,mle,N}\left(\eta/\phi\right)\right) = \frac{2\sigma^4_a}{N\left(N-d\right)} tr\Sigma_{\hat{I}/\sigma_a}.$$

with $\Sigma_{\hat{I}/\sigma_a}$ given by (23).

Writing $\hat{\sigma}^2_a = k_N \sigma^2_{a,mle,N}\left(\eta/\phi\right)$, see (5),

$$\tau_N = \overline{\hat{I}^2} - k_N \sigma^2_{a,mle,N}\left(\eta/\phi\right)\left\{N^{-1} tr\Sigma_{\hat{I}/\sigma_a}\right\} \tag{24}$$

has the variance

$$\frac{2\sigma^4_a}{N^2}\left[tr\left\{\Sigma^2_{\hat{I}/\sigma_a}\right\} - \frac{2k_N - k^2_N}{(N-d)}\left(tr\Sigma_{\hat{I}/\sigma_a}\right)^2\right].$$

For an approximate standard error for the test statistic $\tau^{(1)}_N$, we replace the unknown σ^4_a with $\hat{\sigma}^4_a$ from (5) and take the positive square root to obtain

$$\hat{\sigma}_N(\tau^{(1)}_N) =$$

$$\frac{2^{1/2}\hat{\sigma}^2_a}{N}\left[tr\left\{\Sigma^2_{\hat{I}/\sigma_a}\right\} - \frac{2k_N - k^2_N}{(N-d)}\left(tr\Sigma_{\hat{I}/\sigma_a}\right)^2\right]^{1/2}. \tag{25}$$

For the test statistic $\tau^{(2)}_N$, the formula (25) is modified by leaving $(N-d)$ unchanged but otherwise replacing N elsewhere with $(N-24)$ and by replacing the matrix Δ_N in (23) with the matrix $\Delta^{(2)}_N$ which differs from Δ_N only in its first and last twelve columns by having all entries in these columns equal to 0.

The formulas needed to implement the analogues of the test statistic (13) for the stationary transforms of the seasonal and trend component estimates are given in Proposition 4 of McElroy and Sutcliffe, 2004).

Acknowledgement. We thank William Bell and Michael Shimberg for helpful comments on drafts of this paper and Catherine Hood for supplying the *Export* series.

REFERENCES

Ansley, C. F. and Newbold, P. (1981), "On the Bias in Estimates of Forecast Mean Square Error," *Journal of the American Statistical Association*, **76**, 569-576.

Findley, D. F., Wills, K. C., Aston, J. A., Feldpausch, R. M. and Hood, C. C. (2003), "Diagnostics for ARIMA-Model-Based Seasonal Adjustment," *2003 Proceedings American Statistical Association* [CD-ROM]: Alexandria, VA

Bell, W. R. and Hillmer, S. C. (1988), "A matrix approach to likelihood evaluation and signal extraction for ARIMA component time series models," Bureau of the Census, Statistical Research Division Report Number: Census/SRD/RR-88/22. Also www.census.gov/srd/papers/pdf/rr88-22.pdf

Findley, D. F. and Martin, D. E. K. (2003), "Frequency Domain Analyses of SEATS and X-12-ARIMA Seasonal Adjustment Filters for Short and Moderate-Length Time Series," Statistical Research Division Research Report Series, *Statistics* #2003-02, Census Bureau, Washington, D.C. Also www.census.gov/srd/papers/pdf/rrs2003-02.pdf

Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C. and Chen, B. C. (1998). New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program," *Journal of Business and Economic Statistics* **16**, 127-177 (with discussion).

Gómez, V. and Maravall, A. (1997), "Program SEATS Signal Extraction in ARIMA Time Series: Instructions for the User (beta version: June 1997)," Working Paper 97001, Ministerio de Economía y Hacienda, Dirrectión General de Análysis y Programación Presupuestaria, Madrid.

Maravall, A. (2003), "A Class of Diagnostics in the ARIMA-Model-Based Decomposition of a Time Series," Memorandum, Bank of Spain.

McCullagh, P. (1987), *Tensor Methods in Statistics,* London: Chapman and Hall.

McElroy, T. S. and Sutcliffe, A. (2004), "An Iterated Parametric Approach to Nonstationary Signal Extraction," Statistical Research Division Research Report Series, *Statistics* #2004-05, Census Bureau, Washington, D.C. Also www.census.gov/srd/papers/pdf/rrs2004-05.pdf

Soukup, R. J. and Findley, D. F. (1999). "On the Spectrum Diagnostics Used by X-12-ARIMA to Indicate the Presence of Trading Day Effects after Modeling or Adjustment," *Proceedings of the Business and Economic Statistics Section*, 144–149, American Statistical Association. Also www.census.gov/pub/ts/papers/rr9903s.pdf