ON THE SPECTRUM DIAGNOSTICS USED BY X-12-ARIMA TO INDICATE THE PRESENCE OF TRADING DAY EFFECTS AFTER MODELING OR ADJUSTMENT

Raymond J. Soukup and David F. Findley

Statistical Research Division, Room 3000-4, U.S. Census Bureau, Washington DC 20233

Key Words: Seasonal Adjustment, Trading day, Spectral analysis, Time series

INTRODUCTION

Most seasonal adjustment programs for monthly and quarterly time series include a capability for estimating and removing trading day effects, repetitive effects associated with the seven days of the week. In monthly series, such effects produce periodic movements, primarily at the frequency associated with the fractional part of the number of weeks (seven day cycles) in an average month. Since spectral analysis can be used to detect the presence of periodic components, it is a natural diagnostic tool for detecting trading day effects as well as seasonal effects. The spectrum is not as powerful a diagnostic as likelihood-ratio based procedures for detecting trading day effects when a good model for the possible trading day effect and the series itself is known, based on fitted regression and ARIMA models. It is more versatile because it does not depend upon a correct model specification.

What is desired, especially for users with large numbers of series, is a set of spectral diagnostics that provides a high rate of detection of actual trading day effects and not too many false alarms. For most analysts, inadequate trading day adjustment of the most recent data is the most important trading-day phenomenon to detect. In the Census Bureau's X-12-ARIMA program (Findley et al. 1998), X-12-ARIMA spectral diagnostics are computed from the last 96 months of the series in default mode.

X-12-ARIMA calculates and plots a spectral estimate of the input series and of as many as three output series. The program prints warning messages when a peak of sufficient amplitude is found at frequencies associated with seasonality or trading-day effects. We only consider monthly flow series in this report, although X-12-ARIMA is capable of handling trading day estimation for stock series and quarterly series also.

One motivation for this study was the concern expressed by some users of earlier versions of X-12-ARIMA about possible over-sensitivity of the program's spectral diagnostics, because of the rather large number of "visually significant" spectral peaks (a criterion determined by the range of the spectrum values —see below) in simulated series without trading day effects, as well as in real series that had no coefficients that were statistically different from zero. Particular attention was given to four issues: 1) the selection of an appropriate spectral estimator, 2) selection of the appropriate output series for spectral estimation, 3) relative importance of the frequencies at which the spectrum is examined for peaks associated with trading day effects, and 4) the selection of the criterion used to define visually significant spectral peaks. Before addressing these issues, we describe the graphical display of spectra in X-12-ARIMA output and the methodology for obtaining detection and false alarm rates for a trading day detection procedure.

Indications of a trading day effect from the spectral diagnostics should stimulate the seasonal adjuster or modeler to include an appropriate trading-day regression component in a regARIMA model and ascertain its statistical significance. Alternatively, the adjuster can determine the extent to which various regression models for trading day effects contribute to the forecasting power of the basic regARIMA model by using out-of-sample forecast error comparison diagnostics described in Findley et al. (1998).

Graphical Display of the Spectrum

Criteria based on visual significance are used because effective statistical tests for the significance of a peak from a periodic component in stationary background noise are not available for series of the length usually considered. We wish to have criteria that are "automatic", in the sense that they have decision procedures that are easy to program and lead to a simple decision of "visually significant" or "not visually significant".

Figure 1 shows an example of a spectral graph that appears in the output of X-12-ARIMA. The columns of stars, left to right, represent frequencies ranging from 0 to 0.5 cycles/month. Each "star" at a particular frequency covers a specific range of amplitude, namely, 1/52nd of the range between the maximum and minimum spectral values across the entire frequency range. To be visually significant, the spectral amplitude at a trading day frequency must exceed both of its neighbors by a number of "stars". This value will be called the detection threshold in units of "stars". A "6-star" detection threshold has been used in all versions of X-12-ARIMA. One of the goals of the study was to determine whether this threshold provides acceptable detection and false alarm rates for our choices of diagnostics.

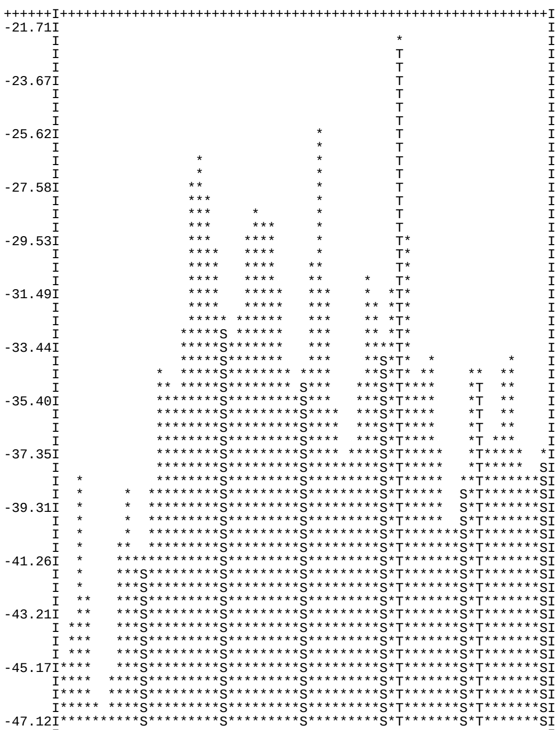

Fig. 1. Spectral graphic from the X-12-ARIMA output file. The graph shows the spectrum of the irregulars for a retail sales series. The "T" columns represent amplitudes at trading day frequencies and the "S" columns represent amplitudes at seasonal frequencies. The y-axis represents log(magnitude).

A heuristic argument suggests examination of spectral peaks at 0.348 cycles/month as a starting point for trading day detection, based on the fractional part of the number of weekly cycles (4.348) that a time series will pass through in a month of average length (30.4375 days/7=4.348). Cleveland and Devlin (1980) created a theoretical model for "flow" series (monthly totals of quantities, e.g. sales, generated throughout the month) that suggested that 0.432 cycles/month is the next most important frequency. Others, for example, McNulty and Huffman (1989), have found that 0.304 cycles/month is associated with certain patterns of daily activity. We will consider the question of how many of these frequencies need to be considered for trading day detection.

Following the BAYSEA program (Akaike 1980) the spectrum is evaluated and plotted at 61 frequencies, $\nu = j/120, j = 0,1,\ldots 60$, except that, in X-12-ARIMA, the trading day frequencies 0.348 and 0.432 are used in place of the frequencies $j/120$ closest to them, and the neighboring frequencies of these two trading day frequencies are defined so that they differ from 0.348 and 0.432 by ± 1/120. The use of 61 frequencies allows for calculation at the exact seasonal frequencies ($\nu = k/12$, k=1,2,3,4,5,6), keeps the output plot within an 80-column limit, and usually produces trading day peaks that are limited to a single frequency. When trading day peaks at 0.304 cycles/month were considered in the study, analogous changes to the neighboring frequencies were made, but these changes are no longer made in the actual X-12-ARIMA software package.

Methods of Obtaining Detection and False Alarm Rates

The issues relating to trading day spectral diagnostics were assessed quantitatively by establishing detection and false alarm rates using the a set of 42 economic series — for details of these series see Soukup and Findley (1999). In all series, a trading-day regression model lowered the value of the AIC by at least 1.0, so we assumed that at least weak trading day effects were present in all series. Thus, ideally, trading day effects would be visible in at least one spectral diagnostic of each of these series, and would be automatically detected in most. The method of establishing a detection rate was to run all series through X-12-ARIMA without estimating a trading day effect and examine the spectra of several output series. The detection rate we report is simply the fraction of series for which X-12-ARIMA found a significant spectral peak at the trading day frequency of interest.

To obtain series for determining false alarm rates that are plausible for real economic time series, we used bootstrapped replicates on some of the 42 series. These replicates were obtained by resampling the irregular component of a multiplicative seasonal, trend and irregular decomposition, $Y_t = T_t S_t I_t$. If I_t' is a series obtained by randomly re-sampling I_t with replacement then $Y_t' = T_t S_t I_t'$ is a *bootstrapped replicate* of the generating series Y_t. The replicate series retains the estimated trend and seasonal behavior of the original series while possessing none of the trading day behavior of the original series. From 50 re-samplings (i.e., 50 different I_t''s) we obtained 50 replicates (Y_t''s). The false alarm rate is defined as the fraction of the 50 replicates for which a visually significant spectral peak occurred at one of the trading day frequencies being considered in the designated output spectra.

Selection of an appropriate spectral estimator

Soukup and Findley (1999) provides a theoretical motivation for the choice of X-12-ARIMA default estimator and shows an example of the performance of different spectral estimators on an actual series. The program's default is an autoregressive model spectral estimator of model order 30. The motivation for this choice is that it can potentially produce the largest number of peaks possible, i.e. 30, in a plot with 61 frequencies. Thus, it has the greatest resolving power. When an model order is selected based on the AIC criterion the spectrum is somewhat smoother, but the contrast between the spectral amplitudes at the trading day frequencies and neighboring frequencies is weaker, and therefore not as suitable for automatic detection. Soukup and Findley (1999) show that the periodogram has trading day detection abilities as good as those of the AR(30) spectrum, but also has a greater false alarm rate. The detection and false alarm results in this report apply to spectra obtained with the AR(30) estimator.

Selection of the appropriate output series for spectral estimation.

X-12-ARIMA output spectra. In its default mode, X-12-ARIMA displays spectral estimates for appropriately differenced and transformed versions of the input series and for two output series that are final products of the seasonal adjustment decomposition. The spectrum of the input series usually contains very large seasonal peaks and is not of interest for trading day detection, because "leakage" from these peaks usually obscures the contribution of trading day components. The current version, in its default setting, provides the spectrum of the first differences of the seasonally adjusted series (the *dsa* spectrum) and the spectrum of the irregular component (*irr*). In this section, we describe the motivation for providing an additional, optional spectrum, namely the spectrum of the residuals of the regARIMA model (*rsd*).

The initial stimulus to examine the *rsd* spectrum came from results from simulated white noise and airline model series with no trading day effect. We had previously observed that earlier versions of X-12-ARIMA, which examined spectral peaks at all three trading day frequencies for visual significance, identified at least one visually significant trading day peak in the spectra of about 20% of simulated airline model series with no trading day component. The *rsd* output is produced by the one-step ahead forecast error filters of the fitted model, and none of X-12-ARIMA's deseasonalizing filters have been applied to produce it. This leads to output spectra that can be quite different from the *dsa* and *irr* spectra. The filtering that produces the *dsa* and *irr* output series imposes a certain structure on the output spectra that leads to more visually significant spectral peaks at some frequencies than others.

Gain function analysis to explain spurious peaks. When time-invariant, linear filters are applied, the term *gain function* is used for the mathematical function that describes the average frequency-dependent effects of the filters. Bell and Monsell (1992) analytically derived gain functions resulting from the symmetric filtering operations of the X-11 seasonal adjustment program (an ancestor of X-12-ARIMA —X-12-ARIMA contains the filters in X-11 and a few additional filters). General characteristics of the X-11 gain functions are 1) low amplitudes at seasonal frequencies and 2) spectral peaks at other frequencies that depended on the choice of trend and seasonal filters.

Since the gain function of a filter is proportional to the spectral density function of the outputs of the filter applied to white noise series, an estimate of the *irr* spectrum of white noise input, averaged over multiple realizations, will approximate the gain function if the series is long enough. Figure 2a shows the average of the *irr* spectra resulting from additive adjustment of 50 white noise series of length 600 with zero mean and a variance of 0.1. As with the analytic results, the obvious feature of this average spectrum is the low amplitude at seasonal frequencies and some spectral peaks between seasonal frequencies. When we introduce a more realistic input series (simulated airline data) and an appropriate series length for trading day estimation (N=96), and examine the *irr* spectra averaged over 50 realizations (Figure 2b), we see a similar overall structure. Note that the average spectrum has smaller amplitude at the frequencies immediately below the trading day frequencies than at the trading day frequencies (0.348 and 0.432 cycles/month) and higher amplitudes at the frequencies immediately after the trading day frequencies.

The average spectra do not have spectral peaks *at* the trading day frequencies, but only at the

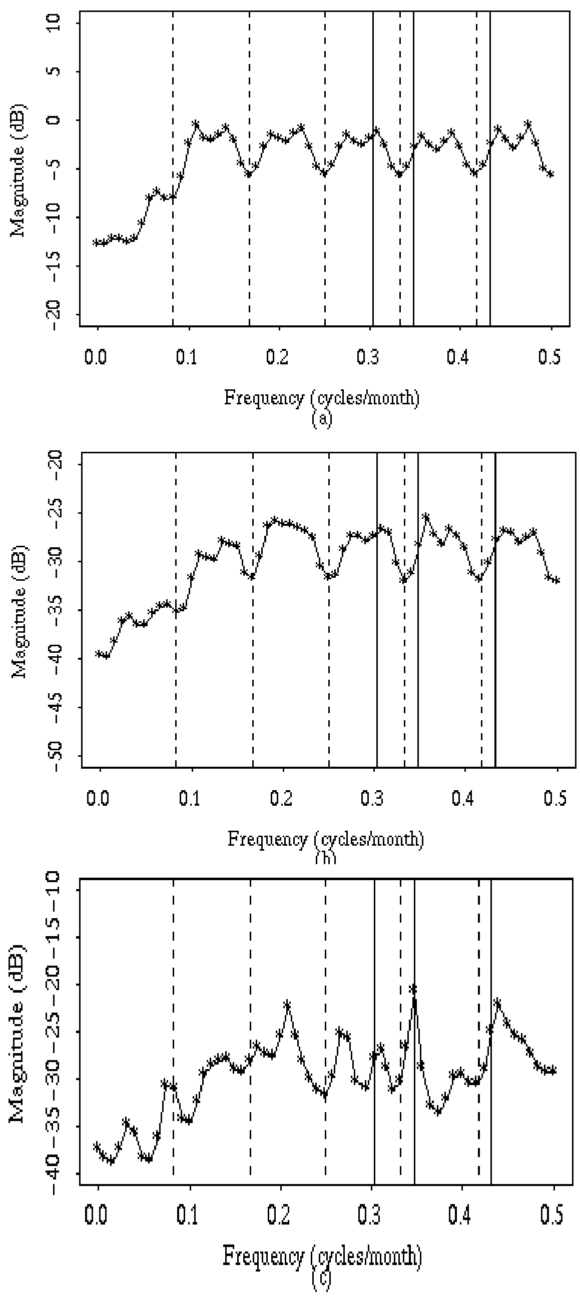

Fig. 2. (a) Logs of the averages of the irregular component spectra of 50 independent white noise series, (b) Logs of the averages of the irregular component spectra of 50 series from a typical (0 1 1)(0 1 1)^{12} "airline" model (θ = 0.4, Θ = 0.6). (c) Log of one of the irregular component spectra contributing to the average in (b).

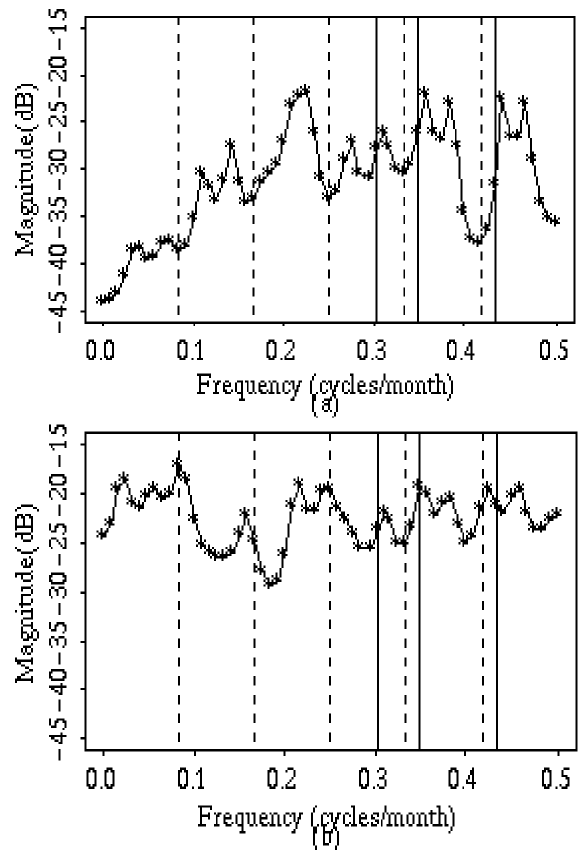

Fig. 3. Spectra of the irregular component (a) and the model residuals (b) for a foreign trade series.

frequencies just beyond them. An example of a single realization (i.e. one of the fifty series that was averaged to create Figure 2b) is given in Figure 2c. Note that the single realization has a visually significant spectral peak at 0.348 cycles/month, although there is no trading day component in the simulated series.[1] The features of the pseudo-gain function are sometimes evident in the output spectra from actual series. Figure 3a presents an *irr* spectrum from the seasonal adjustment of an Imports series with peaks and troughs at about the same locations as those of Figure 2b and similar patterns of increase through the trading day frequencies. Interestingly, this is a series for which the AIC comparison rather

[1] The reader cannot determine from the plots in Figure 2 whether a peak is visually significant, since the amount of log(magnitude), or decibels, that the visual criterion represents depends on the maximum and minimum of the data. The "6-star" criterion typically translates into a requirement that the amplitude at the trading day frequencies is at least 2-3 dB above the neighboring frequencies.

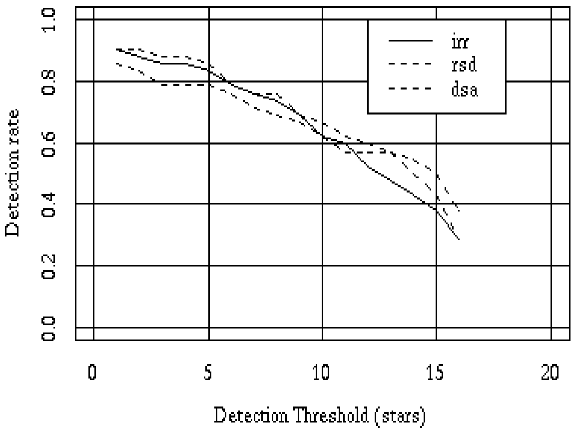

Fig. 4. Detection rates of trading day effects at 0.348 cycles/month in the 42 series as a function of the detection threshold used to define a visually significant peak.

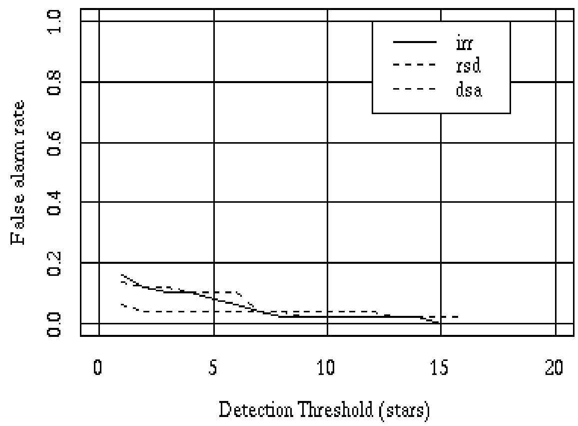

Fig. 5. False alarm rates from the dsa, irr and rsd spectra for a set of 50 bootstrapped versions of a foreign trade series vs. detection threshold at 0.348 cycles/month.

strongly supports trading day adjustment, although only the *rsd* spectrum (Figure 3b) has a trading day peak, albeit not visually significant. Filtering effects near 0.348 cycles/month may have obscured a weak effect.

In addition to detecting unmodeled trading day effects, the *rsd* spectrum can also indicate other kinds of model inadequacy by exhibiting significant peaks at *seasonal* frequencies. Methods of adjusting models to remove such peaks are given in Soukup and Findley (1999). It is sometimes necessary to reduce residual seasonal peak(s) in order to discover visually significant trading peaks in the residual spectrum.

Detection and false-alarm rates. The detection rates at 0.348 cycles/month are shown in Figure 4, as a function of the detection threshold in units of “stars”. From Figure 4, we see that the detection rates with a 6-star threshold for *dsa*, *rsd*, and *irr* are about 0.8. Overall detection rates at a 6-star threshold are much smaller for 0.432 and 0.304 cycles/month. At a “6-star” threshold, the detection rates for these two frequencies are 0.4 and 0.1 respectively.

A more detailed analysis (see Soukup and Findlay 1999) further improved our confidence in the performance of the diagnostics. For three series with missed detections, .the out-of –sample forecast error diagnostic of Findley et al. (1998) indicated that having a trading day regression variable in the model did not improve forecasting. Therefore, these missed detections may not represent any problem with respect to successful forecasting of data for the purpose of seasonal adjustment. Some difficulties in automatic detection can be addressed by user inspection of the spectrum or examination of the regARIMA model. Investigation of the model diagnostics can assist in resolving problems with missed detections in the *rsd* spectrum.

Using the bootstrap sets of 50 series, false alarm rates were derived as described above. False alarm rates at 0.348 cycles/month obtained by using an imports series as the generating series are shown in Figure 5. Overall, the *rsd* spectrum provided significantly lower false alarm rates — this can be seen for the lower values of the detection threshold in Figure 5.

Relative importance of the relevant frequencies for trading day analysis

To reduce false alarms, it was desirable to determine the number of frequencies that are necessary for detection of trading day effects in this set of series. Detections in 34 of the 42 series could be made using the *dsa* and *irr* spectra and the single frequency 0.348 cycles/month. Spectral peak analysis at 0.432 cycles/month only yielded detections in two additional series and that analysis at 0.304 cycles/month yielded no additional detections. In fact, false alarm rates and detection rates at 0.304 cycles/month were similar, so this frequency has little diagnostic value. We consider the effect of examining both 0.348 and 0.432 cycles/month below.

Selection of criteria to define visually significant spectral peaks.

In developing spectral diagnostics for X-12-ARIMA users, we considered a false alarm rate of 0.1 to be acceptable. From plots such as Figure 5, we found that this false alarm rate can be achieved at a detection threshold of 6-stars. Since the detection rate (Figure 4) decreases rapidly for higher values of the threshold, we have retained the “6-star” criterion that was used in previous versions of X-12-ARIMA.

Combining information from different diagnostics

Searching for visually significant peaks in the spectra of all three output series at both 0.348 and 0.432 cycles/month produced unacceptable (> 0.20) false alarm rates, so we used the results of the trading day study to select a subset of output series and frequencies to use in the default mode of X-12-ARIMA.

Use of both the *dsa* and *irr* output spectra does not produce more false alarms as spurious peaks in bootstrapped series nearly always have a significant peak for both spectra. Therefore, we can consider a detection to be the identification of a visually significant peak in either or both of the spectra. Although the *dsa* and *irr* spectra give very similar results, they are both used as default diagnostics because of our experience that seasonal adjusters with different technical background tend to differ which spectrum they prefer.

However, using the *rsd* spectrum in conjunction with the other spectra *does* increase the false alarm rate in the bootstrapping examples (in fact, it almost doubles it), so we have not attempted to consider spectral peaks in all three output series in default mode. The *rsd* spectra rather often have visually significant seasonal peaks that users with little modeling experience may not feel equipped to respond to. For these reasons, we chose to make automatic peak analysis of the model residuals optional. When it is specified, the results are reported independently of the tests of the *dsa* and *irr* spectra.

Consideration of false alarms also leads us to select a single frequency for analysis (0.348 cycles/month). Accepting spectral peaks at either 0.348 cycles/month or 0.432 cycles/month nearly doubles the false alarm rate, and, as previously stated, provides a minimal increase in detection rate.

CONCLUSIONS: SPECTRAL DIAGNOSTICS IN THE RELEASE VERSION OF THE X-12-ARIMA PROGRAM

- An order 30 autoregressive spectral estimator serves as the default. The user can also select the periodogram.
- Spectral estimates are computed from the last 8 years of data by default—this can be modified by the user.
- The default is to assess trading day variability based on the spectra from the 1) differenced, seasonally adjusted series and 2) irregular component. A visually significant peak in either spectrum will produce a warning message.
- The user also has the option to activate diagnostics based on the spectrum of the regARIMA residuals, which also provides a method of checking for modeling of seasonality from the original series.
- The visual significance criteria is set at 6 stars, which led to a false alarm rate of 10-15 percent in the trading day study for the autoregressive spectral estimator.

BIBLIOGRAPHY

Akaike, H. (1980), “Seasonal Adjustment by a Bayesian Modeling**”,** *Journal of Time Series Analysis***,** 1, 1-13.

Bell, W. R. and Monsell, B. C. (1992), “X-11 Symmetric Linear Filters and their Transfer Functions,” U.S. Census Bureau, Statistical Research Division, Research Report RR-92/15.

Cleveland, W. S. and Devlin, S. J. (1980), “Calendar Effects in Monthly Time Series: Detection by Spectrum Analysis and Graphical Methods,” *Journal of the American Statistical Association*, 75, 487-496.

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.

McNulty, M. S. and Huffman, W. E. (1989), “The Sample Spectrum of Time Series with Trading Day Variation,” *Economics Letters,* 31**,** 367-370.

Soukup, R.J. and Findley, D.F., “Using the Spectrum to Automatically Detect Trading Day Effects after Modeling or Seasonal Adjusstment,” U.S. Census Bureau, Statistical Research Division, Research Report RR-99/03.