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

March 23, 2005

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 diagnostic of SEATS depends on variance estimates that assume an infinitely long filter has been applied. This results in substantial bias toward indicating underestimation. Our modified diagnostics are calculated from time-varying variances associated with the finite-length filters actually used. Tests for the statistical significance of any indicated misestimation are presented and analyzed. Our diagnostics and tests also apply to structural model-based approaches to seasonal decomposition.

Key Words. Signal Extraction, Auto Regressive Integrated Moving Average model, Wiener-Kolmogorov Filtering.

Disclaimer. This report is distributed to inform about ongoing research and to encourage discussion. Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

1 Introduction

The most widely used ARIMA model-based approach to seasonal decomposition of a time series is the "canonical" decomposition approach of Hillmer and Tiao (1982) and Burman (1980), as implemented in SEATS (Gómez and Maravall, 1997) and in other software under

development – see Monsell, Aston and Koopman (2003). There are also the "structural" ARIMA model-based approaches of DECOMP (Kitagawa 1981, 1985) and STAMP (Koopman, Harvey, Doornik, and Shepherd, 1995). All of these approaches assume that, after a logarithmic transformation if necessary, and after removal of any trading day, holiday, and outlier effects estimated with the regression component of the fitted regARIMA model, the time series to be seasonally adjusted, Y_t, $1 \leq t \leq N$, can be decomposed into a sum of ARIMA seasonal, trend, and irregular components denoted by S_t, T_t, and I_t respectively,

$$Y_t = S_t + T_t + I_t, \ 1 \leq t \leq N. \tag{1.1}$$

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

$$\hat{I}_t = E\left(I_t | Y_s, 1 \leq s \leq N\right) = \sum_{j=t-N}^{t-1} c_{j,t}^I\left(N\right) Y_{t-j}. \tag{1.2}$$

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 – specified or estimated from the data via maximum likelihood – as if they were correct. $\hat{S}_t$ and $\hat{T}_t$ are defined analogously. (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. The additivity property of conditional means yields the seasonal decomposition $Y_t = \hat{S}_t + \hat{T}_t + \hat{I}_t$, $1 \leq t \leq N$. 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, Monsell, Bell, Otto, and Chen (1998).

SEATS (Gómez and Maravall, 1997) does not yet have spectrum diagnostics for detecting residual seasonality. Instead, it provides a diagnostic for detecting "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, while 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 signal extraction filters used do not adequately suppress the other component or components. This would 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, e.g., 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 Figure 1 below. There are no formal studies of the consequences of overestimation and underestimation, but Table 10 of Findley, Monsell, Shulman, and Pugh (1990) indicates that both can greatly increase the amount of seasonal adjustment instability observed when more recent observations are added to the time series and the earliest observations are deleted.

In Section 2, after describing the SEATS diagnostic for detecting underestimation and overestimation, we show it is biased toward identifying underestimation even when the estimate is optimal. Simulation analyses of two modified diagnostics that are essentially unbiased are provided in Section 3. We develop significance tests for the modified diagnostics in Section 4 and in the Appendix, for the situation in which the distribution of the stationary time series obtained from differencing operations has neither skewness nor kurtosis. (Specifically, we assume that all third and fourth cumulants are zero.) In Section 5, the significance tests are applied with some success to two official time series for which misestimation is suspected. Section 6 discusses some conclusions and extensions. Certain required formulas are derived in the Appendix.

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, as are the formulas of the over/underestimation diagnostics. (For the seasonal, trend, and

3

seasonally adjusted components, the SEATS diagnostic is calculated for the "stationary transformation" of each component, meaning the output of the differencing operation specified by the component's ARIMA model.) Because of this greater simplicity, Agustín Maravall attaches more importance to the over/underestimation diagnostic of the irregulars than to the corresponding diagnostics for the other components (personal communication). Our study seems to be the first systematic investigation of this kind of diagnostic.

In SEATS, the term *estimator* is used for the theoretical Wiener-Kolmogorov estimate that applies with bi-infinite data. To have a specific formula for the *variance of the estimator* of the irregular component used by SEATS, suppose the (estimated or fixed) ARIMA model for the Y_t is written in the usual backshift operator polynomial notation as

$$\delta(B)\phi(B)Y_t = \eta(B)a_t. \tag{2.1}$$

Thus $\delta(B) = 1 - \delta_1 B - \cdots - \delta_d B^d$ denotes the differencing operator which transforms Y_t to stationarity, e.g. $\delta(B) = (1-B)^2(1+B+\cdots+B^{s-1})$, 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 $\eta(z)$. The spectral density of the model for the differenced series $y_t = \delta(B)Y_t$ is

$$g(\lambda) = \frac{\sigma_a^2}{2\pi}\left|\frac{\theta(e^{-i\lambda})}{\phi(e^{-i\lambda})}\right|^2. \tag{2.2}$$

For $j = 0, \pm 1, \cdots$ with

$$c_j^I = \frac{\sigma_I^2}{4\pi^2}\int_{-\pi}^{\pi}\frac{\cos j\lambda}{g(\lambda)}\left|\delta(e^{-i\lambda})\right|^2 d\lambda \tag{2.3}$$

the Wiener-Kolmogorov *estimator* $I_{WK,t}$ of I_t from bi-infinite data, i.e., the Gaussian conditional expectation $E(I_t|Y_s, -\infty < s < \infty)$, has the formula $I_{WK,t} = \sum_{j=-\infty}^{\infty}c_j^I Y_{t-j}$ SEATS' *variance of the estimator*, $\sigma_{WK,I}^2 = EI_{WK,t}^2$, has the formula $\sigma_{WK,I}^2 = \sigma_I^2 c_0^I$. The *estimate* of I_t is the model's conditional mean from the data:

$$\hat{I}_t = E(I_t|Y_s, 1 \leq s \leq N) = \sum_{j=t-N}^{t-1}c_{j,t}^I Y_{t-j}. \tag{2.4}$$

SEATS' *variance of the estimate* is defined to be the sample second moment[1]

$$\overline{\hat{I}^2} = \frac{1}{N} \sum_{t=1}^{N} \hat{I}_t^2, \tag{2.5}$$

and, conceptually for SEATS, *overestimation* is indicated when $\overline{\hat{I}^2} > \sigma_{WK,I}^2$ and *underestimation* when $\overline{\hat{I}^2} < \sigma_{WK,I}^2$. However, the calculations that produce the component models in SEATS and other software yield variances such as σ_I^2 and $\sigma_{WK,I}^2$ calculated as though the innovation variance of (2.1) were equal to one. We shall denote these unscaled variances by σ_I^2/σ_a^2 and $\sigma_{WK,I}^2/\sigma_a^2$. Thus, in place of $\sigma_{WK,I}^2$, for some estimate $\hat{\sigma}_a^2$, the scaled quantity $\hat{\sigma}_a^2 \sigma_{WK,I}^2/\sigma_a^2$ is used. Let $\sigma_{a,mle}^2$ denote the maximum likelihood estimate of σ_a^2 given by (A.1) in the Appendix. Following SEATS, we use the bias-corrected estimate of Ansley and Newbold (1981),

$$\hat{\sigma}_a^2 = \frac{N - n_{\delta,\phi}}{N - n_{\delta,\phi} - n_{coeffs}} \sigma_{a,mle}^2 \tag{2.6}$$

where $n_{\delta,\phi} = d$ (or d plus the degree of $\phi(z)$ if SEATS' conditional estimate of $\phi(z)$ is used), and n_{coeffs} is the number of estimated ARMA coefficients in the model. Note that when $n_{coeffs} = 0$, e.g. in the fixed-coefficient case, then $\hat{\sigma}_a^2 = \sigma_{a,mle}^2$. SEATS' criterion is $\overline{\hat{I}^2} > \hat{\sigma}_a^2 \sigma_{WK,I}^2/\sigma_a^2$ for overestimation, and $\overline{\hat{I}^2} < \hat{\sigma}_a^2 \sigma_{WK,I}^2/\sigma_a^2$ for underestimation.

The following simple Proposition shows that SEATS' over/underestimation criterion is biased toward indicating underestimation.

Proposition 1 *For the estimates $\hat{I}_t$ (2.4) of the stationary irregular component I_t, we have*

$$E\overline{\hat{I}^2} \leq \sigma_{WK,I}^2 \tag{2.7}$$

with $\sigma_{WK,I}^2 - E\overline{\hat{I}^2} = N^{-1} \sum_{t=1}^{N} \left(E(I_t - \hat{I}_t)^2 - E(I_t - I_{WK,t})^2 \right)$. Strict inequality necessarily holds in (2.7) when the moving average polynomial $\eta(B)$ in (2.1) has positive degree.

Proof. For covariance calculations based on the assumed model, the estimate $\hat{I}_t$ and its error $I_t - \hat{I}_t$, which sum to I_t, are treated as uncorrelated. The same is true for $I_{WK,t}$ and $I_t - I_{WK,t}$. Thus we have

$$\sigma_I^2 = E\hat{I}_t^2 + E(I_t - \hat{I}_t)^2 = \sigma_{WK,I}^2 + E(I_t - I_{WK,t})^2. \tag{2.8}$$

---

[1]Instead of $\overline{\hat{I}^2}$, SEATS uses the sample variance, $\sum_t (\hat{I}_t - \overline{\hat{I}_t})^2/(N-1)$ where $\overline{\hat{I}_t}$ is the sample mean. This is $N/(N-1) \cong 1$ times the smaller quantity $\overline{\hat{I}^2} - (\overline{\hat{I}_t})^2$. Therefore, the bias results of Proposition 1 apply to the sample variance as well as to the sample second moment.

It follows from (2.8) that (2.7) is valid, because $E(I_t - I_{WK,t})^2 \le E(I_t - \hat{I}_t)^2$ due to $E(I_t - \hat{I}_t)^2 = E(I_t - I_{WK,t})^2 + E(\hat{I}_t - I_{WK,t})^2$. This identity holds because, for covariance calculations based on the assumed model, $I_{WK,t}$, being the minimum mean square estimator from the larger data set, is treated as uncorrelated with $\hat{I}_t - I_{WK,t}$. Strict inequality holds in (2.7) whenever (2.1) has a moving average component, because then infinitely many coefficients in (2.3) are nonzero, so that $E(\hat{I}_t - I_{WK,t})^2 > 0$. $\square$

Suppose the model considered is the Box-Jenkins airline model,

$$(1 - B)(1 - B^{12})Y_t = (1 - \theta B)(1 - \Theta B)a_t. \tag{2.9}$$

Let $\sigma^2_{WK,I}$ denote the variance of the canonical irregular component of the sort produced by SEATS for this model. For the case $\theta = 0.6$, Table 1 shows values of the *relative bias* $\sigma^2_{WK,I}/E\overline{\hat{I}^2}$ for sample sizes $N = 72, 144$ and various Θ. The relative bias values are quite similar for other values of θ with $0.1 \le \theta \le 0.9$ (not shown).

Table 1. Relative Bias $\sigma^2_{WK,I}/E\overline{\hat{I}^2}$ for Various Θ ($\theta = 0.6$) and N

Θ	$N = 72$	$N = 144$
0.1	1.1900	1.0875
0.2	1.1726	1.0795
0.3	1.1588	1.0736
0.4	1.1462	1.0685
0.5	1.1365	1.0639
0.6	1.1293	1.0599
0.7	1.1274	1.0563
0.8	1.1363	1.0546
0.9	1.1633	1.0614

3 Modification of SEATS' Diagnostic

The above results, documenting substantial bias, suggest that in place of the constant variance $\sigma^2_{WK,I}$ that applies to the bi-infinite case, the average of the time-varying variances $\sigma^2_t = E\hat{I}^2_t$ should be used, calculated as though the estimated or specified model were correct. That is, $\overline{\hat{I}^2}$ should be compared to $N^{-1}\sum_{t=1}^{N}\sigma^2_t$ instead of to $\sigma^2_{WK,I}$. However, the values usually available for σ^2_t are calculated as though the estimated or specified model for Y_t had

innovation variance one, so the values must be scaled by a factor $\hat{\sigma}_a^2$. We will use σ_t^2/σ_a^2 to denote a model-specified variance of I_t that requires scaling; σ_t^2/σ_a^2 can be obtained directly from (A.7). Thus, the role of $N^{-1}\sum_{t=1}^{N}\sigma_t^2$ is replaced by

$$\frac{1}{N}\sum_{t=1}^{N}\frac{\sigma_t^2}{\sigma_a^2}, \tag{3.1}$$

multiplied by the scaling factor $\hat{\sigma}_a^2$. That is,

$$\overline{\hat{I}^2} > \frac{\hat{\sigma}_a^2}{N}\sum_{t=1}^{N}\frac{\sigma_t^2}{\sigma_a^2} \tag{3.2}$$

indicates overestimation and

$$\overline{\hat{I}^2} < \frac{\hat{\sigma}_a^2}{N}\sum_{t=1}^{N}\frac{\sigma_t^2}{\sigma_a^2} \tag{3.3}$$

indicates underestimation.

3.1 Mean performance of (3.2) and (3.3)

To obtain a "proof of concept" of our modification (3.2) and (3.3) of SEATS' diagnostic, we compare the means of their left and right hand sides for zero mean Gaussian data Y_t satisfying (3.4) with $\tilde{\theta} = 0.6 = \tilde{\Theta}$,

$$(1-B)(1-B^{12})Y_t = \left(1 - \tilde{\theta}B\right)\left(1 - \tilde{\Theta}B\right)\tilde{a}_t, \tag{3.4}$$

where the variance of $\tilde{a}_t$ is $\sigma_{\tilde{a}}^2$. We consider irregular component estimates from models 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 $\tilde{\Theta}$. The mean of the left hand side, $E^{true}\overline{\hat{I}^2}$, is given by (A.6) in the Appendix. With $\sigma_{a,mle}^2(\theta, \Theta)$ denoting the value of $\sigma_{a,mle}^2$ (see (A.1)) specified by (2.9), the right hand side mean is the product

$$\left\{\frac{1}{N}\sum_{t=1}^{N}\frac{\sigma_t^2}{\sigma_a^2}\right\}E^{true}\sigma_{a,mle}^2(\theta, \Theta) \tag{3.5}$$

which can be calculated from (A.2) and (A.7) of the Appendix. Values of the ratio of $E^{true}\overline{\hat{I}^2}$ divided by (3.5) are given in Table 2. Graphs of the squared gains $\left|\sum_{j=t-N}^{t-1}c_{j,t}^{I}(N)e^{i2\pi j\lambda}\right|^2$, $-0.5 \leq \lambda \leq 0.5$ from SEATS' decompositions of (2.9) show that overestimation occurs when $\Theta < \tilde{\Theta}$ and underestimation occurs when $\Theta > \tilde{\Theta}$; see Figure 1 for the case $t = [N/2] + 1$ (midpoint) and $\tilde{\Theta} = 0.6$.

Table 2 shows that $\Theta < 0.6$ leads to ratio values greater than one and therefore to (3.2) on average, whereas $\Theta > 0.6$ leads to ratio values less than one, hence to (3.3) on average. Thus, on average, (3.2) and (3.3) 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, from $\Theta = 0.9$, is more difficult to detect with this approach than the less extreme underestimation from $\Theta = 0.7$ or 0.8.

Table 2. True Means of $\overline{\hat{I}^2}$ from Series with $\tilde{\theta} = \tilde{\Theta} = 0.6$

Compared to (3.5) for Various Θ

Θ	$E^{true}\overline{\hat{I}^2}$	(3.5)	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 2 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

$$\bar{\sigma}_t^2 = \left\{ \sigma_t^2 / \sigma_a^2 \right\} E^{true} \sigma_{a,mle}^2 (\theta, \Theta) \tag{3.6}$$

when there is underestimation and greater than $\bar{\sigma}_t^2$ when there is overestimation? We found for the models of Table 2 that such a local property holds only away from the ends of the series. In the first and final years and sometimes further in, particularly with underestimation, the inequality goes in the opposite direction. Thus, when there is underestimation, the values of $E^{true}\hat{I}_t^2$ are consistently *larger* than $\bar{\sigma}_t^2$ near the ends of the series. Figures 2 – 8 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 $E^{true}\hat{I}_t^2 > \bar{\sigma}_t^2$ holds are substantially wider in Figure 8 for the case $\Theta = 0.9$ than in Figure 7 for $\Theta = 0.8$.

For sample sizes not too much smaller than $N = 144$, these results suggest that the sum in (2.5) and the sum

$$\frac{1}{N} \sum_{t=1}^{N} \frac{\sigma_t^2}{\sigma_a^2}, \tag{3.7}$$

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

$$\tau_N^{(1)} = \overline{\hat{I}^2} - \frac{\hat{\sigma}_a^2}{N} \sum_{t=1}^{N} \frac{\sigma_t^2}{\sigma_a^2} \tag{3.8}$$

being the statistic based on (2.5) and (3.7), we define $\tau_N^{(2)}$ to be analogous statistic obtained by restricting the sums in this way.

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,

$$\tau_N^{SEATS} = \overline{\hat{I}^2} - \hat{\sigma}_a^2 \left(\sigma_{WK,I}^2 / \sigma_a^2 \right) \tag{3.9}$$

are correct. 5000 independent realizations of (3.4) with $\tilde{\theta} = \tilde{\Theta} = 0.6$ of length 144 were obtained from simulated standard normal innovations $\tilde{a}_t$. Always using $\theta = 0.6$, overestimated irregulars were obtained from SEATS by specifying decomposition via a model (2.9) with $\Theta < 0.6$, underestimated irregulars by specifying $\Theta > 0.6$. Table 3 lists the underestimation percents. The SEATS diagnostic has a strong bias toward incorrectly indicating underestimation when $0.4 \leq \Theta \leq 0.6$. Thus 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. Thus, as suggested by Table 1, underestimation is more difficult to detect than overestimation in the situations considered.

9

Table 3. Percents of Simulated Airline Model Series with $\tilde{\theta} = \tilde{\Theta} = 0.6$ for Which Underestimation is Indicated by Diagnostics of Adjustments Produced from Estimated θ and Θ and with Specified Incorrect Θ's

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	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 right hand side of (A.9) in the Appendix. This estimate accounts for the variances of $\overline{\hat{I}^2}$ and $\hat{\sigma}_a^2$ and their covariance by treating the model coefficients as non-random, as they are for the simulation results. 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 $\mathcal{N}(0,1)$ 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

$$\tau_N^{(i)} > z_{1-\alpha}\,\hat{\sigma}_N(\tau_N^{(i)}) \tag{4.1}$$

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

$$\tau_N^{(i)} < -z_{1-\alpha}\,\hat{\sigma}_N(\tau_N^{(i)}) \tag{4.2}$$

occurs, indicating underestimation, for $i = 1, 2$; we simulated 1000 series of length $N = 144$ from the model (3.4) with $\tilde{\theta} = \tilde{\Theta} = 0.6$.

Table 4. Specified vs. Observed Sizes of (4.1) and (4.2): $\theta = 0.6, \Theta = 0.4$

$\alpha \backslash \tau$	proportion of (4.1)		proportion of (4.2)	
	$\tau_{144}^{(1)}$	$\tau_{144}^{(2)}$	$\tau_{144}^{(1)}$	$\tau_{144}^{(2)}$
.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 5. Specified vs. Observed Sizes of (4.1) and (4.2): $\theta = 0.6, \Theta = 0.9$

$\alpha \backslash \tau$	proportion of (4.1)		proportion of (4.2)	
	$\tau_{144}^{(1)}$	$\tau_{144}^{(2)}$	$\tau_{144}^{(1)}$	$\tau_{144}^{(2)}$
.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

Table 6. Specified vs. Observed Sizes of (4.1) and (4.2): $\theta = 0.6, \Theta = 0.6$

$\alpha \backslash \tau$	proportion of(4.1)		proportion of (4.2)	
	$\tau_{144}^{(1)}$	$\tau_{144}^{(2)}$	$\tau_{144}^{(1)}$	$\tau_{144}^{(2)}$
.05	.045	.049	.041	.052
.10	.101	.093	.090	.103
.15	.137	.138	.145	.152
.20	.193	.193	.186	.197
.25	.242	.252	.235	.244

Tables 4 and 5 show power properties for detecting overestimation and underestimation respectively. 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 (4.1) 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 4. However, for (4.2) and $\Theta \geq 0.8$, $\alpha \geq 0.20$ must be used, and for $\Theta = 0.9$, $\tau_N^{(2)}$ is required; see Table 5.

11

Table 6 indicates that, when the filter is obtained from the true model, each ratio $\bar{\tau}_N^{(i)} = \tau_N^{(i)}/\hat{\sigma}_N(\tau_N^{(i)})$ for $i = 1, 2$ behaves approximately like a standard normal variate. Moreover, standard tests applied to the histograms of the $\bar{\tau}_N^{(1)}$ and $\bar{\tau}_N^{(2)}$ values indicate that the distributions associated with Table 6 are symmetric with mean zero, which further motivates the use we make of (4.1) and (4.2) in the next Section for significance testing. However, as discussed below, the histograms of $\bar{\tau}_N^{(1)}$ and $\bar{\tau}_N^{(2)}$ can be skewed when parameters are estimated for each simulated series rather than being fixed.

For the other values of Θ, the histograms have sample means and sample skewness coefficients that are, in a statistically significant sense, positive for overestimation and negative for underestimation. The mean seems to contribute more to the power to detect misestimation than does the skewness. These 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

5.1   Application to Two Time Series

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 Figure 9, the only frequency at which the spectrum of the (differenced, logged) original series has a seasonal peak (not shown). The peak in Figure 9 is visually significant according to the criterion used by X-12-ARIMA; see Soukup and Findley (1999).

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. $p_N^{(i)} = Prob\left\{z \leq \bar{\tau}_N^{(i)}\right\}$ if $\bar{\tau}_N^{(i)} < 0$ and $p_N^{(i)} = Prob\left\{z \geq \bar{\tau}_N^{(i)}\right\}$ if $\bar{\tau}_N^{(i)} > 0$. The first row of Table 7 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 (3.9), 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 7. 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 $(012)(010)$ 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 7 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, τ_N^{SEATS} has the value 0.000 and so gives no indication of overestimation.

5.2 A Simulation-Based Alternative to Table 7

The $p_N^{(i)}$ values presented in Table 7 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 7. 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 $(012)(010)$ 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 7. The $\bar{p}_N^{(i)}$ value associated with each $\bar{\tau}_N^{(i)}$ is shown in Table 8.

The simulation-based p-values of Table 8 have the advantages over those of Table 7 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 7 is quite reassuring. Both series represent challenging cases for the detection of misestimation.

Table 8. 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

From the results of Tables 3–6, we are encouraged by the performance of the $\bar{\tau}_N$ statistics and recommend that they replace the over and underestimation diagnostics currently in SEATS.

6   Conclusions and Extensions

By using finite-data signal extraction variances (in conjunction with the bias-corrected estimate of Ansley and Newbold (1981) of the innovation variance), one can obtain essentially unbiased analogues of the diagnostics of SEATS for detecting over- or underestimation and, in addition, test statistics for one-sided tests of the presence of these phenomena. In the preceding Sections, this was demonstrated in the context of an extensive analysis of the diagnostic for the irregular component of a seasonal adjustment decomposition. However, there are broad generalizations of the diagnostic (3.8) and its standard error, with slightly more complex formulas, that apply to time series which can be modeled as a sum of "independent" components, each with an ARIMA model. Such formulas are given in (A.15) and (A.16) of Appendix A for stationary transforms $\hat{u}_t = \delta^U(B)\hat{U}_t$ of the estimates $\hat{U}_t$ of the component U_t of a general two component decomposition $Y_t = U_t + V_t, 1 \le t \le N$. By specializing these formulas to the other components of a seasonal decomposition $Y_t = S_t + T_t + I_t$, i.e., setting U_t equal to S_t, T_t, or $T_t + I_t$, (the target seasonal adjustment), analogues of (4.1) and (4.2) can be obtained to provide tests for replacements for the other over- and under-adjustment diagnostics of SEATS. The exact formulas needed to implement the analogues of the test statistic (3.8) for the stationary transforms of the seasonal and trend component estimates are given in Proposition 4 of McElroy and Sutcliffe (2004). A change in the notation of the proof of Proposition 1 shows that SEATS' diagnostic for the stationary transforms of the seasonal, trend, and seasonally adjusted series are also biased toward indicating underestimation. Feldpausch, Hood, and Wills (2004) present simulation results that show this bias.

In future work, after these tests are programmed into the X-13 program discussed in Monsell et al. (2003), we plan to apply them to large numbers of Census Bureau time series to learn more about what features of series and seasonal decompositions they are sensitive to. Another future project is the establishment of a central limit theorem for $\tau_N^{(1)}/\hat{\sigma}_N(\tau_N^{(1)})$ and its generalizations.

A   Appendix: Technical Results

A.1   Formulas for $\sigma^2_{a,mle}(\theta, \Theta)$ and $E^{true}\sigma^2_{a,mle}\left(\tilde{\theta}, \Theta\right)$

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

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

Let $\Sigma(\tilde{g})$ denote the associated covariance matrix of order $N-d$, i.e., $\Sigma(\tilde{g}) = [\gamma_{j-k}(\tilde{g})]_{1 \le j,k \le N-d}$.
For the model (2.1), the spectral density of $y_t = \delta(B) Y_t$ is $g(\lambda) = (\sigma_a^2/2\pi) |\eta(e^{-i\lambda})|^2 |\phi(e^{-i\lambda})|^{-2}$.
If we define $\mathbf{y}_{d+1:N} = \begin{bmatrix} y_{d+1} & \cdots & y_N \end{bmatrix}'$ and $g_1(\lambda) = (1/2\pi) |\eta(e^{-i\lambda})|^2 |\phi(e^{-i\lambda})|^{-2}$, 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

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

and also that, with tr denoting trace of a matrix,

$$E^{true}\sigma_{a,mle}^2(\eta/\phi) = \frac{1}{N-d} tr\left\{\Sigma^{-1}(g_1)\Sigma(\tilde{g})\right\}. \tag{A.2}$$

For (2.9) and (3.5), we denote (A.1) by $\sigma_{a,mle}^2(\theta,\Theta)$ and (A.2) by $E^{true}\sigma_{a,mle}^2(\theta,\Theta)$.

A.2   Formulas associated with $\hat{I}_t$

To obtain standard errors of test statistics, variances and covariances of $\overline{\hat{I}^2}$ and $\sigma_{a,mle}^2(\eta/\phi)$
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 denoting the $N-d \times N$ band matrix of the form

$$\Delta = \begin{bmatrix} -\delta_d & \cdots & -\delta_1 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & -\delta_d & \cdots & -\delta_1 & 1 \end{bmatrix}, \tag{A.3}$$

the formula for the vector $\hat{I} = \hat{I}(g, N)$ of estimates $\hat{I}_t, 1 \le t \le N$ from (2.3) under Assumption
A of Bell and Hillmer (1988) – see subsection $A.4$ – is

$$\hat{I} = \frac{\sigma_I^2}{\sigma_a^2} \Delta'\Sigma^{-1}(g_1)\mathbf{y}_{d+1:N}, \tag{A.4}$$

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

$$\Sigma_{\hat{I}}^{true} = \left(\frac{\sigma_I^2}{\sigma_a^2}\right)^2 \Delta'\Sigma^{-1}(g_1)\Sigma(\tilde{g})\Sigma^{-1}(g_1)\Delta. \tag{A.5}$$

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

$$E^{true}\overline{\hat{I}^2} = N^{-1}tr\Sigma_{\hat{I}}^{true}. \tag{A.6}$$

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

$$\Sigma_{\hat{I}/\sigma_a} = \left(\frac{\sigma_I^2}{\sigma_a^2}\right)^2 \Delta'\Sigma^{-1}(g_1)\Delta, \tag{A.7}$$

and (3.7) is given by $N^{-1}tr\Sigma_{\hat{I}/\sigma_a}$.

A.3 Test statistic standard errors

We now derive formulas for the variances of $\overline{\hat{I}^2}$ and $\sigma^2_{a,mle}(\eta/\phi)$ 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

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

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

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

with $\Sigma_{\hat{I}/\sigma_a}$ given by (A.7).

With $\hat{\sigma}_a^2 = c_N\sigma^2_{a,mle}(\eta/\phi)$, see (2.6),

$$\tau_N = \overline{\hat{I}^2} - c_N\sigma^2_{a,mle}(\eta/\phi)\left\{N^{-1}tr\Sigma_{\hat{I}/\sigma_a}\right\} \tag{A.8}$$

therefore has the variance

$$\frac{2\sigma_a^4}{N^2}\left[tr\left\{\Sigma_{\hat{I}/\sigma_a}^2\right\} - \frac{2c_N - c_N^2}{(N-d)}\left(tr\Sigma_{\hat{I}/\sigma_a}\right)^2\right].$$

For an approximate standard error for the test statistic $\tau_N^{(1)}$, we replace the unknown σ_a^4 with $\hat{\sigma}_a^4$ from (2.6) and take the positive square root to obtain

$$\hat{\sigma}_N(\tau_N^{(1)}) = \frac{\sqrt{2}\hat{\sigma}_a^2}{N}\left[tr\left\{\Sigma_{\hat{I}/\sigma_a}^2\right\} - \frac{2c_N - c_N^2}{(N-d)}\left(tr\Sigma_{\hat{I}/\sigma_a}\right)^2\right]^{1/2}. \tag{A.9}$$

For the test statistic $\tau_N^{(2)}$, the formula (A.9) is modified by leaving $(N-d)$ unchanged but otherwise replacing N elsewhere with $(N-24)$ and by replacing the matrix Δ in (A.7) with the matrix $\Delta^{(2)}$ which differs from Δ in its first and last twelve columns by having all entries in these columns equal to 0.

A.4 Diagnostic and test statistic formulas for a general decomposition $Y_t = U_t + V_t$

Suppose the observed time series is modeled as having a decomposition $Y_t = U_t + V_t$ with components that obey ARIMA models whose "differencing" polynomials $\delta^U(B)$ and $\delta^V(B)$, of

degrees d_U and d_V respectively, have no common zeros. (If U_t, say, is covariance stationary, set $\delta^U(B) = 1$ and $d_U = 0$.) The differenced series $u_t = \delta^U(B)U_t$ and $v_t = \delta^V(B)V_t$ are assumed to be uncorrelated with one another and to have mean zero. These assumptions imply an ARIMA model for Y_t with differencing polynomial $\delta(B) = \delta^U(B)\delta^V(B)$ of degree $d = d_U + d_V$. The initial value conditions associated with Assumption A of Bell (1984) or Bell and Hillmer (1988) are assumed to hold for the ARIMA difference equations for U_t and V_t. Assumption A states that the initial d values of Y_t, i.e., the variables $Y_1, Y_2, \cdots, Y_d$, are independent of $\{u_t\}$ and $\{v_t\}$. This ensures that the standard formulas such as (2.3) yield minimum mean square linear estimates. The diagnostics are defined for the stationary transforms of the components, e.g. u_t.

As in the development of Section $A.3$, we need the covariance matrix of $\hat{u}_t$. Let Δ_V denote an $N - d \times N - d_U$ band matrix of the form

$$
\Delta_V = \begin{bmatrix} -\delta_{d_V}^V & \cdots & -\delta_1^V & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & -\delta_{d_V}^V & \cdots & -\delta_1^V & 1 \end{bmatrix}. \tag{A.10}
$$

Following Bell and Hillmer (1988), our estimate of the differenced signal is the vector

$$
\hat{u} = \Sigma_{u/\sigma_a} \Delta_V' \Sigma^{-1}(g_1) \mathbf{y}_{d+1:N} \tag{A.11}
$$

where Σ_{u/σ_a} is the covariance matrix of u under the assumption that y_t has unit innovation variance. Therefore,

$$
\Sigma_{\hat{u}}^{true} = \Sigma_{u/\sigma_a} \Delta_V' \Sigma^{-1}(g_1) \Sigma(\tilde{g}) \Sigma^{-1}(g_1) \Delta_V \Sigma_{u/\sigma_a} \tag{A.12}
$$

so that $E^{true}\hat{u}_t^2$ is the t-th diagonal entry of $\Sigma_{\hat{u}}^{true}$, and

$$
E^{true}\overline{\hat{u}^2} = N^{-1} tr\Sigma_{\hat{u}}^{true}. \tag{A.13}
$$

The scaled model-based variance of $\hat{u}$ is given by

$$
\Sigma_{\hat{u}/\sigma_a} = \Sigma_{u/\sigma_a} \Delta_V' \Sigma^{-1}(g_1) \Delta_V \Sigma_{u/\sigma_a}. \tag{A.14}
$$

So we can define our diagnostic to be

$$
\tau_N = \overline{\hat{u}^2} - c_N \, \sigma_{a,mle}^2 (\eta/\phi) \left\{ N^{-1} tr\Sigma_{\hat{u}/\sigma_a} \right\} \tag{A.15}
$$

which, in a derivation analogous to that of (A.9), has standard error

$$
\hat{\sigma}_N(\tau_N) = \frac{\sqrt{2}\hat{\sigma}_a^2}{N} \left[tr\left\{ \Sigma_{\hat{u}/\sigma_a}^2 \right\} - \frac{2c_N - c_N^2}{(N-d)} \left(tr\Sigma_{\hat{u}/\sigma_a} \right)^2 \right]^{1/2} \tag{A.16}
$$

where $\hat{\sigma}_a$ comes from (2.6).

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

References

[1] 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.

[2] Bell, W. (1984). Signal Extraction for Nonstationary Time Series. *The Annals of Statistics* **12**, 646 –664.

[3] 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

[4] Burman, J. (1980), "Seasonal Adjustment by Signal Extraction," *Journal of the Royal Statistical Society, Ser. A,* **143**, 321–337.

[5] Feldpausch, R. M., Hood, C., and Wills, K. (2004), "Diagnostics for Model-Based Seasonal Adjustment," *2004 Proceedings American Statistical Association* [CD-ROM]: Alexandria, VA

[6] 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

[7] 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

[8] Findley, D., Monsell, B., Shulman, H., and Pugh, M. (1990), "Sliding Spans Diagnostics for Seasonal and Related Adjustments," *Journal of the American Statistical Association,* **85**, 345–355.

[9] 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).

[10] 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 Análisis y Programación Presupuestaria, Madrid.

[11] Hillmer, S. and Tiao, G. (1982), "An ARIMA-Model-Based Approach to Seasonal Adjustment," *Journal of the American Statistical Association*, **77**, 63 – 70.

[12] Kitagawa, G. (1981), "A Nonstationary Time Series Model and its Fitting by a Recursive Filter," *Journal of Time Series Analysis,* **2**, 103–116.

[13] Kitagawa, G. (1985), DECOMP. In TIMSAC-84 Part 2, Computer Science Monographs No. 23, The Institute of Statistical Mathematics, Tokyo.

[14] Koopman, S., Harvey, A., Doornik, J., and Shepherd, N. (1995), *Stamp 5.0: Structural Time Series Analyser, Modeller, and Predictor,* London: Chapman and Hall.

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

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

[17] 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

[18] Monsell, B., Aston, J. and Koopman, S. (2003), "Toward X-13?" *Proceeding of the Business and Economic Statistics Section of the American Statistical Association.*

[19] 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
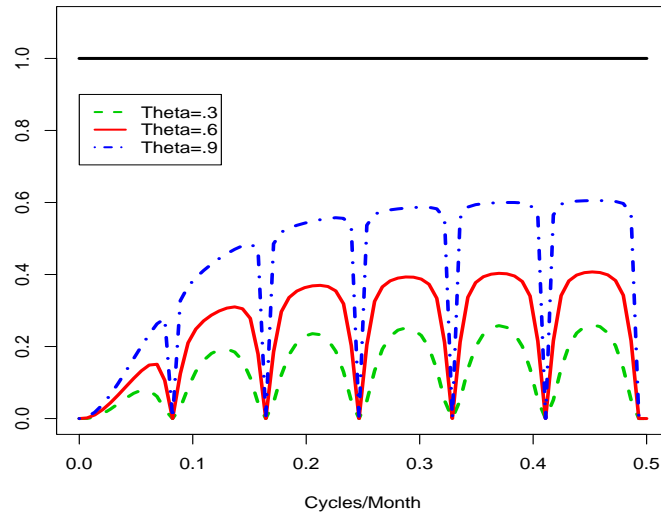
Figure 1: Squared gain function of the midpoint irregular filters of length 144 for $\theta = 0.6$ and $\Theta = 0.3, 0.6, 0.9$.
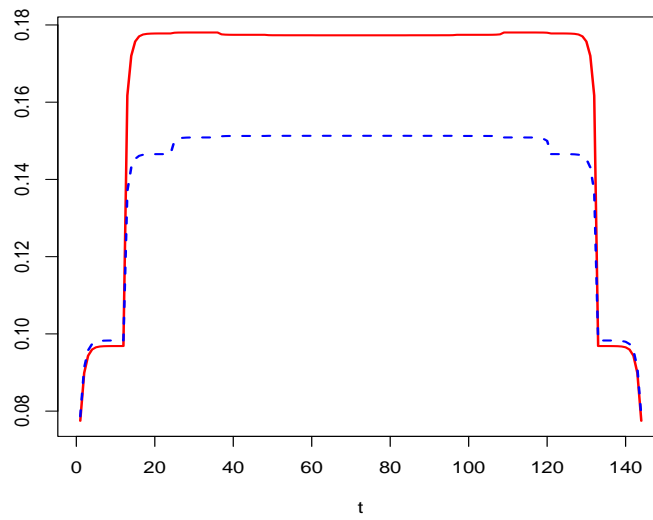

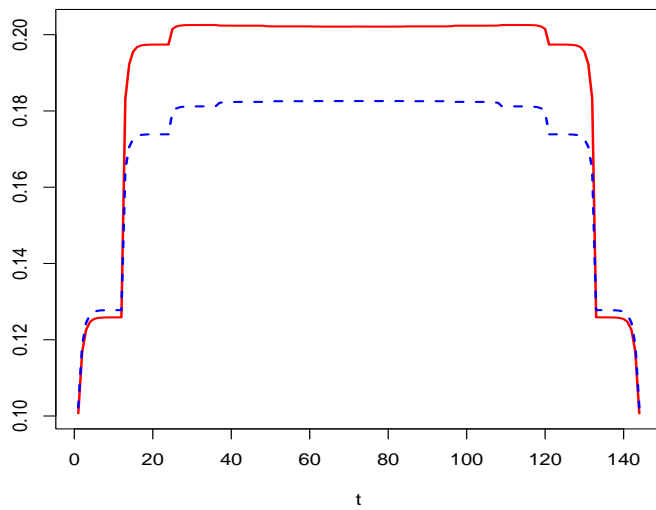Figure 2: $E^{true}\hat{I}_t^2$ (solid) and $\bar{\sigma}_t^2$ (dots) from $\Theta = 0.3$

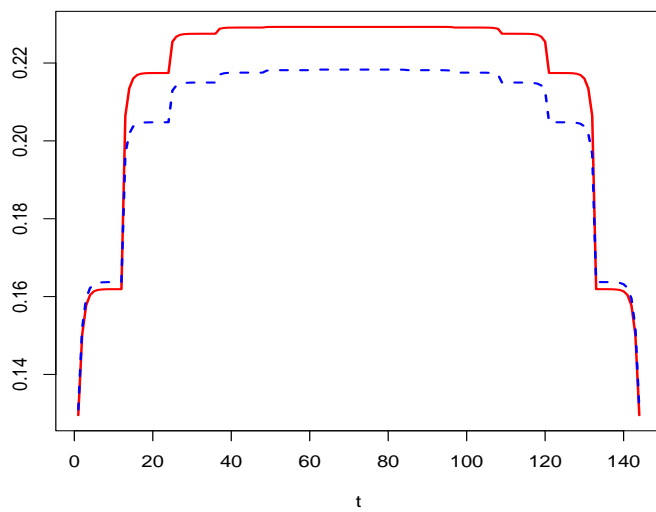Figure 3: $E^{true}\hat{I}_t^2$ (solid) and $\bar{\sigma}_t^2$ (dots) from $\Theta = 0.4$

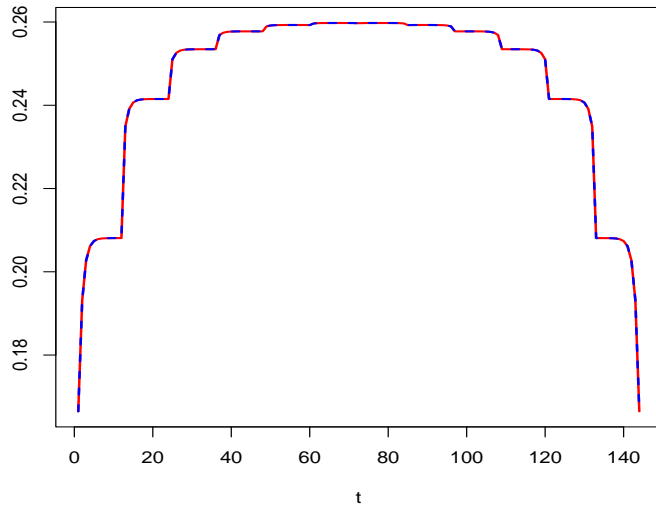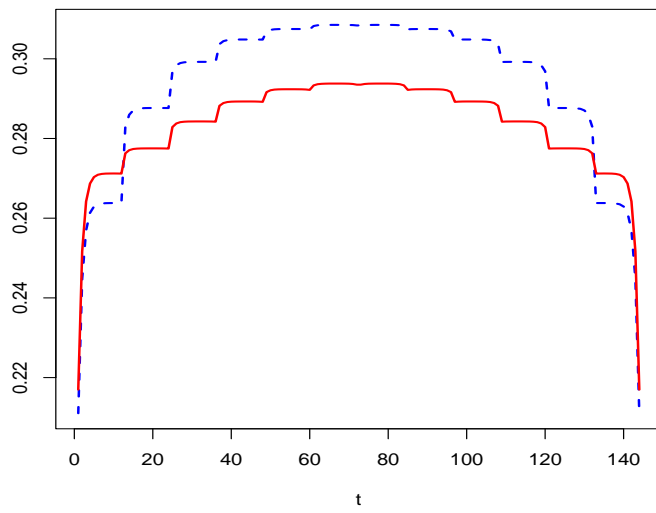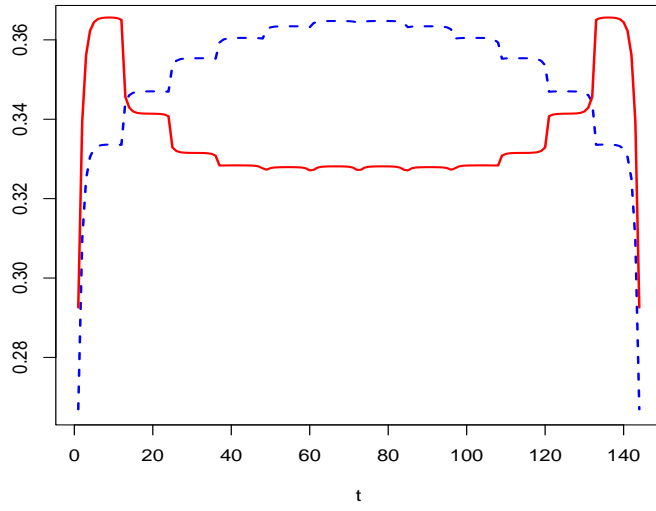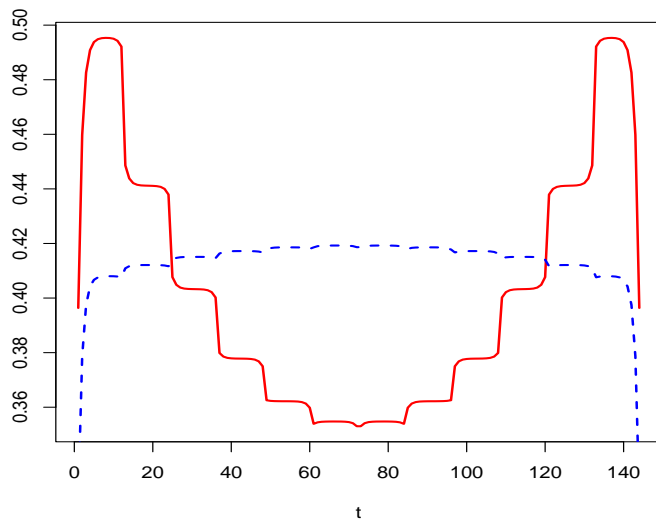Figure 4: $E^{true}\hat{I}_t^2$ (solid) and $\bar{\sigma}_t^2$ (dots) from $\Theta = 0.5$

22

Figure 5: $E^{true}\hat{I}_t^2$ (solid) and $\bar{\sigma}_t^2$ (dots) from $\Theta = 0.6$

Figure 6: $E^{true}\hat{I}_t^2$ (solid) and $\bar{\sigma}_t^2$ (dots) from $\Theta = 0.7$

Figure 7: $E^{true}\hat{I}_t^2$ (solid) and $\bar{\sigma}_t^2$ (dots) from $\Theta = 0.8$

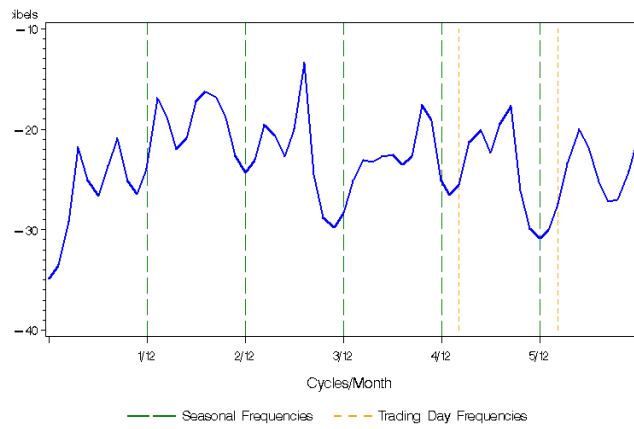Figure 8: $E^{true}\hat{I}_t^2$ (solid) and $\bar{\sigma}_t^2$ (dots) from $\Theta = 0.9$

24

Figure 9: Spectrum of the Irregulars of *Export*. Note the seasonal peak at 6/12 cycles/month.