Diagnostics for ARIMA-Model-Based Seasonal Adjustment

D. F. Findley, K.C. Wills, J. A. Aston, R. M. Feldpausch, and C. C. Hood
U.S. Census Bureau

KEY WORDS Residual seasonality, spectrum, SEATS, underestimation, overestimation, sliding spans

Abstract

We present two types of diagnostics for SEATS and similar programs. We start with modifications of the diagnostic used by SEATS to detect underestimation or overestimation, meaning inadequate or excessive supression of frequency components near the target frequencies, e.g. the seasonal frequencies in the case of seasonal adjustment. The modifications use time-varying variances of the finite-length filter output instead of the constant variance associated with the infinite-length filter. The SEATS diagnostic is shown to be substantially biased toward indicating underestimation even when the estimation is optimal, a situation where the modified diagnostics are unbiased. The second type of diagnostic considered is an adaptation of the widely used sliding spans diagnostic of X-12-ARIMA. The adaptation is a method for determining the span length appropriate for model-based-adjustment as a function of the ARIMA model's seasonal moving average parameter.

1. Introduction

The most widely used ARIMA model-based approach to seasonal adjustment is the "canonical" decomposition approach of Hillmer and Tiao (1982) and Burman (1980) as implemented in SEATS (Gomez and Maravall, 1997) and in other software under development, see Monsell, Aston and Koopman (2003). We present modifications of SEATS' diagnostic for determining underestimation and overestimation in Section 2. These modified diagnostics could also be used in the "structural" ARIMA model-based approaches of DECOMP (Kitagawa 1981, 1985) and STAMP (Koopman, Harvey, Doornik and Shepherd, 1995). In Section 3, we provide an adaptation of the sliding spans diagnostic of Findley, Monsell, Shulman and Pugh (1990) for the canonical decomposition approach.

The model-based approaches referenced above assume that, after removal of any trading day, holiday and/or outlier effects, the time series to be seasonally adjusted, $Y_t, 1 \leq t \leq N$, (or $\log Y_t$) can be decomposed into a sum of seasonal, trend and irregular components denoted by S_t, T_t and I_t respectively,

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

The model-based methods provide an indirectly or directly specified model for each component and produce estimates $\hat{S}_t$, $\hat{T}_t$, and $\hat{I}_t$ that are Gaussian conditional means, e.g.

$$\begin{aligned} \hat{S}_t &= E\left(S_t | Y_s, 1 \leq s \leq N\right) \\ &= \sum_{j=t-N}^{t-1} c_{j,t}^{S}\left(N\right) Y_{t-j}. \end{aligned} \tag{2}$$

Thus the $c_{j,t}^{S}\left(N\right)$ are the t- and N-dependent minimizers of $E\left(S_t - \sum_{j=t-N}^{t-1} c_{j,t}^{S} Y_{t-j}\right)^2$ under assumptions that make it possible to evaluate this expectation by treating the values of model parameters estimated from $Y_s, 1 \leq s \leq N$ as if they were correct. $\hat{T}_t$ and $\hat{I}_t$ are defined analogously. (For simplicity, we will usually suppress the dependence of the estimates on N, the models, and the decomposition.) Here and elsewhere, $E\left(\cdot\right)$ denotes the mean calculated according to the model specified for the data. 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 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. The program DECOMP provides spectrum estimates of $\hat{A}_t$ and $\hat{I}_t$ for detecting residual seasonality (and residual trading day effects). Similar estimates are provided in X-12-ARIMA and the programs discussed in Monsell, Aston and Koopman (2003).

SEATS 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 to mean that its estimate does not capture all of the seasonal variation. Overestimation means that too much variation has been assigned to this component. Underestimation is SEATS' main indicator of residual seasonality. For the seasonally adjusted component $\hat{A}_t$ or the irregular component $\hat{I}_t$, underestimation can be interpreted to mean that the seasonal adjustment or irregular filters do not adequately suppress frequency components around the seasonal frequencies. That is, the dips in their gain functions at the seasonal frequencies are too narrow. Overestimation means these dips are too wide, with the result that too much suppression occurs. (The irregular filters provide suppression both around the seasonal frequencies and also around low frequencies associated with trend, so mis-estimation of I_t could also result from inappropriate suppression of low-frequency components).

2. Variance of Estimator vs. Variance of Estimate (SEATS)

We now describe the SEATS diagnostic intended to detect overestimation and underestimation. We show it is biased toward identifying underestimation even when the estimate is optimal. Then we present analyses of a modified diagnostic that is unbiased. For simplicity, we focus on estimation of the irregular component, because the model for this component is white noise and therefore specifies a constant variance $\sigma_I^2 = E\left(I_t^2\right)$. (For the seasonal, trend and seasonally adjusted components, the diagnostic is calculated for the "stationary transformation" of each component, meaning the output of the "differencing" operation specified by the component's ARIMA model.) Our limited study seems to be the first systematic investigation of the SEATS 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 ARIMA model for the data is expressed in the usual notation as

$$\delta\left(B\right)\phi\left(B\right)Y_t = \theta\left(B\right)a_t. \qquad (3)$$

Thus $\delta\left(B\right)$ denotes the differencing operator which transforms Y_t to stationarity, e.g. $\delta\left(B\right) = (1-B)^2\left(1+B+\cdots+B^{p-1}\right)$, a_t is the innovations process (with variance σ_a^2), etc. Also, B is the backshift operator and p is the number of observations per year: $p = 12$ in our analyses. The pseudo-spectral density of the model for Y_t is

$$f\left(\lambda\right) = \frac{\sigma_a^2}{2\pi}\left|\frac{\theta\left(e^{i\lambda}\right)}{\delta\left(e^{i\lambda}\right)\phi\left(e^{i\lambda}\right)}\right|^2.$$

The spectral density of the associated model for I_t is $\sigma_I^2/2\pi$. Then, with

$$c_j^I = \frac{\sigma_I^2}{4\pi^2}\int_{-\pi}^{\pi}\frac{\cos j\lambda}{f\left(\lambda\right)}d\lambda,\ j = 0, \pm 1, \ldots, \qquad (4)$$

the *estimator* of I_t is given by the Gaussian conditional expectation defined by the covariance structure specified by the Wiener-Kolmogorov formula

$$\begin{aligned} I_{WK,t} &= E\left(I_t|Y_s, -\infty < s < \infty\right) \qquad (5)\\ &= \sum_{j=-\infty}^{\infty} c_j^I Y_{t-j}, \end{aligned}$$

under the assumption that the initial values of Y_t used to determine the solution to the difference equation (3) are independent of the stationary time series $\delta\left(B\right)Y_t$, see Bell (1984). The *variance of the estimator* $\sigma_{WK,I}^2 = E\left(I_{WK,t}^2\right)$ has the formula $\sigma_{WK,I}^2 = \sigma_I^2 c_0^I$.

With $Y_t, 1 \leq t \leq N$ denoting the data available for seasonal adjustment, SEATS defines the *estimate* of I_t to be the model's conditional mean from the available N observations

$$\begin{aligned} \hat{I}_t\left(=\hat{I}_t\left(N\right)\right) &= E\left(I_t|Y_s, 1 \leq s \leq N\right) \qquad (6)\\ &= \sum_{j=t-N}^{t-1} c_{j,t}^I\left(N\right)Y_{t-j}. \end{aligned}$$

In principle, SEATS' *variance of the estimate* is defined to be the sample second moment*

$$\overline{\hat{I}^2} = \frac{1}{N} \sum_{t=1}^{N} \hat{I}_t^2, \tag{7}$$

and *underestimation* is indicated when $\overline{\hat{I}^2} < \sigma^2_{WK,I}$, *overestimation* when $\overline{\hat{I}^2} > \sigma^2_{WK,I}$. In practice, the calculations that produce the component models in SEATS and other software yield variances like σ^2_I and $\sigma^2_{WK,I}$ scaled as though the innovation variance of (3) were equal to one, quantities we shall denote by σ^2_I/σ^2_a and $\sigma^2_{WK,I}/\sigma^2_a$. An estimate of $\sigma^2_{WK,I}$ is obtained by multiplying the latter ratio by an essentially unbiased estimate $\hat{\sigma}^2_a$ of σ^2_a. We define $\hat{\sigma}^2_a$ to be the the maximum likelihood estimate $\hat{\sigma}^2_{a,mle}$ when the model coefficients leading to (6) are fixed independently from the data, as in most of our simulation experiments. When coefficients are estimated, SEATS' estimate is used, $\hat{\sigma}^2_a = \hat{\sigma}^2_{a,mle} \left(N - n_{\delta,\phi}\right) / \left(N - n_{\delta,\phi} - n_{coeffs}\right)$, where $n_{\delta,\phi}$ is the order of $\delta(B)$ plus that of $\phi(B)$ (when SEATS' (conditional) estimates of $\phi(B)$ are used), n_{coeffs} is the number of ARMA coefficients in the model. So SEATS criterion is $\overline{\hat{I}^2} < \hat{\sigma}^2_a \frac{\sigma^2_{WK,I}}{\sigma^2_a}$ for underestimation, $\overline{\hat{I}^2} > \hat{\sigma}^2_a \frac{\sigma^2_{WK,I}}{\sigma^2_a}$ for overestimation.

In correct and incorrect model cases with the correct $\delta(B)$, σ^2_a can be interpreted as the model's asymptotic mean square one-step-ahead prediction error, and it can be shown that $\hat{\sigma}^2_a - \sigma^2_a \to 0$ in probability and in mean, see (b) of Proposition 5.2 of Findley, Pötscher and Wei (2003). When $E\left(\hat{\sigma}^2_a/\sigma^2_a\right) \simeq 1$, the following simple Proposition shows that SEATS's under/overestimation criterion is biased toward indicating underestimation when the estimate is actually optimal, i.e. when the model is correct. A slight modification of the Proposition's proof can be used to obtain the analogous results for the stationary transforms of the seasonal, seasonally adjusted and trend components, which we will not discuss further.

Proposition 1 *When the ARIMA model (3) and variance σ^2_I used for the calculation of the estimates $\hat{I}_t$ in (6) and of $\sigma^2_{WK,I}$ are correct, then*

$$E\left(\overline{\hat{I}^2}\right) \leq \sigma^2_{WK,I}. \tag{8}$$

Strict inequality necessarily holds when the moving average polynomial $\theta(B)$ in (3) has positive degree.

Proof. The estimate $\hat{I}_t$ and its error $I_t - \hat{I}_t$ sum to I_t, and, because the model is correct, are uncorrelated. The same is true for $I_{WK,t}$ and $I_t - I_{WK,t}$. Thus. setting $\sigma^2_t = E\left(\hat{I}^2_t\right)$, we have

$$\begin{aligned} \sigma^2_I &= \sigma^2_t + E\left(I_t - \hat{I}_t\right)^2 \\ &= \sigma^2_{WK,I} + E\left(I_t - I_{WK,t}\right)^2. \end{aligned} \tag{9}$$

It follows that

$$\sigma^2_t \leq \sigma^2_{WK,I}, \tag{10}$$

because $E\left(I_t - \hat{I}_t\right)^2 \geq E\left(I_t - I_{WK,t}\right)^2$, since $I_{WK,t}$ is the optimal estimator from the larger data set $Y_s, -\infty < s < \infty$. Further, strict inequality always holds when (3) has a moving average component, because then infinitely many coefficients (4) are nonzero. The assertions of the proposition follow from this fact, (10), and another consequence of the model's correctness,

$$E^{true}\left(\overline{\hat{I}^2}\right) = N^{-1} \sum_{t=1}^{N} \sigma^2_t, \tag{11}$$

where $E^{true}(\cdot)$ denotes expectation with the respect to the true distribution of the data. ■

For the case in which Y_t follows the Box-Jenkins airline model (AL)

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

with $\theta = 0.6$, Table 1 shows the *relative bias* $\sigma^2_{WK,I}/E^{true}\left(\overline{\hat{I}^2}\right)$ for sample sizes $N = 72, 144$ and various Θ.

The relative bias values are very similar for other values of θ with $0.1 \leq \theta \leq 0.9$. These results, documenting substantial bias, suggest that in place of the constant variance $\sigma^2_{WK,I}$ that applies to the infinite case (5), the average of the time-varying t-dependent variances $\sigma^2_t = E\left(\hat{I}^2_t\right)$ should be used, calculated as

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

though the estimated or specified model is correct. Thus, the resulting *modified diagnostic* compares $\overline{\hat{I}^2}$ to

$$\frac{1}{N}\sum_{t=1}^{N}\sigma_t^2. \tag{13}$$

instead of to $\sigma^2_{WK,I}$. Note that σ_t^2 can be obtained from $\sigma_t^2 = \sigma_I^2 - E\left(I_t - \hat{I}_t\right)^2$, cf. (9). The two r.h.s. quantities are available from model-based adjustment programs that use state space methods, usually in units of σ_a^2. When only these scaled quantities, which we denote by σ_t^2/σ_a^2, are available, the role of (13) is taken by

$$\frac{\hat{\sigma}_a^2}{N}\sum_{t=1}^{N}\frac{\sigma_t^2}{\sigma_a^2}. \tag{14}$$

Thus

$$\overline{\hat{I}^2} < \frac{\hat{\sigma}_a^2}{N}\sum_{t=1}^{N}\frac{\sigma_t^2}{\sigma_a^2} \tag{15}$$

indicates underestimation and

$$\overline{\hat{I}^2} > \frac{\hat{\sigma}_a^2}{N}\sum_{t=1}^{N}\frac{\sigma_t^2}{\sigma_a^2} \tag{16}$$

indicates overestimation. SEATS has no significance test for under/overestimation. It is difficult to derive a test that accounts for the variability of both $\overline{\hat{I}^2}$ and $\hat{\sigma}_a^2$.in the difference between $\overline{\hat{I}^2}$ and (14). We plan to investigate a test that accounts for the variability of $\overline{\hat{I}^2}$. Maravall (2003) illustrates such a test for the SEATS diagnostic assuming $\overline{\hat{I}^2}$ has been obtained from (5).

Table 1. Relative Bias $\sigma^2_{WK,I}/E^{true}\left(\overline{\hat{I}^2}\right)$ 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

2.1 Performance of the modified diagnostic

We now present a limited exploration of the ability of the modified diagnostic based on (14) to detect underestimation and overestimation. 5000 independent replicates of series of length 144 were generated by inputting pseudo-$N(0,1)$ innovations a_t into the difference equation (12) with $\theta = \Theta = 0.6$. Underestimated irregulars were obtained from SEATS by specifying adjustment with an airline model with $\theta = 0.6$ and $\Theta > 0.6$. This results in seasonal adjustment filters and irregular filters that do less than optimal suppression of frequency components around the seasonal frequencies. Overestimated irregulars were obtained by specifying $\Theta < 0.6$ (and $\theta = 0.6$) to force greater than optimal suppression of such frequency components. See Findley and Martin (2003) for plots of squared gains of seasonal adjustment filters for various Θ. (The corresponding plots for irregular filters are similar except near $\lambda = 0$.) For each specification, Table 2 lists the percentage of the 5000 series classified as underestimated. The percents for the case in which (12) is estimated for each series should be near fifty. The SEATS diagnostic results reveal strong bias toward indicating underestimation when there is none in this case and in misspecified Θ cases. Thus it is not a reliable diagnostic for residual seasonality. The modified diagnostic that uses (14), designated Mod1 in Table 2, does much better in the overestimation and estimated cases, and does fairly well identifying underestimation. The Mod2 column presents percentages for a further modification described in Subsection 2.2.

Table 2. Percent of Simulated Airline Model Series with $\theta = \Theta = 0.6$ for Which (15) Holds for Diagnostics of Adjustments Produced with Various Incorrect Θ's

Incorrect Θ	SEATS	Mod1	Mod2
0.3	12.1	1.6	2.2
0.4	32.2	6.3	8.2
0.5	62.7	22.1	24.6
estimated θ, Θ	100	48.0	48.6
0.7	96.6	74.2	73.3
0.8	99.1	83.7	83.5
0.9	98.4	66.4	80.8

The local properties considered next indicate that the decreases in Table 2 at $\Theta = 0.9$

are due to end point behavior.

2.2 Local properties and a further modification

When the ARIMA model specified to obtain the estimate $\hat{I}_t$ is incorrect, the model-based quantities σ_t^2 and their average (13) will generally be different from the true means of $\hat{I}_t^2$ and $\overline{\hat{I}^2}$ respectively. In fact, the results of Table 2 suggest that (13) will be larger than the mean of $\overline{\hat{I}^2}$ when there is underestimation and smaller when there is overestimation for the examples considered. This conjecture is supported by Table 3 which presents the results of a simulation experiment with $10,000$ independent replicates of series of length 144 satisfying (12) with $\theta = \Theta = 0.6$. We use $\hat{E}^{true}\left(\hat{I}_t^2\right)$ resp. $\hat{E}^{true}\left(\overline{\hat{I}^2}\right)$ to denote the simple average over the replicates of $\hat{I}_t^2$ resp. $\overline{\hat{I}^2}$.

Table 3. Simulation Means of $\overline{\hat{I}^2}$ from 10000 Series with $\theta = \Theta = 0.6$ Compared to (13)

Θ	$\hat{E}^{true}\left(\overline{\hat{I}^2}\right)$	(13)
0.3	0.1633	0.1236
0.5	0.2147	0.2003
0.6	0.2455	0.2455
0.7	0.2834	0.2965
0.8	0.3356	0.3534
0.9	0.4036	0.4135

A natural next question is whether the averages over all t inherit these properties from analogous local properties. That is, for every t, is the mean of $\hat{I}_t^2$ less than σ_t^2 when there is underestimation, and greater than σ_t^2 when there is overestimation? An analysis using $\hat{E}^{true}\left(\hat{I}_t^2\right)$ from 5000 simulations to represent the true means indicates that a local property may hold in the case of overestimation but does not in the case of underestimation, where the values of $\hat{E}^{true}\left(\hat{I}_t^2\right)$ are consistently larger than σ_t^2 near the ends of the series. Figures 2–6 present the graphs of $\hat{E}^{true}\left(\hat{I}_t^2\right)$ and σ_t^2 that provide these conclusions for the incorrect model cases of Table 3. Figure 1 shows that the $\hat{E}^{true}\left(\hat{I}_t^2\right)$ are very close to the true means $E^{true}\left(\hat{I}_t^2\right) = \sigma_t^2$ in the correct model case $\Theta = 0.6$.

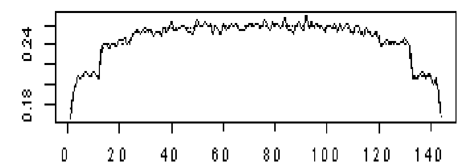

Figure 1: $\hat{E}^{true}\left(\hat{I}_t^2\right)$ (solid) and σ_t^2 (dotted) from $\Theta = 0.6$

The time intervals at the ends of the series over which $\hat{E}^{true}\left(\hat{I}_t^2\right) > \sigma_t^2$ are substantially wider in Figure 6 for the case $\Theta = 0.9$ than in Figure 5 for the case $\Theta = 0.8$. Also, the values of $\hat{E}^{true}\left(\hat{I}_t^2\right) - \sigma_t^2$ over these outer intervals are visually greater on average, and the values between these intervals do not compensate for this difference, as is shown in Table 3 by the fact that the percentage increase from

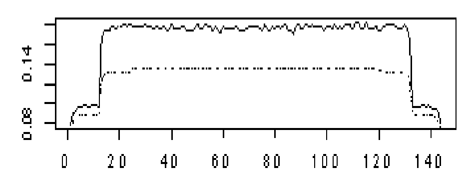

Figure 2: $\hat{E}^{true}\left(\hat{I}_t^2\right)$ (solid) and σ_t^2 (dotted) from $\Theta = 0.3$

$\hat{E}^{true}\left(\overline{\hat{I}^2}\right)$ to (13) for $\Theta = 0.9$ is less than half of what it is for $\Theta = 0.8$ (2.4% vs. 5.3%). These results help to explain the finding from Table 2 that, even though the underestimation is more extreme when $\Theta = 0.9$ than when $\Theta = 0.8$, it is less detectable with the diagnostic. They also suggest, for the case $N = 144$ at least, that modifying the sums in (7) and (14) to run from 13 to $N-12$ might yield a diagnostic that identifies underestimation more reliably with little or no loss in identifications of overestimation. Results for this diagnostic, listed in the Mod2

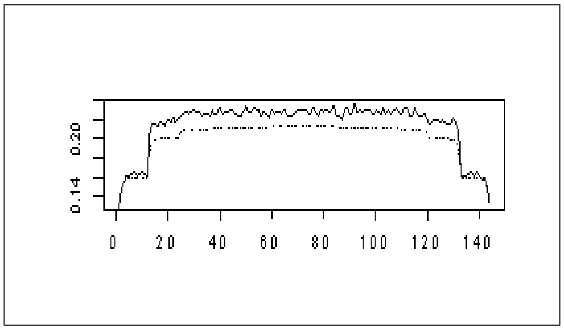

Figure 3: $\hat{E}^{true}\left(\hat{I}_t^2\right)$ (solid) and σ_t^2 (dotted) from $\Theta = 0.5$

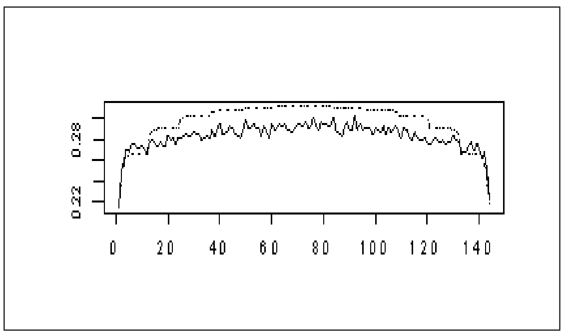

Figure 4: $\hat{E}^{true}\left(\hat{I}_t^2\right)$ (solid) and σ_t^2 (dotted) from $\Theta = 0.7$

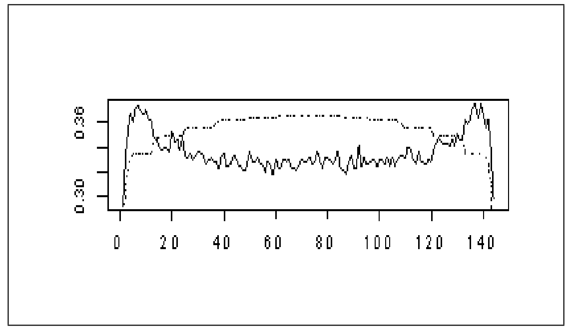

Figure 5: $\hat{E}^{true}\left(\hat{I}_t^2\right)$ (solid) and σ_t^2 (dotted) from $\Theta = 0.8$

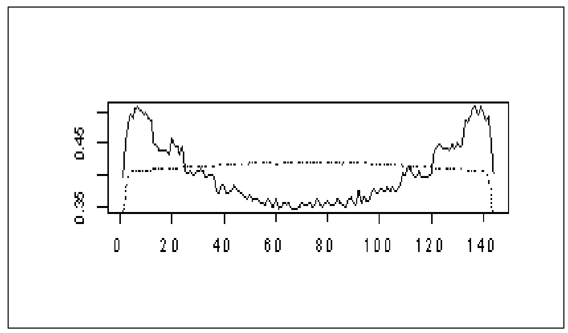

Figure 6: $\hat{E}^{true}\left(\hat{I}_t^2\right)$ (solid) and σ_t^2 (dotted) from $\Theta = 0.9$

column of Table 2, confirm this conjecture.

In summary, we have shown that modified diagnostics based on finite-filter variances, particularly the second diagnostic just described, perform better than SEATS' biased diagnostic which uses an infinite-filter variance. This fits the pattern of the conclusion of Findley and Martin (2003) that finite-filter frequency domain diagnostics are more informative than the corresponding diagnostics of the Wiener-Kolmogorov filter (5). Another finding is that underestimation is more difficult to identify than overestimation.

3. Sliding Span Lengths For SEATS

We now consider a suite of diagnostics not currently in SEATS. The sliding spans diagnostics of X-12-ARIMA compare seasonal adjustments (or trends, etc.) of the same month calculated from four overlapping equal-length (sub)spans of the data. The adjustment options (ARIMA model, holiday regressors, filters, etc.) specified for each span are the options specified for the entire time series. The diagnostics calculated from these comparisons by X-12-ARIMA are fundamental for deciding when a seasonal adjustment must be rejected because of inconsistency/instability, see Findley, Monsell, Shulman and Pugh (1990) and Findley et al. (1998) for details.

These diagnostics can be used if $L + 3$ years of data are available, where L, the span length in years, is large enough that the specified adjustment options can yield adjustments of reasonable quality for a span of length L, under moderately favorable circumstances. Seasonal adjustment filters whose coefficients decay slowly require longer spans. The "effective" length of the seasonal adjustment filter is the length needed to obtain adjustments that will change little when more data become available. The filters for adjustment of months near the center of the series usually produce the best adjustments. Therefore, for specified software options, we seek to determine the shortest series length such that adjustments near the series' center will change little if the time series is extended with consistent data, past or future.

For example, when a default X-11 seasonal adjustment is obtained with the 3×5 seasonal filter (and the 13-term Henderson trend filter), the non-zero weights of the symmetric seasonal adjustment filter for the center of the series span 169 months if the series is this long or longer. However, it is usually the case that almost the same seasonal adjustment at and near the center is obtained for a series of length 85 months with the same center. Consequently, when the series is 96 or 97 months long, seasonal adjustments for months within a year or so of its center are usually little changed when the series is extended with additional data. The sizes of such changes increase with increasing distance from the center of the series and can be large within two years of the ends of the series. From such considerations, when the 3×5 seasonal filter is used, eight years is taken to be the appropriate span length for sliding spans analysis, so eleven years of data are needed to calculate sliding spans diagnostics. By similar reasoning, six years would be used as the span length with the 3×3 seasonal filter and twelve years with the 3×9 seasonal filter.

The model-based seasonal adjustment filters of SEATS are always as long as the data span being adjusted (when the ARIMA model specified has a moving average component). We present an approach for determining "effective" filter/span lengths for the sliding spans diagnostic from an analysis of filters associated with the airline model. (This is the model chosen for about half the series adjusted by SEATS, see Gomez and Maravall, 1997.) Using the fact that values of θ and Θ are known for which the SEATS seasonal adjustment filters have gain and phase-shift properties very close to those of the X-11 filters (see Planas and Depoutot, 2002 and Findley and Martin, 2003), we calibrate the span lengths used for SEATS to coincide with the span lengths used for the X-11 filters when the two types of filters are close. In this way, the span length specifications used for SEATS adjustments are anchored to the successful experiences obtained with the X-11 filters.

Let t_c denote the midpoint of an observation interval for Y_t of length 241 months. For $2 \leq L \leq 20$ and $t_c - 6L \leq t \leq t_c + 6L$, consider the seasonal adjustment from a data span of length $12L + 1$ centered at t_c,

$$A_t^L = E\left(A_t | Y_s,\ t_c - 6L \leq s \leq t_c + 6L\right).$$

We focus on the adjustment one year past the center of the span, $A_{t_c+12}^L$, and consider its root mean square difference from the adjustment $A_{t_c+12}^{20}$ from the 241 month span as a fraction of the root mean square error of $A_{t_c+12}^{20}$,

$$relerr(L) = \left\{ \frac{E\left(A_{t_c+12}^L - A_{t_c+12}^{20}\right)^2}{E\left(A_{t_c+12}^{20} - A_{t_c+12}\right)^2} \right\}^{1/2}.$$

We have chosen $A_{t_c+12}^{20}$, the adjustment available with twenty years (and one month) of data, as the closest value to the truth A_{t_c+12} the user might ever see and therefore the reference adjustment for evaluating the performance of $A_{t_c+12}^L$. The quantity $relerr(L)$ is easily calculated from state-space smoothing algorithms using the decomposition $E\left(A_{t_c+12}^L - A_{t_c+12}\right)^2 = E\left(A_{t_c+12}^L - A_{t_c+12}^{20}\right)^2 + E\left(A_{t_c+12}^{20} - A_{t_c+12}\right)^2$, which results from the fact that for $L \leq 20$, $A_{t_c+12}^L - A_{t_c+12}^{20}$ is a linear combination of $Y_s,\ t_c - 120 \leq s \leq t_c + 120$ and is therefore uncorrelated with the error of the optimal estimate $A_{t_c+12}^{20}$ of A_{t_c+12} from these Y_s.

When Y_t obeys the airline model, $relerr(L)$ depends on θ and Θ, but the dependence on θ is weak. For a given Θ (and θ between 0.2 and 0.8), define the span length for sliding spans analysis to be

$$L_\Theta = \min\left(L : relerr(L) \leq 0.25\right). \quad (17)$$

Table 4. $\Theta \geq 0$ at which the span length L_Θ increases to the value indicated.

Θ	L_Θ (in years)
0.160	5
0.325	6
0.490	7
0.535	8
0.620	9
0.640	10
0.695	11
0.710	12
0.750	13
0.760	14
0.795	15
0.805	16
0.840	17
0.850	18
> 0.910	19

The span lengths of 6, 8, and 12 years used by X-12-ARIMA with the 3×3, 3×5, and 3×9 filters coincide with the lengths L_Θ for the Θ's that define the best airline model approximations to these seasonal adjustment filters in Table 1 of Planas and Depoutot (2002), namely $\Theta = 0.38$, 0.55, and 0.73, respectively. (The θ values paired with these Θ in their Table 1 are between 0.58 and 0.60.)

Our Table 4, obtained from (17) with airline models, provides the span lengths L_Θ (in years) we currently recommend for sliding spans analyses of SEATS seasonal adjustments from a model with seasonal differencing and a seasonal moving average factor $\Theta \geq 0.160$. For $0 \leq \Theta < 0.160$, use $L_\Theta = 4$.

With four spans of the specified length, sliding spans diagnostics are to be interpreted as indicated in Findley, Monsell, Bell, Otto and Chen (1998). In preliminary studies at the Census Bureau involving simulated series with known seasonal decomposition components, this sliding spans diagnostic implemented for SEATS provided more reliable indications of adjustment inaccuracy than all other diagnostics considered: on average, the greater the inaccuracy, the larger the maximal changes in adjustments over the spans. This approach to determining span lengths does not provide useful results for $\Theta \geq 0.85$. In fact, it requires series lengths that will often be impractical for $\Theta \geq 0.71$. We have ideas for approaches to detect excessive instability with spans of length 10 years or less for large values of Θ.

Acknowledgement. We thank Doug Bond, Donald Martin and Brian Monsell for helpful comments on this paper.

REFERENCES

Bell, W. R. (1984), "Signal extraction for nonstationary time series," Annals of Statistics, 12, 646-664.

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

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. Available from 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.

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

Findley, D. F., Pötscher, B. M., and Wei, C. Z. (2003), "Modeling of Time Series Arrays by Multistep Prediction or Likelihood Methods," *Journal of Econometrics* 118, 151-187.

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.

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

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

——————— (1985), DECOMP. In TIMSAC-84 Part 2, Computer Science Monographs No. 23, The Institute of Statistical Mathematics, Tokyo.

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

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

Monsell, B. C., Aston, J. A. and Koopman, S. J. (2003), "Toward X-13?," Proceedings of the Business and Economic Statistics Section of the American Statistical Association (to appear).

Planas, C. and Depoutot, R. (2002), "Controlling Revisions in ARIMA-Model-Based Seasonal Adjustment," *Journal of Time Series Analysis* 23, 193-214.