Editor's Note: The following article was the invited address presented at the Joint Statistical Meetings, Chicago, Illinois, August 3–8, 1996.

New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program

David F. FINDLEY
Statistical Research Division, Bureau of the Census, Washington, DC 20233-9100 (dfindley@census.gov)

Brian C. MONSELL
Statistical Research Division, Bureau of the Census, Washington, DC 20233-9100 (bmonsell@census.gov)

William R. BELL
Statistical Research Division, Bureau of the Census, Washington, DC 20233-9100 (wbell@census.gov)

Mark C. OTTO
Office of Migratory Bird Management, U.S. Fish and Wildlife Service, Laurel, MD 20708-4016

Bor-Chung CHEN
Statistical Research Division, Bureau of the Census, Washington, DC 20233-9100

X-12-ARIMA is the Census Bureau's new seasonal-adjustment program. It provides four types of enhancements to X-11-ARIMA—(1) alternative seasonal, trading-day, and holiday effect adjustment capabilities that include adjustments for effects estimated with user-defined regressors; additional seasonal and trend filter options; and an alternative seasonal-trend-irregular decomposition; (2) new diagnostics of the quality and stability of the adjustments achieved under the options selected; (3) extensive time series modeling and model-selection capabilities for linear regression models with ARIMA errors, with optional robust estimation of coefficients; (4) a new user interface with features to facilitate batch processing large numbers of series.

KEY WORDS: Model selection; RegARIMA models; Trading-day models.

The Census Bureau's well known X-11 program was introduced in 1965 (Shiskin, Young, and Musgrave 1967). It was the product of over a decade of development beginning with "Method I" in 1954, followed by 12 experimental variants (X-0, X-1, etc.) of "Method II," culminating in X-11 (Shiskin 1978). X-11 followed in a long tradition of empirical smoothing and seasonal-adjustment procedures (Bell and Hillmer 1984), particularly the "ratio-to-moving-average" method of Macaulay (1931). The early Census Bureau methods were the first computerized seasonal-adjustment methods. X-11 became something of a standard that was used by statistical agencies around the world. Important features of X-11 that contributed to its widespread use are its treatment of atypical ("extreme") observations, its variety of moving averages for estimating evolving trend and seasonal components (and its methods and diagnostics for selecting among these), its refined asymmetric moving averages for use near the ends of time series, and its method for estimating trading-day effects.

Statistics Canada's X-11-ARIMA seasonal-adjustment program (Dagum 1980) contained all the capabilities of X-11 and provided important improvements. The most important is X-11-ARIMA's ability to extend the time series with forecasts and backcasts from autoregressive integrated moving average (ARIMA) models prior to seasonal adjustment. The use of forecast and backcast extensions results in initial seasonal adjustments whose revisions are smaller, on average, when they are recalculated after future data become available; see Huot, Chiu, Higginson, and Gait (1986) and Bobbitt and Otto (1990), for example. Extension overcomes deficiencies in the preliminary X-11 trend-estimation procedure at the ends of the series, especially in the first and last half-year. In the additive decomposition case, extension with optimal forecasts and backcasts for the half length of the symmetric seasonal filter used minimizes revisions in a mean squared sense. The history of this optimality property and an elegant derivation were given by Cleveland (1983).

Other X-11-ARIMA improvements include its more systematic and focused diagnostics for assessing the quality of its seasonal adjustments, which enable users to get good results more easily. And X-11-ARIMA offers diagnostics for comparing indirect and direct seasonal adjustments of series that are aggregates of multiple component series. X-11 did not calculate indirect adjustments.

The Census Bureau's new X-12-ARIMA program includes essentially all the capabilities of the latest version of X-11-ARIMA, X-11-ARIMA/88 (Dagum 1988), including all the capabilities of X-11. The major improvements in X-12-ARIMA address inadequacies of X-11 not targeted by X-11-ARIMA/88, as well as limitations in the modeling and diagnostic capabilities of X-11-ARIMA/88. These major improvements are the focus of this article; they are discussed and illustrated in Sections 1–5. We shall outline these sections, but first we briefly discuss the general structure of X-12-ARIMA.

Plans for X-12-ARIMA developed around the operation-flow diagram of Figure 1. This posits a regARIMA (linear regression model with ARIMA time series errors) modeling subprogram that can provide forecasts, backcasts, and prior adjustments for various effects before the seasonal-adjustment subprogram in the central box is invoked. The final box in Figure 1 represents a set of post-adjustment diagnostic routines that can be used to obtain indicators of the effectiveness of both the modeling and the seasonal-adjustment options chosen. The seasonal-adjustment methodology symbolized by the central box is an enhanced version of the X-11 methodology. A significant number of the enhancements were suggested by seasonal-adjustment experts at statistical offices and central banks in the United States, Canada, the United Kingdom, Germany, New Zealand, and Japan. The improvements introduced in X-11-ARIMA/88 were also influential.

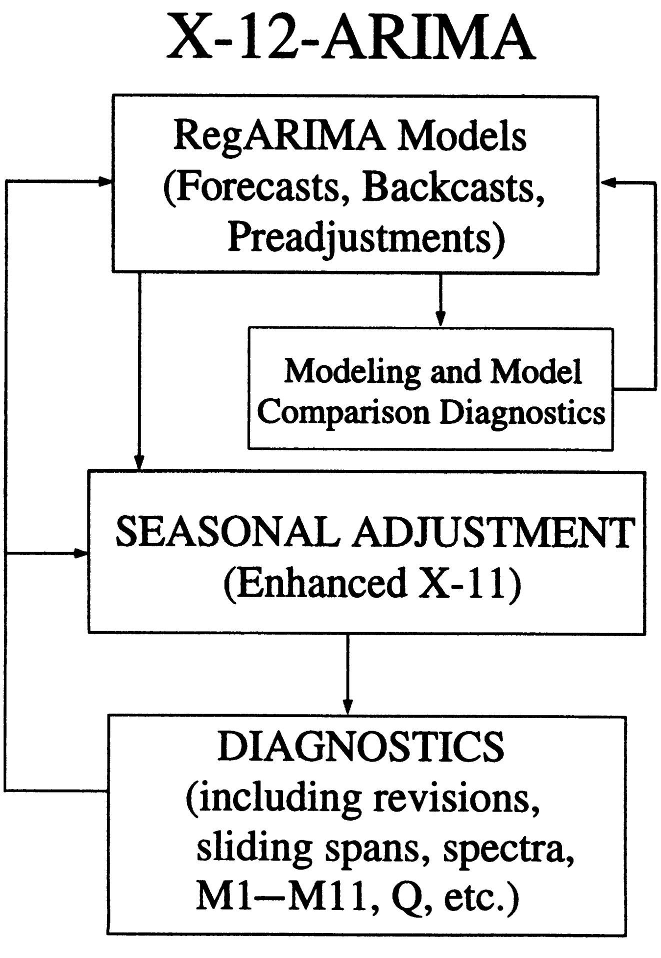

Figure 1. Flow Diagram for Seasonal Adjustment With X-12-ARIMA.

The major methodological improvements of X-12-ARIMA fall into three general groups that are discussed in the sections indicated—new X-11 adjustment options (Sec. 1), new diagnostics (Sec. 2), and new modeling capabilities emphasizing regARIMA modeling and model selection (Sec. 3). Section 4 illustrates how these modeling capabilities can address real problems that arise in seasonal adjustment. Section 5 briefly discusses another major improvement of X-12-ARIMA—its new user interface. Section 6 provides concluding remarks and an ftp address for obtaining the program. We now give a more detailed overview.

Section 1 discusses how new options in X-12-ARIMA provide additional flexibility in the basic seasonal-adjustment methodology of X-11 and X-11-ARIMA. New filter options include a longer seasonal moving average, allowance for user specification of Henderson trend filters of any (odd) length, and slight modifications to some of X-11's asymmetric moving averages so that more are derived from a single optimization principle (outlined in the Appendix). The program also provides a "pseudo-additive" decomposition that has been found useful for series with periodically small or zero values. Finally, improvements were made in how trading-day and other regression effects, including user-defined effects (a new capability), are estimated from a preliminary version of the irregular component. (Alternatively, such effects can be estimated directly from the observed time series using the program's regARIMA modeling capabilities.)

Section 2 discusses significant diagnostic capabilities X-12-ARIMA provides beyond those of X-11 and X-11-ARIMA. These include spectrum estimates for detection of seasonal and trading-day effects and also sliding spans (Findley and Monsell 1986; Findley, Monsell, Shulman, and Pugh 1990) and revisions history diagnostics for assessing the stability of seasonal adjustments. We were motivated in this development by our experience that, although the diagnostics of X-11-ARIMA are an important advance beyond those of X-11, they sometimes fail to identify series that cannot be satisfactorily adjusted. They also sometimes give an incorrect indication as to whether the direct or an indirect adjustment of an aggregate series should be preferred (see the examples in these articles).

Other important features of X-12-ARIMA derive from its regARIMA modeling capabilities; these are discussed in Section 3 and illustrated in Section 4. X-11-ARIMA lacks the capability to add regression effects to the models used for forecast extension. Although preadjustment for trading-day and other regression effects estimated from irregulars (the approach taken by X-11-ARIMA/88) may usually do as well for point forecasts, this approach is more limited than use of regARIMA models, as our later discussion will show. X-12-ARIMA's use of regARIMA models can potentially improve forecasts and backcasts and, through its outlier detection capabilities, help robustify model parameter estimates and model forecasts against additive outliers and level shifts.

The focus in Sections 3 and 4 is not, however, on advantages of using regARIMA models for forecast extension. Rather, it is on a variety of important *direct* applications for regARIMA models in seasonal adjustment. These include the following: (1) regARIMA models for trading-day and holiday effects (Bell and Hillmer 1983) provide more reliable diagnostics for the presence of such effects than do F statistics of regression models fit to the irregular component of the seasonal decomposition as in X-11 (see Secs. 1.4 and 3.3). (2) Chang and Tiao (1983) and Bell (1983) showed how regARIMA models can be used to detect additive outliers (AO's) and level shifts (LS's). (See also Chang, Tiao, and Chen 1988, Sec. 3.2, and Appendix C.) Allowance for such outlier effects in a model can help protect the model's coefficient estimates and forecasts against corruption (Burman and Otto 1988; Ledolter 1989). (3) The ability to handle AO's provides a capability for dealing with small amounts of missing data: Bruce and Martin (1989) observed that exact treatment of missing observations is approximately the same as replacing missing observations by their estimated AO effects (see Sec. 4.2). (4) Preadjustment for LS's (before seasonal adjustment by X-11) can overcome one of the most troubling common sources of difficulty for X-11—the inability of its trend filters to track sudden changes in level. For example, Figure 2 shows the graph of the series of net income from U.S. retail sales and the modified series resulting from the use of an LS regressor in a regARIMA model of the log series to remove the precipitous drop in level in the first quarter of 1982. (This drop was caused by a governmental action, called the Paperwork Reduction Act, that took smaller companies out of the survey universe.) (5) regARIMA models can be used to test for changes in seasonal pattern, in trading-day effects, and so forth. Note from Figure 2 that the net income series from the reduced universe appears to have a different, more stable seasonal pattern than the pre-1982 series from the larger universe. In Section 4.1, we shall show how regARIMA models can be used to test this series for a change in seasonal pattern.

To complement its regARIMA modeling capabilities X-12-ARIMA also provides extensive model-selection diagnostics, including recently developed diagnostics based on out-of-sample forecast performance. The need for such diagnostics in seasonal adjustment will become clear in Sections 3 and 4: Many of the model comparisons that arise naturally within the rich class of regARIMA models appropriate for time series with seasonal and calendar effects are not addressed by standard statistical tests.

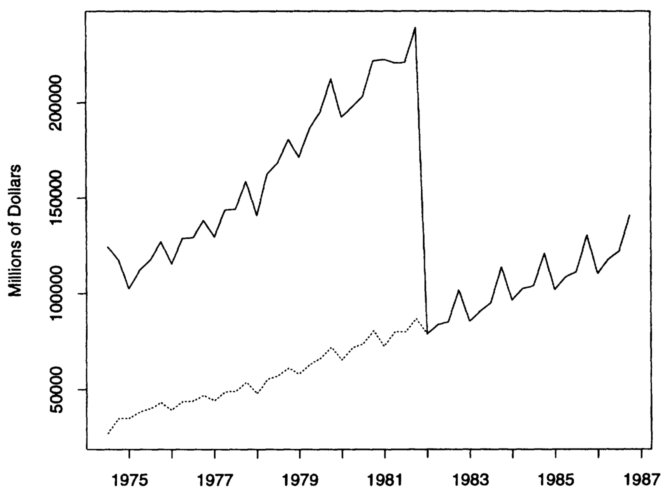

Figure 2. Net Income (Sales − Costs) From U.S. Retail Sales With and Without LS Adjustment: ———, Original; · · ·, Adjusted.

Finally, Section 5 briefly illustrates the new user interface of X-12-ARIMA. This interface, which uses a simple, self-descriptive command language, greatly simplifies the program's use in both production and research environments.

1. NEW X–11 ADJUSTMENT OPTIONS

We begin with a review of the decomposition procedures of X-11. This serves as background for the discussion of the program's new seasonal and trend moving average options in Section 1.2 (and Appendix B) and its new decomposition option in Section 1.3. The final Section 1.4 explores issues surrounding the estimation of regression effect components, such as trading-day components, from the irregulars. It includes a derivation of X-11's deseasonalized model for multiplicative trading-day effects and discussion of how the derivation's model-deseasonalization approach is extended in X-12-ARIMA to the additive and other decompositions and to other regression effects.

1.1 Decompositions for Seasonal Time Series

The basic seasonal-adjustment procedure of X-11 and X-11-ARIMA decomposes a monthly or quarterly time series into a *product* of (estimates of) a *trend* component, a *seasonal* component, and a residual component, called the *irregular* component. Such a *multiplicative decomposition* is usually appropriate for series of positive values (sales, shipments, exports, etc.) in which the size of the seasonal oscillations increases with the level of the series, a characteristic of most seasonal macroeconomic time series. Under the multiplicative decomposition, the *seasonally adjusted series* is obtained by dividing the original series by the estimated seasonal component. The values of the estimated seasonal component are called *seasonal factors.* There is also an analogous *additive decomposition,* which decomposes the series into a sum of trend, seasonal, and irregular components, with the seasonally adjusted series obtained by subtracting away the estimated seasonal component. Although analyses of the properties of X-11 often focus on the additive decomposition (e.g., Cleveland and Tiao 1976; Wallis 1982; Ghysels, Granger, and Siklos 1996), the multiplicative decomposition is used far more frequently.

X-12-ARIMA retains the basic multiplicative and additive decompositions. Moreover, in common with X-11-ARIMA, the X-12-ARIMA program can calculate a second multiplicative decomposition by exponentiating the additive decomposition of the logarithms of the series being adjusted. This is called the *log-additive decomposition.* It is used mainly for research purposes, because it requires a bias correction for its trend estimates (due to geometric means being less than arithmetic means) as well as a different calibration for extreme value identification based on the lognormal distribution. Section 1.3 describes a new, fourth decomposition, the *pseudo-additive decomposition,* that was developed at the U.K. Central Statistical Office.

Following X-11, the default scheme of X-12-ARIMA for obtaining the various three-component decompositions of a time series is a three-stage procedure. This is presented in Appendix A for the simplified situation of a series with no extreme values. It is further assumed that the series has been extended far enough by forecasts and backcasts that the data required by the formulas in Appendix A are available for all months t in the span of the observed series. The only calculations whose role may not be clear are those of Step (d) in Stages 1 and 2. Their effect is usually to make 12-month totals of the adjusted series be close to the corresponding totals of the unadjusted series. [The log-additive decomposition is not explicitly presented in Appendix A, because its computations parallel those of the additive decomposition. In X-12-ARIMA, the log-additive decomposition includes a bias-correction due to Thomson and Ozaki (1992), which is applied to the exponentiated trend component.]

1.2 X-11 Seasonal Adjustment and Trend Filters

1.2.1 Symmetric Seasonal Filters. The symmetric seasonal moving averages used in step (c) of Stages 1 and 2 in Appendix A have a similar structure: They are simple 3-term moving averages, of simple averages of odd length, $2n+1$, of SI ratios (detrended series values) from the same calendar month as month t,

$$S_t^{3\times(2n+1)} = \frac{1}{3}\left(S_{t-12}^{2n+1} + S_t^{2n+1} + S_{t+12}^{2n+1}\right)$$

with

$$S_t^{2n+1} = \frac{1}{2n+1}\sum_{j=-n}^{n} \mathrm{SI}_{t+12j}.$$

$S_t^{3\times(2n+1)}$ is referred to as the $3\times(2n+1)$ seasonal moving average or seasonal filter. In the default setting of X-11, the 3×3 seasonal moving average is used at step (c) of Stage 1 and the 3×5 seasonal moving average at step (c) of Stage 2. X-12-ARIMA and X-11-ARIMA/88 differ from X-11 in that step (c) of Stage 2 uses a criterion due to Lothian (1984) to select from among four filters—the $3\times 3, 3\times 5$, and 3×9 moving averages and the average of all SI ratios from the same calendar month as t, the *stable seasonal average.* Optionally, in all three programs the user can specify any of these moving averages for use in any calendar month. The chosen averages are then used in step (c) of both Stages 1 and 2. In X-12-ARIMA, there is also an optional 3×15 seasonal moving average. This filter was used in X-10 and in a customized version of X-11 at the German Bundesbank as an alternative to the stable seasonal average for series of length at least 20 years. The appropriateness of longer seasonal moving average filters has been suggested by researchers investigating ARIMA-model-based signal-extraction seasonal adjustments. (See Bell and Hillmer 1984, pp. 308–309).

1.2.2 Symmetric Trend Filters. The symmetric Henderson trend (or "trend-cycle") moving averages used in step (a) of Stages 2 and 3 will perfectly reproduce a cubic polynomial. Moreover, their "weights" $h_j^{(2H+1)}$ change with j as smoothly as possible in a sense we explain in Appendix B, where their formula is given. In X-11 and X-11-ARIMA/88, either the user or the automatic "variable trend cycle curve routine," discussed at the end of Appendix B, chooses among Henderson filters of length 9, 13, and 23.

In X-12-ARIMA, the automatic selection procedure is the same, but the user can alternatively specify any odd-number length $2H+1$. The specified Henderson filter is then used in step (a) of both Stages 2 and 3. In recent years the Australian Bureau of Statistics has been using 15-term and 17-term Henderson filters in their customized version of X-11 as alternatives to the 13-term filter.

Figure 3 displays the squared gain functions (up to frequency $\lambda = .25$)

$$\left|\sum_{j=-H}^{H} h_j^{(2H+1)} e^{-i2\pi j\lambda}\right|^2 = \left(h_0^{(2H+1)} + 2\sum_{j=1}^{H} h_j^{(2H+1)}\cos 2\pi j\lambda\right)^2, \quad 0 \le \lambda \le .5,$$

of the 13- and 17-term Henderson filters ($H = 6, 8$), together with the squared gain functions of the resulting X-11 additive-decomposition trend-component extraction filters for a monthly series. These trend extraction filters are obtained by combining (convolving) all of the additive decomposition's linear operations in Stages 1–3 used to obtain the final trend estimates $T_t^{(3)}$. Recall that the product of the squared gain function and the spectral density of the filter's input series gives the spectral density of the output series when the input series is stationary (see Koopmans 1974, p. 86). Thus, at frequencies at which the gain function is close to 0, the variance components of the input series are suppressed. Figure 3 shows that the Henderson filters suppress the higher-frequency components of a stationary input series and essentially preserve the magnitudes of the components whose frequency is close enough to 0. A similar effect can be expected with nonstationary input series (see Oppenheim and Schafer 1975, p. 110). As Figure 3 shows, the squared gain function of the 13-term Henderson filter has substantial power beyond the first seasonal frequency 1/12. This results in the peak just beyond this frequency in the squared gain of the associated trend extraction filter. (The preceding dip down to 0 at 1/12 comes from the seasonal-adjustment operations applied before the application of the Henderson trend filters.) Because of this peak, it has been claimed that X-11's final trend estimate from the 13-term Henderson filter exaggerates short-term cyclical behavior (Schips and Stier 1995). The 17-term Henderson filter is the shortest that does not result in a significant peak beyond the first seasonal frequency in the squared gain function of the trend extraction filter.

1.2.3 Asymmetric Filters. Now we consider briefly the asymmetric filters used near the beginning and end of a

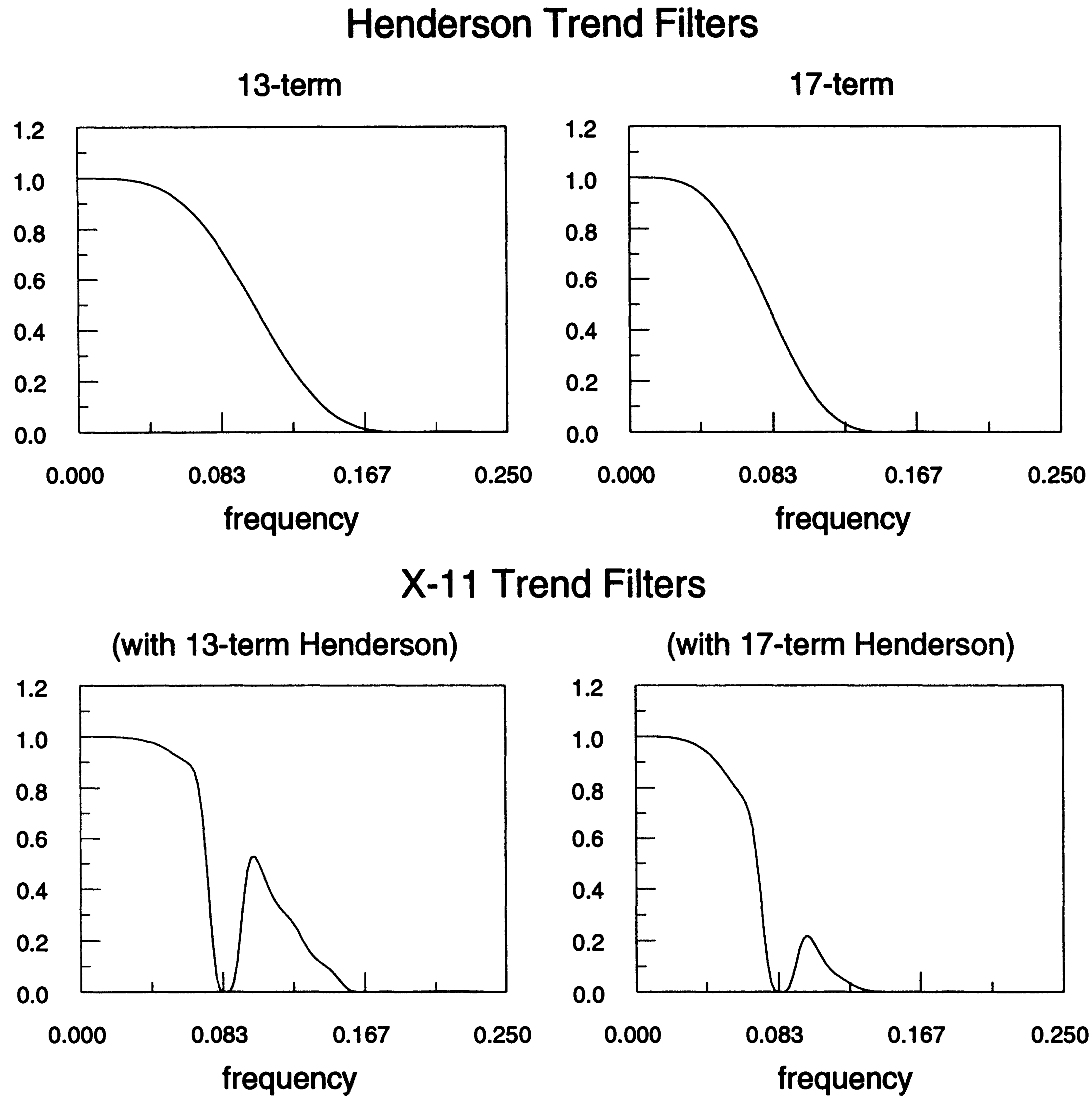

Figure 3. Henderson and Additive X-11 Trend Filter Squared Gains for Frequencies [0, .25]. The deseasonalization prior to final trend estimation produces the 0 at $\lambda = 1/12$ in the trend filter gains, with the result that some higher-frequency components will be suppressed less than lower-frequency components near $\lambda = 1/12$. When the 17-term Henderson filter is used, the second peak in the trend filter's squared gain is quite small. Consequently, trends from this filter have negligible susceptibility to displaying anomalous higher-frequency oscillations compared to trends obtained by use of the 13-term Henderson filter.

series that is not extended, or not fully extended, by forecasts and backcasts. In X-12-ARIMA, the coefficients of the asymmetric filters associated with the 3×9 seasonal filter are slightly modified versions of the filters in X-11 and X-11-ARIMA. The modifications were done to obtain filters that are derivable from an unpublished optimization principle developed by Musgrave (1964) that is detailed in Appendix B. There it is explained that the asymmetric replacements for both the 3×9 seasonal filter and the Henderson filters are determined by values chosen for a certain "noise-to-signal ratio." For the Henderson filters, the X-12-ARIMA user can change this ratio to obtain different asymmetric filters. This is one of the program's "rarely used options," intended for the researcher or specialist rather than for the general user. An unpublished formula of M. Doherty for the exact solution of Musgrave's optimization [given as (B.3) in Appendix B] made it easy for us to implement both this option and the option to allow the user to specify Henderson filters of any odd length (replaced by appropriate asymmetric filters near the ends of the series).

X-12-ARIMA can produce a smoothed version ("trend") of a *nonseasonal* series through application of any of its Henderson trend filters directly to the input series, or to the series modified by regression preadjustments (if, for example, there are outliers).

1.3 The Pseudo-Additive Decomposition

The pseudo-additive decomposition has the form $Y_t = T_t(S_t + I_t - 1) = T_t(S_t - 1) + T_t I_t$. The algorithm for its calculation is summarized in Appendix A. According to M. Baxter of the U.K. Office for National Statistics, where it has been used for almost 20 years, this procedure was developed for seasonally adjusting nonnegative time series that have quite small, possibly zero values in the same month or months each year. Such months have seasonal factors close to 0, and dividing by such very small factors produces unsatisfactory results. Adjustment of these months by subtraction of an estimate of $T_t(S_t - 1) \approx -T_t$ is more likely to give an estimate close to the trend of the series, because $Y_t \approx 0$. Agricultural products that are available only at certain times of year can give rise to such series. So can institutional behavior such as the shutdown of factories because of summer vacations, as the graph of an Italian car-production series in Figure 4 illustrates.

Figure 4 shows both the additive and the pseudo-additive adjustments of recent years of this series. The first impression might be that the additive adjustment is reasonable except in August of the last year. In this month, the additively adjusted series incorrectly suggests that a very low level of production, essentially unchanged from the two preceding Augusts, represents a substantial increase. The pseudo-additive adjustment provides a more plausible, neutral value for this month. It also presents the Augusts of 1989 and 1990 as having significantly increased production, which they do have relative to other Augusts, a feature not indicated as clearly by the additive adjustment. When we calculated the revisions history diagnostics (presented in Sec. 2.2) for both adjustments of this series, however, the results (not given in this article) showed that the pseudo-additive adjustments of Augusts are much more likely than the additive adjustments to experience large revisions as future data are added to the series. (Multiplicative adjustments are more volatile still and give implausible adjustments.)

It is an unusual aspect of the pseudo-additive decomposition that the adjustment quantities removed by the adjustment operation are not the level-independent quantities S_t as in the other decompositions but are instead the level-dependent quantities $T_t(S_t - 1)$; see the steps (e) in Appendix A. Thus, difficulties in estimating T_t at the ends of series (see Sec. 2.2) can be expected to increase the variability of the adjustments there. This decomposition can yield negative adjustments for nonnegative series.

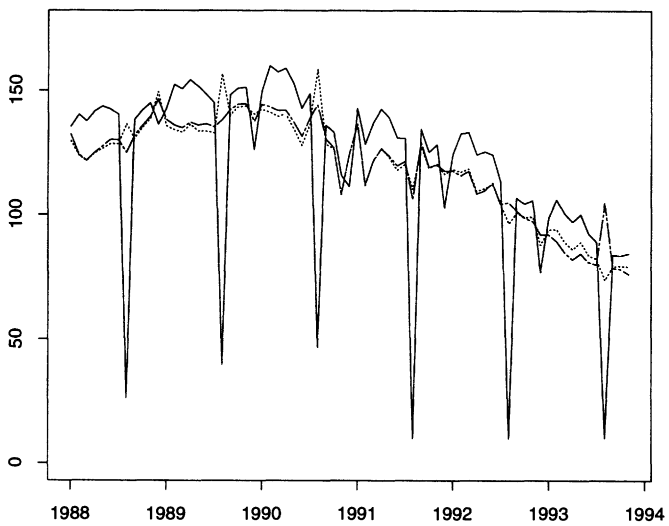

Figure 4. Italian Car Production (———) With Additive (- - -) and Pseudo-Additive (· · ·) Seasonal Adjustments. In Italy August is the main month for vacationing, and the resulting very low levels of car production make this series unsuitable for multiplicative adjustment. The graphs show that the pseudo-additive adjustment more accurately reflects the increased production in August of 1989 and 1990. Moreover, unlike the additive adjustment, the pseudo-additive adjustment does not suggest that the very low August 1993 value, which differs little from the August 1991–1992 values, represents a substantial increase.

1.4 Extracting Regression-Effect Components From the Irregulars

The concern in this section is with the estimation of calendar effects and other effects by means of regression models for the irregular component. Trading-day effects are estimated this way in X-11 and X-11-ARIMA. More general regression modeling of the irregular component is possible in X-12-ARIMA, which offers Easter-holiday and other calendar-effect regressors, as well as indicator variables to identify extreme irregulars and diminish their influence when other regression effects are estimated. User-defined regression models can also be estimated. Alternatively, X-12-ARIMA can estimate all of these effects by means of regARIMA models for the observed time series. This latter approach has important advantages, which we shall elaborate later, for making inferences about the regression effects. Our decision to retain and enhance the older approach of modeling the irregulars was motivated by its historical success, by practical considerations mentioned later, and by the requests of statistical agencies and central banks in different countries who wish to be able to estimate their own country-specific working-day and holiday effects in this way.

The irregulars series, being the residual component after deseasonalization and detrending, is a natural series from which to estimate further components. Being an almost uncorrelated series, it has the appealing simplicity of being a candidate for ordinary least squares (OLS) regression estimation of additional components. There is a complication, however: Its deseasonalized and detrended nature implies that regression models for the irregulars should also be deseasonalized and detrended. In Section 1.4.1 we illustrate how this is done for a natural model of trading-day effects. We obtain thereby both a derivation of the trading-day model of Young (1965) used by X-11 and X-11-ARIMA and also a derivation of X-12-ARIMA's default regARIMA regression model for trading-day effects estimated from the logarithms of the observed time series. It is an important feature of this model that the effect of month length is known in advance and does not require estimation. The estimation of other calendar effects from the irregulars is discussed briefly in Section 1.4.2.

1.4.1 Trading-Day Effects and Young's Model. We begin with a brief explanation of trading-day effects. In addition to seasonal effects, monthly time series that are totals ("flows") of daily economic activities are often influenced by the weekday composition of the month. The presence of such an effect is revealed when the series values for a given calendar month depend in a consistent way over time on which days of the week occur five times in the month. With retail grocery sales, for example, there is usually lower volume on Mondays, Tuesdays, and Wednesdays than on days later in the week. Thus, sales in March, say, will be relatively lower in a year in which March has an excess of early weekdays and higher when March has five Fridays and Saturdays. To a lesser extent, series of "stocks" measured on the same day each month, such as inventories or unfilled orders as of the last day of the month, are sometimes sensitive to the day of the week on which their value is obtained. Finally the average daily effect in flow series can give rise to a length-of-month effect. Because the length of February is not the same every year, this effect is not completely absorbed by the seasonal component. The residual effect left in Februaries is called the Leap Year effect.

Recurring weekday composition effects in monthly (or quarterly) economic time series are called *trading-day effects.* Flow trading-day effects were discussed by Young (1965) and stock trading-day effects by Cleveland and Grupe (1983) (see also Bell 1984; Chen and Findley 1996a). Like seasonal effects, trading-day effects can make it difficult to compare series values across months or to compare movements in one series with movements in other series. For this reason, when estimates of trading-day effects are statistically significant, they are usually adjusted out of the series when seasonal adjustment is performed. In this adjustment context, they form a fourth decomposition component, the *trading-day* component.

To obtain a model for trading-day effects in monthly flow series, suppose that the jth day of the week has effect α_j, where $j = 1$ designates Monday, $j = 2$ Tuesday, $\ldots, j = 7$ Sunday. Then if D_{jt} denotes the number of occurrences of day j in month t, the cumulative effect for the month will be $\sum_{j=1}^{7} \alpha_j D_{jt}$. Set $\bar{\alpha} = \sum_{j=1}^{7} \alpha_j/7$ and $N_t = \sum_{j=1}^{7} D_{jt}$, the length of month t. Because $\sum_{j=1}^{7}(\alpha_j - \bar{\alpha}) = 0$, we have

$$\begin{aligned}\sum_{j=1}^{7} \alpha_j D_{jt} &= \bar{\alpha} N_t + \sum_{j=1}^{7} (\alpha_j - \bar{\alpha}) D_{jt} \\ &= \bar{\alpha} N_t + \sum_{j=1}^{6} (\alpha_j - \bar{\alpha})(D_{jt} - D_{7t}), \qquad (1)\end{aligned}$$

a decomposition into a length-of-month effect and the net effect of the daily contrasts $(\alpha_j - \bar{\alpha})$. Replacing D_{jt} in the center expression of (1) by $D_{jt} - 4$ changes nothing and makes it clear that this second component is equal to the sum of the $(\alpha_j - \bar{\alpha})$ for those weekdays j that occur five times in month t. We shall obtain a deseasonalized and level-neutral version of (1) by removing calendar-month means.

The monthly calendar repeats itself over any 28-year cycle (until the year 2100 when the 29th of February is omitted). Consequently, the variables D_{jt} are periodic with period 336 ($= 12 \times 28$) months, and the calendar-month means $(1/28) \sum_{k=1}^{28} D_{j,t+12k}$ have the same value for all t and j. It follows that the 28-year calendar-month means of the difference variables $D_{jt} - D_{7t}$ on the right in (1) are 0. This implies that the final expression in (1) involving these differences has a seasonal component of 0 and also a level component (336-month mean) of 0. Thus the seasonal and level components of (1) reside in the calendar-month means of $\bar{\alpha} N_t$. Because $N_{t+48} = N_t$, these are given by $\bar{\alpha} N_t^*$ with $N_t^* = (1/4) \sum_{k=1}^{4} N_{t+12k}$. How these components are removed from the model depends on the type of seasonal decomposition used to obtain the irregulars.

For the usual case of a *multiplicative decomposition,* we then deseasonalize and detrend the trading-day effect by dividing (1) by $\bar{\alpha} N_t^*$. Setting $\beta_j = (\alpha_j/\bar{\alpha}) - 1$, this yields

$$\frac{N_t}{N_t^*} + \sum_{j=1}^{6} \beta_j \left(\frac{D_{jt} - D_{7t}}{N_t^*} \right) = \frac{\sum_{j=1}^{7} (\beta_j + 1) D_{jt}}{N_t^*}. \qquad (2)$$

This is the formula for trading-day effects given without derivation by Young (1965). With $\hat{I}_t$ denoting a preliminary estimate of the irregular component, the X-11 program and its direct descendants estimate $\beta_1, \ldots, \beta_6$ (and thus $\beta_7 = -\sum_{j=1}^{6} \beta_j$) by the OLS fitting of the regression model

$$N_t^* \hat{I}_t - N_t = \sum_{j=1}^{6} \beta_j (D_{jt} - D_{7t}) + e_t. \qquad (3)$$

In X-12-ARIMA, the analog of (3) for the *additive decomposition* is obtained by subtracting $\bar{\alpha} N_t^*$ from (1). This yields

$$\hat{I}_t = \beta_0 (N_t - N_t^*) + \sum_{j=1}^{6} \beta_j (D_{jt} - D_{7t}) + e_t, \qquad (4)$$

where now $\beta_0 = \bar{\alpha}$ and $\beta_j = \alpha_j - \bar{\alpha}$ for $1 \le j \le 6$. Thus, in the additive case, seven coefficients must be estimated instead of six. In X-11 and X-11-ARIMA, the regressor $N_t - N_t^*$ is not used. [Young told us that he agrees that X-11 should have used (4).] For the *pseudo-additive decomposition,* letting $\bar{N} = (1/48) \sum_{k=1}^{48} N_{t+k} = 30.4375$ (the average month length), it can be shown that deseasonalization and detrending lead to trading-day factors of the form $1 + (N_t - N_t^*)/\bar{N} + \sum_{j=1}^{6} \beta_j (D_{jt} - D_{7t})/\bar{N}$.

Finally, to motivate the regARIMA regression model considered in Section 3 for logarithms of the observed series, we need the trading-day factor formula for the *log-additive decomposition.* Taking the logarithm on the left in (2) and using $\log(1 + x) \approx x$, one obtains

$$\begin{aligned}&\log \left\{ 1 + \frac{N_t - N_t^*}{N_t^*} + \sum_{j=1}^{6} \beta_j \left(\frac{D_{jt} - D_{7t}}{N_t^*} \right) \right\} \\ &\qquad \approx \frac{N_t - N_t^*}{N_t^*} + \sum_{j=1}^{6} \beta_j \left(\frac{D_{jt} - D_{7t}}{N_t^*} \right). \qquad (5)\end{aligned}$$

The summation on the right in (5) has 28-year-calendar-month means equal to 0 and thus has no seasonal or trend. Hence, it can be taken as the regression expression for trading-day effects in the additive irregular component of the logs of the time series being adjusted. Exponentiating, using $e^x \approx 1 + x$, and setting $\beta_j' = \beta_j/\bar{N}$, we obtain both an exact and an approximate trading-day factor formula for the log-additive case,

$$\exp\left(\frac{N_t - N_t^*}{N_t^*}\right) \exp\left\{\sum_{j=1}^{6} \beta_j \left(\frac{D_{jt} - D_{7t}}{N_t^*}\right)\right\}$$

$$\approx \frac{N_t}{N_t^*} \exp\left\{\sum_{j=1}^{6} \beta_j'(D_{jt} - D_{7t})\frac{\bar{N}}{N_t^*}\right\}. \quad (6)$$

The approximating second expression, further simplified by treating $\bar{N}/N_t^*$ as if it were equal to 1.0, defines X-12-ARIMA's default regARIMA model factors for trading-day effects in the logarithms of the observed time series. [See (14) in Sec. 3.]

There is an alternative to the deseasonalization and detrending approach just illustrated that is most appealing with additive decompositions. This is the "matched filtering" procedure used for the trading-day regression in the SABL seasonal-adjustment program, described by Cleveland and Devlin (1982) and Cleveland (1983). In this procedure, the "irregular filter" is applied to the regressors D_{jt} in (1) prior to the irregulars series being regressed on them. (The irregular filter corresponds to applying all of the linear operations of Appendix A used to calculate the irregular component of a series.) Because this is obviously another way to deseasonalize and detrend the trading-day effect, this may accomplish much the same thing as the procedure discussed previously. The matched filtering approach is also plausible when an additive decomposition is obtained for a transformed version of the original series, as in the log-additive case. It is unclear how matched filtering applies to the multiplicative or pseudo-additive decompositions.

1.4.2 Other Regressors and Robustification of the Regressions Against Additive Outliers. Since the early 1970s, the versions of the X-11 program used at the Census Bureau also obtained estimates of the effects on retail sales of Easter and of the moving U.S. holidays Labor Day and Thanksgiving. Easter effects, for example, can increase retail sales of clothing in the week or so prior to Easter or decrease factory shipments in certain industries a few days before Easter. [The X-11 procedure for estimating Easter effects was detailed by Chen and Findley (1996b).] X-11-ARIMA/88 estimates an Easter effect from the series of irregulars using a different procedure, described by Dagum, Huot, and Morry (1988). In X-11 and X-11-ARIMA/88, the trading-day and holiday effects are estimated iteratively rather than simultaneously.

In X-12-ARIMA, these effects can be estimated simultaneously from the irregulars. With such a diverse set of regressors, however, the deseasonalization and detrending procedure exemplified previously can lead to *nonlinear* regression models (Chen and Findley 1996a). These can sometimes be linearized easily. [For example, the approximation in (5) is a linearization.] The coefficient estimates can be protected against the effects of extreme irregulars by means of a procedure like the one discussed in Section 3 and Appendix C.

Finally, X-12-ARIMA allows user-defined regressors in the irregulars regression. These regressors may need to be deseasonalized and detrended before being input to the program.

1.4.3 Model the Irregulars or Model the Original Series? Instead of modeling the irregulars series, one can model the original series, as we discuss in Section 3. This has important advantages for making statistical inferences about calendar and other regression effects. Implicit in the use of OLS regressions and the associated F tests of significance is the assumption that the irregular component is a series of constant-variance, independent variates. Rather frequently (with 14 of 71 series in the trading-day modeling study described by Chen and Findley 1993, 1996a), the regression F statistic from a regression model of the irregulars has a spuriously significant value in tests of the null hypothesis of no effect even at the .01 level of significance—an indication that this implicit assumption is often not adequately satisfied. In fact, some autocorrelation is typically found in the irregulars: The sample autocorrelations between irregulars a year apart are almost always negative and larger in magnitude than all other sample autocorrelations (often being close to $-.2$). Moreover, it is clear from the trend filter gain functions that X-11's relatively short-term trends cannot fully capture long-term correlation in the data if it exists. Additionally, there is heteroscedasticity near the ends of the irregulars series because of the time-varying asymmetric filters used to obtain the decomposition near the ends of the series being adjusted. To detect a spuriously significant F statistic, the spectrum and sliding-spans diagnostics discussed in Section 2 can be used, as can the regARIMA model diagnostics that will be discussed in Section 3.2.

One might expect that estimates of calendar and other regression effects would also be better when these come from regARIMA models, both because these models account for the correlation structure of the observed series and because they model the effects directly rather than as a residual component identified after seasonal and trend estimation. We have not found this to be universally true, however. In Section 4.3 we shall show how out-of-sample forecasting performance can be used to demonstrate the superiority, inferiority, or rough equivalence of calendar-effect estimates from regARIMA models versus those from OLS regression models of a preliminary irregular component.

The better inference properties and the typically equivalent or better performance of estimates from regARIMA models lead us usually to prefer using a regARIMA model of the original series to estimate regression effects. There are some series that cannot be modeled well by regARIMA models, however, due, for example, to frequent changes in variability or to erratic trend movements over the course of the series that require more sophisticated detrending procedures than differencing. Finally, many people worldwide who are responsible for producing seasonal adjustments do not have the necessary training to develop regARIMA models for their series.

2. NEW DIAGNOSTICS

X-12-ARIMA provides the diagnostic tables of X-11 and X-11-ARIMA, as well as the M1–M11 quality-control statistics of X-11-ARIMA. It also has important additional diagnostics, including spectrum estimates for the presence of seasonal and trading-day effects (see Sec. 2.1) and the sliding spans and revisions history diagnostics of the stability of seasonal adjustments (see Sec. 2.2). The sliding spans and revisions histories are directly interpretable, whereas M1–M11 are indirect measures, in some cases very indirect, of data features known to be troublesome for the X-11 methodology. Most of the M1–M11 statistics can be calculated for short time series, however, something impossible for the current stability diagnostics of X-12-ARIMA.

2.1 Using Spectrum Estimates to Detect Seasonal Effects and Flow Trading-Day Effects

Sensitive diagnostics are sometimes needed to determine if seasonal or trading-day effects are present in a series. This is especially true for detecting residual effects in a series that has already been adjusted for seasonal and trading-day effects. For a series adjusted by direct application of X-12-ARIMA, residual seasonality can result from inadequacies in the adjustment procedures chosen or from difficult-to-estimate seasonal effects in the series—for example, highly variable effects. With an indirectly adjusted aggregate series, whose adjustment is obtained from its component series (say a national series that is a sum of component regional series), it can happen that some of the component series are not adjusted for one or both of seasonal and trading effects, either because the effects are not detectable or because they are not reliably estimable in these components. This can leave residual effects.

As seasonal and calendar effects are approximately periodic, it is natural to use spectrum estimation to detect their presence. The period that defines seasonal effects is one year. Thus, in monthly series, seasonal effects can be discovered through the existence of prominent spectrum peaks at any of the frequencies $k/12$ cycles per month, $1 \leq k \leq 6$. In quarterly series the relevant frequencies are 1/4 and 1/2 cycles per quarter.

Monthly trading-day effects have a period of 28 years (336 months). This long period leads to an overabundance of frequencies potentially associated with trading-day effect peaks (see McNulty and Huffman 1989). Cleveland and Devlin (1982) demonstrated, however, for flow series that the most sensitive frequencies will typically be .348 cycles/month and .432 cycles/month. The empirical experience of W. P. Cleveland at the Federal Reserve Board showed that peaks at the biweekly-period alias-frequency .304 cycles/month ($.304 = 1 - 2 \times .348$) are also useful indicators of a trading-day effect.

Whenever seasonal adjustment is done (with or without trading-day adjustment), X-12-ARIMA automatically estimates two spectra, (1) the spectrum of the month-to-month differences of the adjusted series modified for extreme values from X-11 output table E2 (or of the first differences of logarithms of this series with a multiplicative adjustment) and (2) the spectrum of the final irregular component adjusted for extreme values, from output table E3. First-differencing is a crude detrending procedure that is usually adequate to enable the spectrum estimate to reveal significant seasonal and trading-day effects. The program compares the spectral amplitude at the seasonal and trading-day frequencies noted previously with the amplitudes at the next lower and higher frequencies plotted. If these neighboring amplitudes are smaller by a margin that depends on the range of all spectrum amplitudes, then plots of the estimated spectra are automatically printed, together with a warning message that gives the number of "visually significant" peaks found at seasonal or trading-day frequencies.

The best known spectrum estimator for detecting nonrandom periodic components is the periodogram [see Priestley (1981, pp. 390–415) for a very informative discussion]. For a series $x_t, 1 \leq t \leq N$, the periodogram, in decibel units, has the formula $10\log_{10}((2/N)|\sum_{t=1}^{N} x_t e^{i2\pi t\lambda}|^2), 0 \leq \lambda \leq .5$. [At the frequencies $\lambda = 2\pi n/N, 1 \leq n \leq [N/2]$, letting A_λ and B_λ denote the least squares estimates of the coefficients of the regression of x_t on $A\cos 2\pi\lambda t + B\sin 2\pi\lambda t$, the periodogram is equal to $10\log_{10}\{(N/2)(A_\lambda^2 + B_\lambda^2)\}$; see Priestley (1981, p. 395).] The periodogram is one of the two spectrum estimators in X-12-ARIMA, the other being the autoregressive spectrum estimator, which in decibel units has the form

$$10\log_{10}\left\{\frac{\sigma_m^2}{2\pi|1 - \sum_{j=1}^{m} c_j e^{i2\pi j\lambda}|^2}\right\}, \qquad 0 \leq \lambda \leq .5. \quad (7)$$

The coefficients c_j are those of the least squares regression of $x_t - \bar{x}$ on $x_{t-j} - \bar{x}, 1 \leq j \leq m$, with $\bar{x} = N^{-1}\sum_{t=1}^{N} x_t$, and σ_m^2 is the sample variance of the resulting regression residuals. For a discussion of this estimator, see Priestley (1981, pp. 600–612). The default spectrum estimator in X-12-ARIMA usually uses $m = 30$, as in the BAYSEA seasonal-adjustment program (Akaike 1980; Akaike and Ishiguro 1983). Although this estimate is somewhat less sensitive to the presence of periodic components than the periodogram, its graphs are much more stable under slight changes in the data window used or in the set of frequencies chosen for its evaluation. The radian frequencies used in the spectrum graphs produced by X-12-ARIMA are $\lambda = k/120, 0 \leq k \leq 60$, except that the three trading-day frequencies .304, .348, and .432, whose spectral amplitudes are plotted with a T, are used in place of their closest neighbors of the form $k/120$. The amplitudes at the seasonal frequencies $1/12, 2/12, \ldots, 6/12$ are plotted with an S. Examples will be given shortly.

The spectrum of any span of data within the series can be estimated. The default span for the automatically calculated spectrum estimates is the most recent eight years of data if the series is at least this long. Data-users are normally most concerned about recent data, and eight years of monthly data are usually enough to achieve reliable estimates of trading-day effects. When the pattern of the effects changes substantially over the course of the series, diagnostics calculated from the full series can lead to decisions that are inappropriate for the recent data, as we now demonstrate.

Figure 5 shows X-12-ARIMA's output spectrum plots of (7) from the irregular component of the construction series of single- and multi-unit housing starts (January 1964 to December 1982) from the Midwest region of the United States, after seasonal adjustment but without trading-day adjustment. The spectrum of Figure 5(a) is from the full irregulars series. It shows a strong peak at the main trading-day frequency .348 and a very slight peak at the frequency .304. The spectrum of Figure 5(b), which is calculated from the last eight years of the irregulars series, has no trading-day peaks. The conclusion is that there is not a significant trading-day effect late in this series. This conclusion was confirmed by model-selection diagnostics (Chen and Findley 1996a).

In a purely diagnostic mode, X-12-ARIMA can calculate a spectrum estimate of the first differences of an input time series (or of its logarithms) and print a spectrum plot without doing any further processing. This feature was designed for use at the Census Bureau in a once-a-year inspection looking for residual seasonal and trading-day effects in major aggregate series that are compiled from component series, some of which might not be seasonally adjusted.

2.2 Diagnostics for the Stability of the Seasonal Adjustments and Trends

A seasonal (and trading-day and holiday) adjustment that leaves detectable residual seasonal and calendar effects in the adjusted series is usually regarded as unsatisfactory. Even if no residual effects are detected, the adjustment will be unsatisfactory if the adjusted values (or important derivative statistics, such as the percent changes from one month to the next) undergo large revisions when they are recalculated as future time series values become available. Frequent, substantial revisions cause data users to lose confidence in the usefulness of adjusted data. Indeed, such instabilities in the adjustments should cause the producers of adjustments to question their meaning. Unstable adjustments can be the unavoidable result of the presence of highly variable seasonal or trend movements in the series being adjusted. They can, however, also be due to inappropriate option specification in the software used to produce the adjustments, in which case they are avoidable.

2.2.1 Sliding Spans. X-12-ARIMA includes two types of stability diagnostics, sliding spans and revision histories. The sliding-spans diagnostics display, and provide summary statistics for, the different outcomes obtained by running

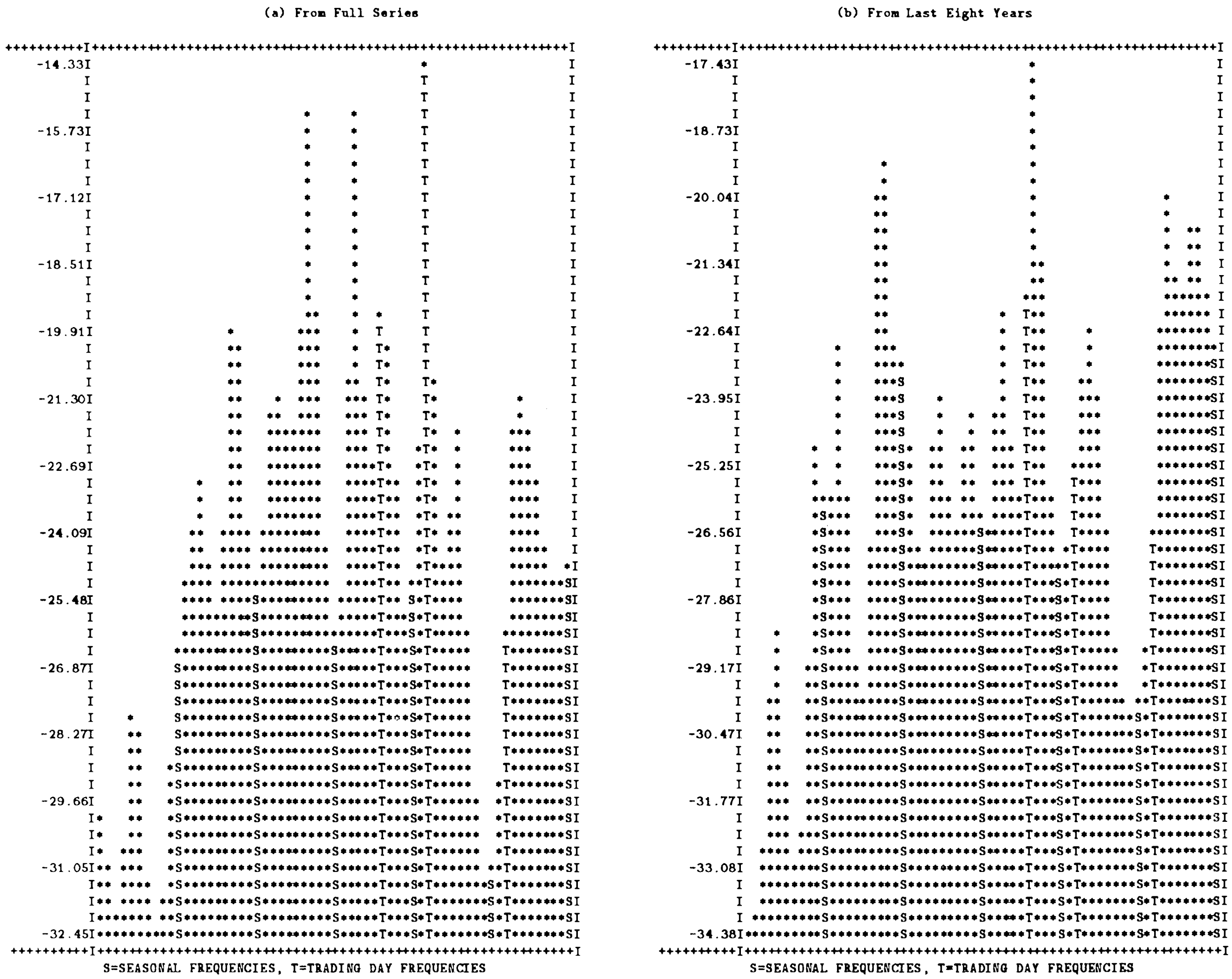

Figure 5. Graphs of the AR Spectrum (7) of the Irregulars of North Central Housing Starts From (a) the Full Series and (b) Its Last Eight Years. The dominant peak at the second trading-day frequency in (a) does not appear in (b). Hypothesis testing confirms the lack of a significant trading-day effect in the last eight years.

the program on up to four overlapping subspans of the series. For each month that is common to at least two of the subspans, these diagnostics analyze the difference between the largest and smallest adjustments of the month's datum obtained from the different spans. They also analyze the largest and smallest estimates of month-to-month changes and of other statistics of interest. Several uses of these diagnostics were demonstrated by Findley et al. (1990). It was shown how they improve on, or complement in important ways, earlier diagnostics for (a) determining if a series is being adjusted adequately, (b) deciding between direct and indirect adjustments of an aggregate series, and (c) confirming option choices such as the length chosen for the seasonal filter or showing that other option choices must be tried. We refer the reader to this article for examples of these uses and others. Battipaglia and Focarelli (1995) performed simulation experiments and concluded that stability statistics from sliding spans were significantly more correlated with adjustment accuracy than the Q statistic of X-11-ARIMA, which is a weighted average of the M1–M11 statistics. Other comparisons between sliding-spans diagnostics and Q were given by Findley and Monsell (1986) and Findley et al. (1990).

2.2.2 Revision Histories. The second type of stability diagnostic in X-12-ARIMA considers the revisions associated with continuous seasonal adjustment over a period of years. The basic revision calculated by the program is the difference between the earliest adjustment of a month's datum obtained when that month is the final month in the series and a later adjustment based on all future data available at the time of the diagnostic analysis. Similar revisions are obtained for month-to-month changes, trend estimates, and trend changes. Sets of these revisions, calculated over a consecutive set of time points within the series, are called revisions histories. We will show how they can suggest the number of years of forecasts to use in forecast extension of the series and how they indicate whether the (final) Henderson trend estimates (from output table D12) are stable enough to serve as an alternative to the seasonal adjustments. (The Australian Bureau of Statistics prefers to publish these trend estimates instead of seasonal adjustments because the trend estimates have fewer changes of direction and therefore seem more interpretable to data users, especially when the seasonal adjustments are quite volatile. For the same reason, some other statistical agencies are also considering publishing the Henderson trends.)

To describe the variety of revisions that can be obtained, we introduce precise notation. Suppose a set of options has been chosen for the application of X-12-ARIMA to the unadjusted time series $Y_t, 1 \leq t \leq N$. For any of these months t, and any integer u in the interval $t \leq u \leq N$, let $A_{t|u}$ denote the seasonally adjusted value for time t obtained with these options when only the data $Y_t, 1 \leq t \leq u$, are used in their calculation ($Y_{u+1}, \ldots, Y_N$ are withheld). For given t, as u increases these adjustments converge to a final adjusted value. When the $3 \times m$ seasonal filter is used, convergence is usually effectively reached in about $1 + m/2$ years. The largest revisions tend to occur when u is the same calendar month as t, specifically $u = t + 12, t + 24, \ldots$, and the next to largest changes a month later, $u = t+1, t+13, t+25, \ldots$. (In the additive decomposition case, the largest weights in the seasonal-adjustment filter combining all of the seasonal-adjustment calculations are at lags $1, 12, 13, 24, 25, \ldots$.) The adjustment $A_{t|t}$ obtained from data through time t is called the *concurrent* adjustment. It is usually the first adjustment obtained for month t. We call $A_{t|N}$ the *most recent* adjustment. In the case of a multiplicative decomposition, the revision from the concurrent to the most recent adjustment for month t is calculated by the program as a percentage of the concurrent adjustment,

$$R_{t|N}^{A} = 100 \times \frac{A_{t|N} - A_{t|t}}{A_{t|t}}.$$

For given N_0 and N_1 with $N_0 < N_1$, the sequence $R_{t|N}^{A}, N_0 \leq t \leq N_1$, is called a *revision history* of the seasonal adjustments from time N_0 to time N_1. We suggest that N_0 be at least as large as the effective length of the seasonal filter used, $12(2+m)$. It should definitely be large enough for reliable estimation of any trading-day or holiday adjustments being performed.

Period-to-period percent changes,

$$\Delta^{\%} A_{t|u} = 100 \times \frac{A_{t|u} - A_{t-1|u}}{A_{t-1|u}},$$

are often as important as the seasonal adjustments. X-12-ARIMA can produce revision histories for them:

$$R_{t|N}^{\Delta^{\%} A} = \Delta^{\%} A_{t|N} - \Delta^{\%} A_{t|t}, \qquad N_0 \leq t \leq N_1.$$

The program also calculates the analogous quantities for final Henderson trends $T_{t|u}$ and for their period-to-period percent changes $\Delta^{\%} T_{t|u}$. These histories are denoted by $R_{t|N}^{T}$ and $R_{t|N}^{\Delta^{\%} T}, N_0 \leq t \leq N_1$. [Note: A slightly larger N_0 is required for the trend-revision histories because the effective length of the trend filters is one or two years longer than that of the adjustment filters; see Bell and Monsell (1992).]

2.2.3 Two Applications of Revision Histories. We now present an example demonstrating how these histories can help with decisions about what kind of forecast extension to use, if any, and whether the Henderson trend is a practical alternative to the seasonal adjustments. To illustrate a variety of issues with a single example, we use a series for which the final Henderson trend estimates and seasonal adjustments have different relative stability properties, depending on which feature of the data is of interest. The series is construction starts of single- and multi-unit dwellings ("housing starts") in the Southern region of the United States beginning in January 1962 and ending in August of 1993. It is adjusted for trading-day effects as well as seasonal effects. For the latter, the X-11 default options are used. The regARIMA model used for forecast extension includes a regression variable to make an adjustment for an additive outlier (see Sec. 3) in December 1989. Figure 6 is a graph of the series from January 1981, along with the seasonally adjusted series and final Henderson trend ob-

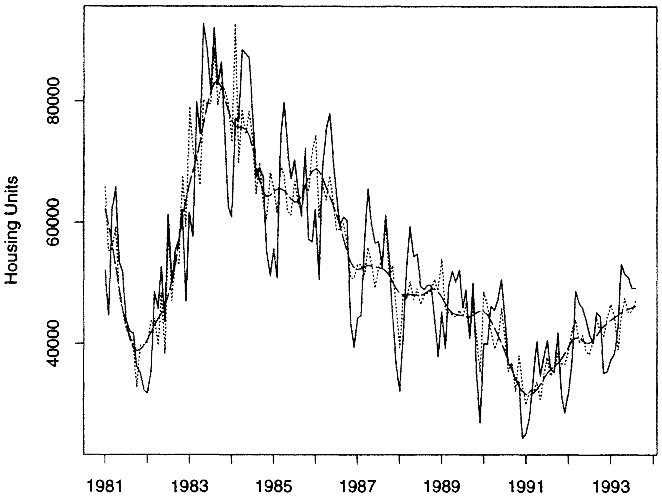

Figure 6. Southern Housing Starts With Seasonal Adjustment and Associated Trend: ——, Original; ---, SA; - - -, Trend.

tained using 42-month forecast extension. The trend is significantly smoother and more appealing to the eye than the seasonal adjustment. The revisions histories begin in January of 1981 and end in December of 1989. This ending date, three and a half years before the most recent datum, was chosen so that all revisions would be final or close-to-final revisions. Thus, they are revisions of similar type, so it is reasonable to consider their average magnitudes.

We start by examining the effect of the length of the forecast extension on the magnitudes of $|R^A_{t|N}|$ and $|R^T_{t|N}|$. The cases considered are no forecast extension, 12-month forecast extension, and 42-month forecast extension, the last length being the effective half-length of the 3×5 seasonal filter used. Bobbitt and Otto (1990) found that the use of such "full forecast extension" can result in smaller average revisions between concurrent and final seasonal adjustments than shorter forecast extension. Table 1 shows that the average magnitudes of the $R^A_{t|N}$ of the housing starts series, denoted avg$|R^A_{t|N}|$, follow this pattern. The table also includes counts of large revisions, which we have defined to be revisions of magnitude greater than 4% (this is more than twice the average magnitude of the seasonal-adjustment revisions).

For the data user interested only in the levels of the seasonally adjusted series, these results suggest that the adjustments obtained with the aid of a 42-month forecast extension are preferable to the other adjustments considered and to the Henderson trends.

We now consider month-to-month changes. The graphs of $\Delta^{\%}A_{t|t}$ and $\Delta^{\%}A_{t|N}$ are given in (a) of Figure 7 and the graphs of $\Delta^{\%}T_{t|t}$ and $\Delta^{\%}T_{t|N}$ in (b). (The quantities graphed were obtained using 42-month forecast extension.) The different scales in Figure 7, (a) and (b), make clear that the month-to-month changes in trend are often much smaller. The revisions, whose magnitudes are indicated by the lengths of the vertical lines connecting the concurrent and most recent estimates, are also smaller for the month-to-month changes in trend. For the different forecast leads

Table 1. Average Absolute Percent Revisions and Numbers of Extreme Revisions Over (1/1981–12/1989) for Seasonal Adjustments and Henderson Trends of Southern Housing Starts (1/1962–8/1993) Obtained Using Different Numbers of Forecasts

No. forecasts	avg$\|R^A_{t\|N}\|$	avg$\|R^T_{t\|N}\|$	No. $\|R^A_{t\|N}\|$ > 4.0%	No. $\|R^T_{t\|N}\|$ > 4.0%
0	2.1	3.8	17	33
12	1.8	3.1	10	33
42	1.5	3.0	2	26

0, 12, 42, the values of avg$|R^{\Delta\% A}_{t|N}|$ and avg$|R^{\Delta\% T}_{t|N}|$ are 2.5, 2.3, 2.2 and 1.7, 1.4, 1.4, respectively.

There is only one visible way in which the revisions of $\Delta^{\%}T_{t|t}$ are less appealing than the revisions of $\Delta^{\%}A_{t|t}$: About twice as often for $\Delta^{\%}T_{t|t}$, the most recent estimate has a different sign from that of the initial estimate. Such revisions, which change month-to-month increases to decreases or decreases to increases, are irritating for many data users. In Figure 7, the vertical connecting lines cross the horizontal axis at level 0 when there is a change in sign, making such changes easy to see. Both the number of sign changes and the sizes of the revisions are much smaller for the trends obtained after a 3-month wait, $T_{t|t+3}$: Compare Figure 8, which graphs the revisions from $\Delta^{\%}T_{t|t+3}$ to $\Delta^{\%}T_{t|N}$, with Figure 7(b). (The seasonally adjusted month-

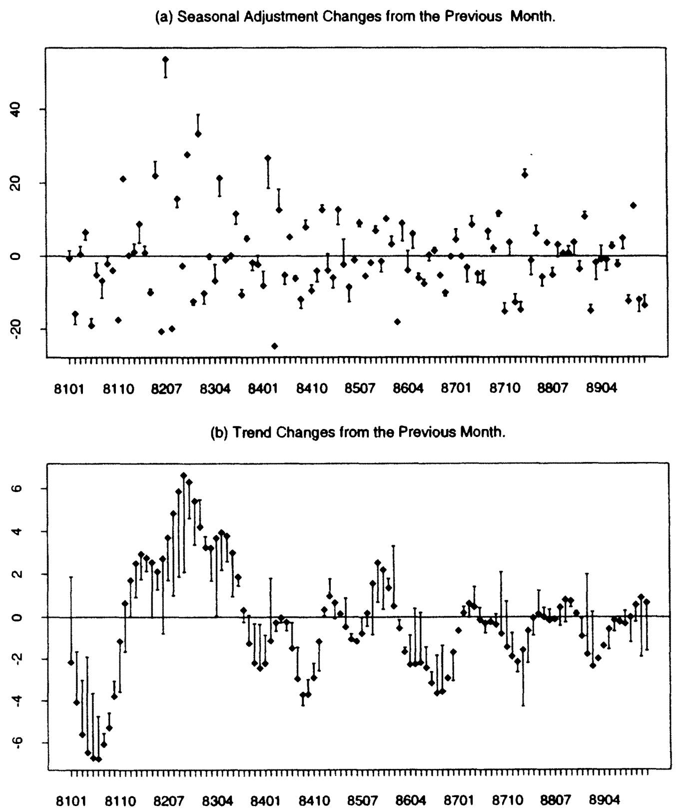

Figure 7. Concurrent (-) and Most Recent (♦) Estimates of Percent Changes From the Previous Month in the Seasonal Adjustment (a) and Trend (b) of Southern Housing Starts. The connecting vertical lines show the size of each revision. [Note that the scales of (a) and (b) are different.] When these lines cross the level zero axis, the revision of the concurrent value includes a change of sign, an unfavorable situation.

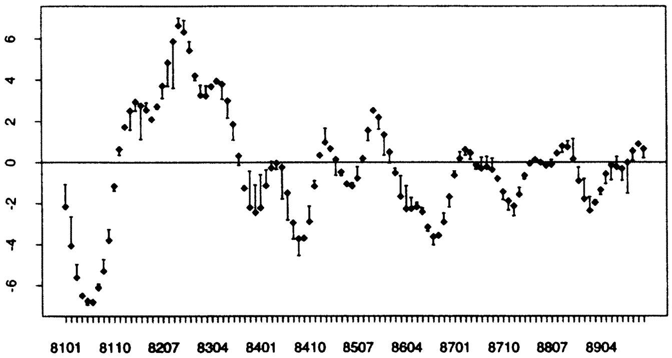

Figure 8. Three-Month-Lagged (-) and Most Recent (♦) Estimates of Percent Changes from the Previous Month in the Trend. Comparison with Figure 7(b) shows that the trend-change estimates calculated three months after the concurrent estimates have much smaller revisions than the latter and fewer changes of sign after revision.

to-month changes obtained with a three-month delay do not have improved revisions and are not graphed.)

The SABL seasonal-adjustment program of Cleveland, Devlin, Schapira, and Terpenning (1981) was the first to calculate revisions series. SABL produces a history of the differences between the seasonal adjustments obtained using seasonal factors projected a calendar year in advance and the concurrent seasonal adjustments. [Projected factor adjustments are much less used now than when SABL was created, having been displaced by concurrent adjustments because the latter generally have smaller revisions; see Dagum (1987) for a survey of the relevant literature. X-12-ARIMA can also calculate revisions of projected adjustment factors to most recent adjustment factors so that users can compare these with the revisions of concurrent factors.]

3. REGARIMA MODELING AND MODEL SELECTION

We now describe the time series modeling and model-selection methodologies of X-12-ARIMA, beginning in Section 3.1 with an overview of regARIMA models and the regressors for them that are included in X-12-ARIMA. Section 3.2 indicates how the program uses regARIMA models to identify automatically AO's and LS's. Section 3.3 deals with model selection. First, log-likelihood-based model-selection criteria are presented in Section 3.3.1, along with the way we use one such criterion, the Akaike information criterion (AIC), for automatically deciding whether or not a trading-day effect is present. Section 3.3.2 shows how the program's ability to "recreate history" is exploited for model selection, especially by withholding data, forecasting these data, and analyzing the resulting out-of-sample forecast errors.

3.1 Overview of regARIMA Modeling in X-12-ARIMA

Given a time series Y_t to be modeled, it is often necessary to take a nonlinear transformation of the series, $y_t = f_t(Y_t)$, to obtain a series that can be adequately fit by a regARIMA model. For example, if Y_t is a positive-valued series with seasonal movements proportional to the level of the series, one would usually take logarithms or, more generally,

$$y_t = \log\left(\frac{Y_t}{d_t}\right) = \log Y_t - \log d_t, \tag{8}$$

where d_t is some appropriate sequence of divisors. Possible divisors include (a) deseasonalized and detrended length-of-month factors N_t/N_t^* from (6), (b) combined trading-day and Easter-holiday effect factors obtained from a regression model of the irregular component of Y_t (obtained from a preliminary run), and (c) user-defined adjustment factors that estimate the effects of unusual economic events. X-12-ARIMA can calculate the transformed series (8) for choices (a) and (b) via user-specified options and for choice (c) by reading in the divisors from a user-specified data file.

The built-in transformations include a one-parameter family of power transformations (modified "Box–Cox" transformations),

$$y_t^{(\lambda)} = \begin{cases} Y_t/d_t, & \lambda = 1 \\ \lambda^2 + [(Y_t/d_t)^\lambda - 1]/\lambda, & \lambda \neq 0, 1 \\ \log(Y_t/d_t), & \lambda = 0, \end{cases} \tag{9}$$

which changes smoothly in λ and preserves positivity if $Y_t/d_t > 1.0$. Although the program permits any value of λ to be used for the purpose of obtaining forecast and backcast extensions, to get regression preadjustments for a seasonal-adjustment decomposition in X-12-ARIMA, λ must be 0 or 1. These are the only values of λ for which it is possible to isolate the effect on Y_t of regression components of y_t.

Let B denote the backshift operator, $By_t = y_{t-1}$. X-12-ARIMA can estimate regARIMA models of order $(p, d, q)(P, D, Q)_s$ for y_t. These are models of the form

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D\left(y_t - \sum_{i=1}^{r}\beta_i x_{it}\right) = \theta_q(B)\Theta_Q(B^s)a_t, \tag{10}$$

where s is the length of the seasonal period, $s = 4$ or 12. The polynomials $\phi_p(z), \Phi_P(z), \theta_q(z), \Theta_Q(z)$ with degrees p, P, q, and Q, respectively, have constant terms equal to 1. For example, if $p \geq 1$, we have $\phi_p(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$. These polynomials are constrained so that the zeros of $\theta_q(z)$ and $\Theta_Q(z)$ have magnitudes greater than or equal to 1, and (in the default estimation procedure) so that the zeros of $\phi_p(z)$ and $\Phi_P(z)$ have magnitudes greater than 1. Because a_t is assumed to be a sequence of independent variables with mean 0 and constant variance σ_a^2, it follows from these constraints that $w_t = (1-B)^d(1-B^s)^D(y_t - \sum_{i=1}^{r}\beta_i x_{it})$ is a covariance stationary time series that satisfies the difference equation $\phi_p(B)\Phi_P(B^s)w_t = \theta_q(B)\Theta_Q(B^s)a_t$. Consequently, we can reexpress the model (10) for y_t as

$$(1-B)^d(1-B^s)^D y_t = \sum_{i=1}^{r}\beta_i\{(1-B)^d(1-B^s)^D x_{it}\} + w_t. \tag{11}$$

This is a regression model with stationary autoregressive moving average (ARMA) errors w_t for suitably differenced y_t. Its regressors result from applying the same differencing operations to the x_{it}. The model (11), together with an assumption that the innovations a_t in the model for w_t are iid $\mathrm{N}(0, \sigma^2)$, determine the likelihood function that is maximized to estimate the regression coefficients β_i, σ^2, and the coefficients of $\phi_p(B), \Phi_P(B^s), \theta_q(B)$, and $\Theta_Q(B^s)$. The default likelihood in X-12-ARIMA is the fully exact Gaussian likelihood. To help circumvent convergence problems in the numerical maximization (which occur rarely), the approximating conditional Gaussian likelihood defined by Box and Jenkins (1976) can optionally be used instead of the exact likelihood. There is also a third option in which the likelihood is conditional for the autoregressive parameters and exact for the moving average parameters (see Hillmer and Tiao 1979). (These two alternative likelihoods do not constrain the zeros of autoregressive polynomials.)

In model estimation, any of the ARMA coefficients can be held at fixed values, such as 0. The program produces asymptotic standards errors, correlations, and t statistics for the estimated coefficients, as well as confidence intervals for forecasts. With the exception of the confidence intervals, these statistics remain valid with non-Gaussian data if the model form is correctly specified; see, for example, Hosoya and Taniguchi (1982).

The set of built-in regressors for monthly series is listed in Table 2. Applications using some of them are given in Section 4. As discussed in Section 3.2, the program has options to add automatically both AO and LS regressors to the set of regression variables in the model. In this way, the regARIMA coefficient estimates and forecasts can be made robust to some kinds of atypical data values and to sudden changes in the level of the series. The user can optionally choose to have such automatically identified outliers and level shifts removed from the data, together with specified other regression effects, before the X-11 procedure outlined in Appendix A is applied. Through such preadjustments, the seasonal factors that are used to adjust the original data can be shielded from distortion.

The extreme value treatments within the X-11-ARIMA procedure, which were described fully by Dagum (1980)

Table 2. Predefined Regression Variables in X-12-ARIMA

Regression effect	Variable definition(s)
Trend constant	$(1-B)^{-d}(1-B^s)^{-D} I(t \geq 1)$, where $I(t \geq 1) = \begin{cases} 1 & \text{for } t \geq 1 \\ 0 & \text{for } t < 1 \end{cases}$
[a]Fixed seasonal	$M_{1,t} = \begin{cases} 1 & \text{in January} \\ -1 & \text{in December} \\ 0 & \text{otherwise} \end{cases}, \ldots, M_{11,t} = \begin{cases} 1 & \text{in November} \\ -1 & \text{in December} \\ 0 & \text{otherwise} \end{cases}$
[a]Fixed seasonal	$\sin(\omega_j t)$, $\cos(\omega_j t)$, where $\omega_j = 2\pi j/12, 1 \leq j \leq 6$ (drop $\sin(\omega_6 t) \equiv 0$)
Trading day (monthly or quarterly flow)	T_{1t} = (no. of Mondays) − (no. of Sundays), ..., T_{6t} = (no. of Saturdays) − (no. of Sundays)
[a]Length-of-month (monthly flow)	$N_t - \bar{N}$, where N_t = length of month t [in days] and $\bar{N} = 30.4375$ [average length of month]
Leap year (monthly flow)	$N_t - N_t^*$, where $N_t^* = (N_t + N_{t-12} + N_{t-24} + N_{t-36})/4$ (Note: This variable is 0 except in February.)
Stock trading day (monthly stock)	$T_{1,t} = \begin{cases} 1 & \tilde{w}^{\text{th}} \text{ day of month } t \text{ is a Monday} \\ -1 & \tilde{w}^{\text{th}} \text{ day of month } t \text{ is a Sunday} \\ 0 & \text{otherwise} \end{cases}, \ldots, T_{6,t} = \begin{cases} 1 & \tilde{w}^{\text{th}} \text{ day of month } t \text{ is a Saturday} \\ -1 & \tilde{w}^{\text{th}} \text{ day of month } t \text{ is a Sunday} \\ 0 & \text{otherwise,} \end{cases}$ where $\tilde{w}$ is the smaller of w and the length of month t. For end-of-month stock series, set w to 31.
[b]Easter holiday (monthly or quarterly flow)	$E(w, t) = 1/w \times$ (no. of the w days before Easter falling in month (or quarter) t) [Note: This variable is 0 except in February, March, and April (or first and second quarter). It is nonzero in February only for $w > 22$.]
[b]Labor Day (monthly flow)	$L(w, t) = 1/w \times$ (no. of the w days before Labor Day falling in month t) [Note: This variable is 0 except in August and September.]
[b]Thanksgiving (monthly flow)	$TC(w, t)$ = proportion of days from w days after Thanksgiving through December 24 that fall in month t (negative values of w indicate days before Thanksgiving) [Note: This variable is 0 except in November and December.]
Additive outlier at t_0	$\mathrm{AO}_t^{(t_0)} = \begin{cases} 1 & \text{for } t = t_0 \\ 0 & \text{for } t \neq t_0 \end{cases}$
Level shift at t_0	$\mathrm{LS}_t^{(t_0)} = \begin{cases} -1 & \text{for } t < t_0 \\ 0 & \text{for } t \geq t_0 \end{cases}$
Temporary ramp, t_0 to t_1	$\mathrm{RP}_t^{(t_0,t_1)} = \begin{cases} -1 & \text{for } t \leq t_0 \\ (t-t_0)/(t_1-t_0) - 1 & \text{for } t_0 < t < t_1 \\ 0 & \text{for } t \geq t_1 \end{cases}$

[a] The variables shown are for monthly series. Corresponding variables are available for quarterly series.
[b] The actual variable used for monthly Easter effects is $E(w, t) - \bar{E}(w, t)$, where the $\bar{E}(w, t)$ are the "long-run" (computed over 38,000 years) monthly means of $E(w, t)$ (nonzero only for February, March, and April). Analogous deseasonalized variables are used for Labor Day and Thanksgiving effects and for quarterly Easter effects.

and nicely flowcharted by Hylleberg (1986, p. 91), provide some protection against AO's for the seasonal factors. The trend filters applied in the course of obtaining the seasonal factors cannot follow sudden, large LS's, however. Thus, estimation of LS's together with preadjustment for them, as illustrated in Figure 2, is an especially important capability of X-12-ARIMA. Other approaches to treating sudden changes in level have been considered. Bruce and Jurke (1996) compared X-12-ARIMA seasonal adjustments of series having preadjustments for LS's and AO's from regARIMA models with seasonal adjustments obtained from a state-space model that uses Gaussian-mixture state- and observation-noise models to deal with disruptions to the level of the series. They concluded that the regARIMA approach succeeds more broadly.

3.2 Automatic Outlier Treatment

The automatic methods for identifying AO's and LS outliers are stepwise regression procedures based on work of Chang and Tiao (1983) (see also Bell 1983; Burman 1983; Chang, Tiao, and Chen 1988). In the default procedure, whose steps are listed in Appendix C, appropriate AO and LS regressors are fit at (almost) all time points of the series (or of a chosen subspan), and their corresponding t statistics are compared against specified critical values. The default critical value is 3.8 for both regressor types. Such large critical values are appropriate because of the large number of regressors to which individual significance tests are applied. Automatic outlier identification is, in this respect, different from the model-selection problem discussed in the next subsection, where less stringent criteria are often used to include other regressors.

3.2.1 Instabilities of Outlier Identification. The set of automatically identified outliers can change if the regressor set or ARIMA model type is changed. For example, in series with a strong date-of-Easter effect, Marches and Aprils are often identified as outliers if no regressors for this effect are included in the model but not if such regressors are used. A second source of instability in the composition of the set of observations defined as AO's or LS outliers is the use of stepwise regression procedures based on specified critical values. Regressors with t-statistic values just below the critical values can have their t statistics increase above the critical values as new data are added to the series over time. Conversely, regressors can drop out of the set of identified outliers as new data are added. The printed output of X-12-ARIMA's automatic-outlier-identification option lists months whose AO or LS regressors are close to the critical values. This is done to enable the user to consider in advance whether to include such regressors in subsequent runs of the program. Instability is a problem with most outlier detection and automatic model-selection schemes. In the context of regressor selection for independent observations, Breiman (1997) proposed some interesting, although computationally expensive, data-perturbation approaches to achieve more stable selections.

3.3 Model Identification and Selection

X-12-ARIMA has an automatic ARIMA modeling option that is patterned after the procedure of X-11-ARIMA/88. Under this option, the program examines the fit of regARIMA models whose ARIMA structures are those with a specified set of orders $(p,d,q)(P,D,Q)_s$. The default set consists of the five models with nonseasonal orders (0 1 1), (0 1 2), (2 1 0), (0 2 2), and (2 1 2), and always the same seasonal order, $(0\ 1\ 1)_s$, exactly as in X-11-ARIMA/88. In X-12-ARIMA, the user can specify an alternative set of models for consideration. Moreover, the user can specify regression variables to be included in the model and can use built-in criteria to decide if trading-day, AO, and LS regressors should be included with any specified regressors. A fitted model whose estimated mean absolute percent forecast error statistic and Box–Ljung portmanteau statistic are below certain thresholds is considered an acceptable model.

For the situation in which none of the automatically tested models is adequate, or where the user wishes to identify or check a model, X-12-ARIMA has options to produce standard modeling diagnostics. For model identification, the program provides the sample autocorrelations and partial autocorrelations of the residuals obtained by doing OLS regression in (11),

$$(1-B)^d(1-B^s)^D y_t - \sum_{i=1}^{r} \hat{\beta}_i^{\mathrm{OLS}}\{(1-B)^d(1-B^s)^D x_{it}\}.$$

For model checking, it produces the sample autocorrelations and partial autocorrelations of the residuals from a fitted regARIMA model [estimates of the a_t in (11)], together with associated portmanteau statistics and histograms of residuals (see Box and Jenkins 1976; Abraham and Ledolter 1983; Vandaele 1983; Bell 1996).

3.3.1 Log-Likelihoods, AIC, and Automatic Trading-Day-Effect Modeling. Suppose that there are competing regARIMA models whose diagnostics seem adequate and that these models differ in the choice of the ARMA model for w_t in (11), or in the choice of regressors other than AO and LS regressors, or in the choice of transformations $f_t(Y_t)$ of Y_t. When the parameters in these models have been estimated by maximizing the *exact* Gaussian likelihoods, then X-11-ARIMA provides several log-likelihood-based model-selection criteria that can be used to select one of the models. Let the logarithm of the maximized exact likelihood of a covariance stationary time series model for $(1-B)^d(1-B^s)^D f_t(Y_t), d+sD+1 \le t \le N$, be denoted by $\hat{L}^f_{d+sD+1,N}$. This can be converted into the logarithm of a likelihood for $Y_{d+sD+1},\ldots,Y_N$ conditional on $Y_1,\ldots,Y_{d+sD}$, denoted by $\hat{L}(Y_N,\ldots,Y_{d+sD+1}|Y_{d+sD},\ldots,Y_1)$, by adding the log of the Jacobian determinant of the transformation $y_t = f_t(Y_t), d+sD+1 \le t \le N$,

$$\hat{L}(Y_{d+sD+1},\ldots,Y_N|Y_1,\ldots,Y_{d+sD}) = \hat{L}^f_{d+sD+1,N} + \sum_{t=d+sD+1}^{N} \log\left|\frac{df_t(Y_t)}{dY_t}\right|. \quad (12)$$

When $d>0$ or $D>0$, we are, in effect, treating the starting values $y_1,\ldots,y_{d+sD}$ as fixed.

Let m denote the number of free parameters estimated in the model. If there are no coefficient constraints in (11), then $m = r + p + q + P + Q + 1$, counting the coefficients and the variance of a_t. The AIC statistic for this model is defined by

$$\mathrm{AIC}_{N|d+sD} = -2\hat{L}(Y_{d+sD+1}, \ldots, Y_N | Y_1, \ldots, Y_{d+sD}) + 2m. \quad (13)$$

Given several competing models with the same $d + sD$, Akaike's *minimum AIC criterion* states that the model with smallest AIC is the best of the models for Y_t. [See Akaike (1973), Findley (1985), and Brockwell and Davis (1987) for technical details and Findley and Parzen (1995) for historical background.] X-12-ARIMA also calculates the small-sample version of AIC derived by Hurvich and Tsai (1989), the Schwarz Bayesian information criterion statistic (Schwarz 1978), and the Hannan and Quinn statistic (Hannan and Quinn 1979). These differ from AIC in the replacement of the term $2m$ in (13) by $2m/\{1-(m+1)/(N-d-sD)\}, m \log(N-d-sD)$, and $2m \log\log(N-d-sD)$, respectively. These replacements are usually larger than $2m$. Therefore these other criteria are less tolerant than AIC of models with more coefficients. (For these criteria too, the smallest value of the statistic over a set of competing models is used to determine a preferred model.)

Comparing different transformations. Rather frequently, it is necessary to compare regARIMA models whose data transformations $f_t(Y_t)$ differ. The most common situation is the one in which two different divisors d_t are used in the log transform (8), say $d_t = N_t/N_t^*$ and $d_t = 1$ as in the next subsection, or "subjective" and "objective" choices of d_t as in Section 4.4. The next most frequent situation is the one in which different choices of λ are considered in (9), usually $\lambda = 0$ and $\lambda = 1$. The choice $\lambda = 0$ suggests that the seasonal-adjustment decomposition should be multiplicative, $\lambda = 1$ that it should be additive.

To help decide between two transformations, either the out-of-sample forecast diagnostics described in Subsection 3.2.2 or one of the log-likelihood-based criteria, such as AIC, can be used. In the case of choosing a power transformation (9) with unrestricted λ when the same ARIMA model type is used with the different λ, the numbers of estimated parameters do not change, so the latter criteria all prefer whichever λ yields a larger log-likelihood. In this situation, one can examine an interval of values and to identify the λ maximizing the log-likelihoods (12) (see Ansley, Spivey, and Wrobleski 1977). Although this procedure appears to yield reasonable results (Shulman and McKenzie 1988), the asymptotic distribution theory on which it rests has not been verified and, if valid, may require subtle arguments for its proof when $d > 0$ or $D > 0$.

Deciding whether to adjust for trading-day effects. We now describe X-12-ARIMA's option for automatically determining if trading-day regressors should be included in the model (11) after the rest of the model has been specified (meaning f_t, d, D, and ARMA model type for w_t, and any other regressors). The models with and without trading-day regressors are estimated. In the default case for multiplicatively decomposed flow series, the model with trading-day regressors also uses the Leap Year effect preadjustments $d_t = N_t/N_t^*$ [see (14)], but the model without these regressors does not. The AIC's for the two models are compared, and the model with the smaller AIC is chosen (for forecast extension and for estimating any requested regression preadjustments).

As we indicated in Section 1.4.1 after (6), the default regARIMA trading-day model for a multiplicative decomposition of a seasonal monthly series Y_t has the form

$$(1-B)^d(1-B^{12})^D \left\{ \log\left(\frac{N_t^*}{N_t} Y_t\right) - \sum_{i=1}^{6} \beta_i (D_{it} - D_{7t}) - \sum_{i=8}^{r} \beta_i x_{it} \right\} = w_t. \quad (14)$$

If there are only trading-day regressors in the model, the second sum is omitted. With $\beta_7 = -\sum_{i=1}^{6} \beta_i$, the trading-day factors obtained from (14) have the form

$$\frac{N_t}{N_t^*} \exp \sum_{i=1}^{6} \beta_i (D_{it} - D_{7t}) = \frac{N_t}{N_t^*} \prod_{i=1}^{7} e^{\beta_i (D_{it} - 4)}. \quad (15)$$

The alternative model with no trading-day effects is

$$(1-B)^d(1-B^{12})^D \left\{ \log Y_t - \sum_{i=8}^{r} \beta_i x_{it} \right\} = w_t. \quad (16)$$

Thus, $f(Y_t) = \log Y_t$ is used instead of $f_t(Y_t) = \log(N_t^* Y_t / N_t)$ in (14).

Our experience is that comparing the AIC's of (14) and (16) to decide if a trading-day effect is present is substantially more reliable than X-11's F test of the hypothesis $\beta_1 = \cdots = \beta_6 = 0$ in (3). As we mentioned in the first paragraph of Section 1.4.3, in the empirical study summarized by Chen and Findley (1996a), this F test falsely indicated significant trading-day effects in 14 of 71 series. The automatic procedure just described classified these 14 series as not having estimable trading-day effects, in agreement with the forecast comparison procedure described next.

3.3.2 Historical Output for Comparing Models: Out-of-Sample Forecasting Performance, AIC Histories. We return to the option discussed in Section 2.2 under which the program recreates history. Recall that it performs a sequence of runs on increasing spans of data within the series. Starting from an initial segment of the series, the spans grow with each new run by the addition of one observation until the full series is included.

Out-of-sample forecasts. To obtain information about a model's forecasting performance, the available time series data outside each span can be regarded as future data to be forecasted from a model fit to the span. These forecasts can be compared to the actual series values or, for series values identified as outliers, to the outlier-adjusted values. As an option, the X-12-ARIMA program calculates the resulting out-of-sample forecast errors and stores them for later analysis, along with their accumulating sums of squares. When forecast errors are available from two competing models, the sequence of differences between the accumulating sums of squared errors can be an effective model-selection diagnostic, as we shall demonstrate in Section 4.3 (see also Chen and Findley 1996a,b). We now give a detailed description of this diagnostic.

Assume that we are interested in h-step(-ahead) forecasting of the time series $Y_t, 1 \le t \le N$. Suppose that a regARIMA model has been proposed for the transformed series $y_t = f(Y_t)$. Let N_0 be a number less than $N-h$ that is large enough that the data $y_t, 1 \le t \le N_0$, can be expected to yield reasonable estimates of the model's coefficients. For each t in $N_0 \le t \le N-h$, let $y_{t+h|t}$ denote the forecast of y_{t+h} obtained by estimating the regARIMA model using only the data $y_s, 1 \le s \le t$, and by using this estimated model to forecast h steps from time t. Then the *out-of-sample* h-step forecast of Y_{t+h} is defined to be $Y_{t+h|t} = f^{-1}(y_{t+h|t})$. We define the associated forecast error by $e_{t+h|t} = Y_{t+h} - Y_{t+h|t}$ if all AO, LS, and ramp regressors in the regARIMA model for the full series $y_1, \ldots, y_N$ have the value 0 at time $t+h$. Otherwise, we define $e_{t+h|t} = f^{-1}(\bar{y}_{t+h}) - Y_{t+h|t}$, where $\bar{y}_{t+h}$ is obtained by subtracting from y_{t+h} all such regression effects. The main diagnostic calculated by the program is the sequence of accumulating sums of squared out-of-sample forecast errors,

$$\mathrm{SS}_{h,M} = \sum_{t=N_0}^{M} e^2_{t+h|t}, M = N_0, \ldots, N-h. \tag{17}$$

Suppose there are two competing models, Model 1 and Model 2, with forecast errors $e^{(1)}_{t+h|t}$ and $e^{(2)}_{t+h|t}$ and with sums $\mathrm{SS}^{(1)}_{h,M}$ and $\mathrm{SS}^{(2)}_{h,M}$, respectively. Then we plot a standardized version of the differences $\mathrm{SS}^{(1)}_{h,M} - \mathrm{SS}^{(2)}_{h,M}$ defined by

$$\mathrm{SS}^{1,2}_{h,M} = \frac{\mathrm{SS}^{(1)}_{h,M} - \mathrm{SS}^{(2)}_{h,M}}{\mathrm{SS}^{(2)}_{h,N-h}/(N-h-N_0)} \tag{18}$$

against M, for $M = N_0, \ldots, N-h$. The recursion formula

$$\mathrm{SS}^{1,2}_{h,M+1} = \mathrm{SS}^{1,2}_{h,M} + \frac{(e^{(1)}_{M+1+h|M+1})^2 - (e^{(2)}_{M+1+h|M+1})^2}{\mathrm{SS}^{(2)}_{h,N-h}/(N-h-N_0)},$$

$$M = N_0, \ldots, N-h-1,$$

shows that over intervals of values of M in which the graph of (18) goes up, the forecast performance of Model 2 is better; if it goes down, Model 1 is better; and if it has no general direction, neither model's forecast performance dominates. The denominator in (18) provides a scale for the interpretation of jumps in the graph.

This diagnostic has the important virtue of not requiring the assumption that any of the models being compared is correct. Its use is not limited to situations in which forecasting is the main goal of modeling, as the examples of Section 4.3 will show.

AIC histories. Suppose the minimum AIC criterion is being used to decide between two models for $Y_1, \ldots, Y_N$ with AIC statistics $\mathrm{AIC}^{(1)}_{N|d+sD}$ and $\mathrm{AIC}^{(2)}_{N|d+sD}$. Then the preference is determined by the sign of the difference

$$\begin{aligned}\mathrm{AIC}^{1,2}_{N|d+sD} &= \mathrm{AIC}^{(1)}_{N|d+sD} - \mathrm{AIC}^{(2)}_{N|d+sD} \\ &= 2\{\hat{L}^{(2)}_{N|d+sD} - \hat{L}^{(1)}_{N|d+sD}\} + 2\{m^{(1)} - m^{(2)}\},\end{aligned} \tag{19}$$

where $\hat{L}^{(i)}_{N|d+sD}$ denotes the maximized log-likelihood (12) and $m^{(i)}$ the number of estimated parameters of the ith model, $i = 1, 2$. Often one wishes to know something about the stability of such a model choice. In the classical situation in which Model 1 is a constrained version of Model 2, under the assumption that Model 1 has the correct form, the large-sample distribution of $2\{\hat{L}^{(2)}_{N|d+sD} - \hat{L}^{(1)}_{N|d+sD}\}$ is chi-squared with $m^{(2)} - m^{(1)}$ df from which a probability value can be calculated for $\mathrm{AIC}^{1,2}_{N|d+sD}$. The assumptions required by this approach are too restrictive, however, not least because so many naturally occurring time series model comparisons are like the comparison of (14) and (16) in Subsection 3.3.1: Neither model is a constrained version of the other. X-12-ARIMA's AIC history option offers a somewhat more versatile diagnostic of the stability of minimum AIC model selections. For each model, the sequence of AIC values reestimated from subspans of data $Y_1, \ldots, Y_M, N_0 \le M \le N$, can be obtained. From these, the AIC difference sequence

$$\mathrm{AIC}^{1,2}_{M|d+sD} = \mathrm{AIC}^{(1)}_{M|d+sD} - \mathrm{AIC}^{(2)}_{M|d+sD},$$

$$N_0 \le M \le N, \tag{20}$$

can be calculated and examined for constancy of sign. An application of this diagnostic will be given in Section 4.4.

4. USING MODELS TO SOLVE ADJUSTMENT PROBLEMS: FOUR EXAMPLES

We present four applications of regARIMA models and the model-selection diagnostics discussed in Section 3.3 to problems encountered in seasonal adjustment.

4.1 Using Regressors to Verify a Change in Seasonal Pattern

For the regressors of Table 2 that model fixed seasonal effects and trading-day effects and for their quarterly analogs, X-12-ARIMA has a built-in procedure for modeling a *change of regime* at a user-specified point in time. We illustrate the procedure for the fixed quarterly seasonal variables M_{it} (which are defined like the fixed monthly seasonal variables of Table 2) and a changepoint designated N_c. For $i = 1, 2, 3$, define

$$M^c_{it} = \begin{cases} M_{it}, & 1 \le t \le N_c \\ 0, & N_c < t \le N, \end{cases}$$

where N denotes the length of the modeled series. The program models a change of regime by including both the M_{it} and the M^c_{it} in the regressor set.

We consider again the net income series of Figure 2. To verify a change in seasonal pattern at the time point N_c

corresponding to the first quarter of 1982, X-12-ARIMA was applied to estimate a model of the form

$$(1-B)(1-B^{12})\left\{\log Y_t - \beta_0 \mathrm{LS}_t^{(1982.1)} - \sum_{i=1}^{3}\beta_i M_{it} - \sum_{i=1}^{3}\beta_i^c M_{it}^c - \sum_{i=4}^{6}\beta_i x_{it}\right\} = (1-\theta B)(1-\Theta B^{12})a_t,$$

in which the regressors x_{it} model an additional LS identified in the second quarter of 1980 and AO's identified in the third quarter of 1974 and the first quarter of 1981. The AIC of this model, calculated as in (13), has the value $\mathrm{AIC}^c = 993.1$. The model of the same form but without the M_{it}^c can be used to represent the hypothesis of no change of seasonal pattern. It has three fewer estimated parameters but a much larger AIC value, AIC $= 1{,}028.8$. Hence, by the minimum AIC criterion, the model with a change in seasonal pattern is preferred.

A standard hypothesis test leads to the same conclusion. Because the second model is a constrained version of the first (with constraints $\beta_i^c = 0$), two times the difference of log-likelihoods, which is equal to $\mathrm{AIC} - \mathrm{AIC}^c + 2 \times 3 = 41.7$ [see (19)] can be compared to values of a chi-squared distribution with 3 df under the hypothesis of no change of regime. The value 41.7 strongly contradicts this hypothesis, being extraordinarily large for this chi-squared distribution.

4.2 Using AO Regressors to Replace Missing Data

The regARIMA modeling capability of X-12-ARIMA makes possible a rather simple approach to circumventing or estimating missing observations. The procedure requires the user to supply values for the missing observations (any values will do) and to then fit a regARIMA model to the completed dataset with AO regressors at the times of the missing observations. (The value −99999. in the input series always denotes a missing value and causes the program to insert the appropriate AO regressor automatically.) If Y_{t_0} is the value specified for the missing observation at time t_0, and if β_{t_0} is the estimated coefficient of the regressor $\mathrm{AO}_t^{(t_0)}$ in the fitted model, then the regression-adjusted value,

$$\hat{Y}_{t_0} = Y_{t_0} - \beta_{t_0}\mathrm{AO}_t^{(t_0)}, \tag{21}$$

provides an estimate of the missing datum that the program can use for calculating forecasts and seasonal decompositions. If the user requests autocorrelations and partial autocorrelations of the differenced data to help identify a model, then the OLS estimate $\beta_{t_0}^{\mathrm{OLS}}$ is used in (21) to provide an estimate needed to calculate these statistics.

There is an alternative procedure that estimates a missing datum via a regARIMA model's Gaussian conditional expectation of the missing datum given the available data. This procedure is optimal if the estimated model is the true model and the data are Gaussian. It is implemented in the regARIMA model-based signal-extraction seasonal-adjustment programs TRAMO and SEATS of Gomez and Maravall (1994a,b). [Their procedure is equivalent to the modified Kalman filter of Kohn and Ansley (1986), which extends the approach proposed by Jones (1980) to the case of models with differencing and missing data in the first $d + sD$ time points.] In the case of independent observations, it gives the same replacement values as (21) (Cook and Weisberg 1982, p. 33). With regARIMA time series models, theoretical calculations show that its values can be expected to be well approximated by those of (21) (Bruce and Martin 1989; Ljung 1993). We compared TRAMO's "optimal" estimates with X-12-ARIMA's estimates from (21) for several series from which observations were deleted at random after a regARIMA model had been identified for the full series. Observations that had been identified as outliers were not candidates for deletion. The estimates of the missing values from both procedures were always very close to each other. They were also usually quite close to the value of the deleted datum (< 2% error). The worst error observed, about 6%, occurred with the series of values of manufacturers' shipments of electrical appliances. The estimation results for the three observations deleted from this series are given in Table 3.

TRAMO also implements the procedure of (21). We followed TRAMO's use of −99999. as the missing value designator.

4.3 Comparing Trading-Day Estimation Procedures

We now wish to illustrate the versatility of the model-comparison diagnostic (18). It is not obvious how to compare estimates of effects from regression models of the irregular component (Sec. 1.4) with analogous estimates from regARIMA models of the observed series—for example, the trading-day factors (2) and (15). Model-selection procedures like those based on AIC comparisons are inapplicable because the models are fit to different time series. We shall show that forecast comparisons are possible because forecasts of calendar effects estimated from a regression model of the irregulars can be used to obtain forecasts of the observed series. This enables us to call on the model-comparison principle that a model that produces better forecasts can reasonably be assumed to produce calendar-effect estimates that better describe what is present in the series.

To begin, we need to explain how X-12-ARIMA calculates out-of-sample h-step-ahead forecasts when an effect is estimated from the irregular component. It will be sufficient to discuss the case of trading-day factors (2) estimated from the model (3). Given estimates of the coefficients $\beta_1, \ldots, \beta_6$, the factors (2) can be calculated for all times t. We use $\mathrm{TD}_t^{(M)}$ to denote the factors, when these coefficients have been estimated from the reduced dataset $Y_1, \ldots, Y_M$

Table 3. "Optimal" and AO Regressor Estimates of Deleted Observations From Shipments of Electrical Appliances

Date	Value	"Optimal" est.	Error (%)	Est. from (21)	Error (%)
5/1977	661.	660.61	.06	660.48	.08
9/1979	1,088.	1,156.25	6.27	1,159.07	6.53
11/1981	1,397.	1,396.55	.04	1,396.07	.07

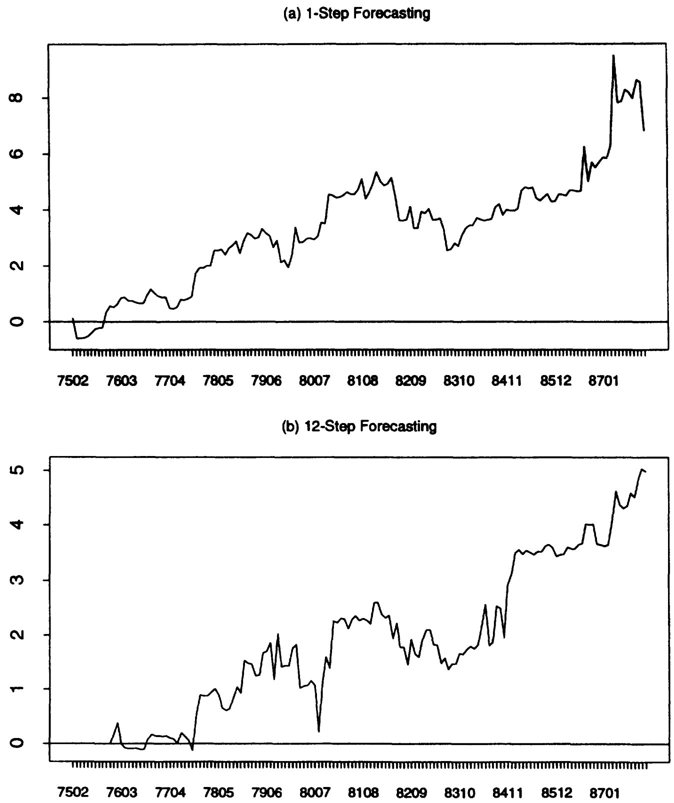

Figure 9. Comparison of Two Trading-Day-Effect Estimation Procedures for Retail Shoe Sales via (18). Graphs are given for forecast leads (a) h = 1 and (b) h = 12. Model 1 in (18) uses irregular component regression estimates of the form (3) and Model 2 the regARIMA estimates of the form (14). These graphs of the accumulating squared forecast error differences show that, at both forecast leads, the squared forecast errors of Model 1 are persistently larger.

(obtained without regARIMA forecast extension). A regARIMA model can be identified for the preadjusted series $Z_t^{(N)} = Y_t/\mathrm{TD}_t^{(N)}, 1 \le t \le N$, possibly after nonlinear transformation to $f(Z_t^{(N)})$. Then, with this model, out-of-sample forecasts $Y_{M+h|M}$ for each $M = N_0, \ldots, N-h$ can be calculated by the following steps:

1. Do trading-day estimation from the irregulars series of $Y_1, \ldots, Y_M$ to obtain $\mathrm{TD}_t^{(M)}$ for $t = 1, \ldots M$ and $t = M+h$.
2. Let $Z_t = Y_t/\mathrm{TD}_t^{(M)}, 1 \le t \le M$. Calculate the out-of-sample forecast $Z_{M+h|M}$ as in Section 3.2.2 from the regARIMA model after estimating its parameters using only the data $z_t = f(Z_t), 1 \le t \le M$.
3. Calculate the forecast $Y_{M+h|M} = \mathrm{TD}_{M+h}^{(M)} Z_{M+h|M}$.

The error $e_{M+h|M}$ associated with this forecast is defined to be $Y_{M+h} - Y_{M+h|M}$ provided that all AO, LS, and ramp regressors in the regARIMA model for $z_1, \ldots, z_N$ are 0 at time $M + h$. Otherwise, define

$$e_{M+h|M} = \mathrm{TD}_{M+h}^{(M)} f^{-1}(\tilde{z}_{M+h}) - Y_{M+h|M},$$

where $\tilde{z}_{M+h}$ denotes the result of subtracting these regression effects from z_{M+h}. The sequence $\mathrm{SS}_{h,M}$ of accumulating sums of the squared errors $e^2_{N_0+h|N_0}, \ldots, e^2_{N|N-h}$ is defined by (17). Comparisons between competing trading-day estimation approaches are made with graphs of the normalized differences $\mathrm{SS}_M^{1,2}$ defined in (18). In the comparisons we present, the model for Y_t that incorporates preadjustment by the X-11 trading-day factors (2) is designated Model 1, and the model of the form (14) is Model 2. Therefore, decreasing graphs favor the irregulars-regression component estimates and increasing graphs favor the regARIMA model estimates.

In the study by Chen and Findley of X-12-ARIMA's various regARIMA trading-day models (Chen and Findley 1993, 1996a), there were 41 series for which the regARIMA analog (14) of (3) was preferred over models that gave estimates of a coefficient of the Leap Year regressor of Table 2 or that ignored length-of-month effects. For these 41 series, it is natural to compare the approaches (14) and (3). This was done via graphs of (18) for lags $h = 1, 12$. Only for eight series was one approach found better than the other: The regARIMA trading-day model was favored five times and the irregulars-regression model three times. We present two examples of graphs of (18), one for each preference.

Figure 9 shows that, for the series of retail sales from U.S. shoe stores up to 1989, the regARIMA trading-day estimates lead to persistently better one-month and twelve-month forecasts than the irregulars-regression estimates. By contrast, Figure 10 reveals that, for the series of values of U.S. factory shipments of communications equipment up to 1983, the one-month forecasts via the irregulars-regression

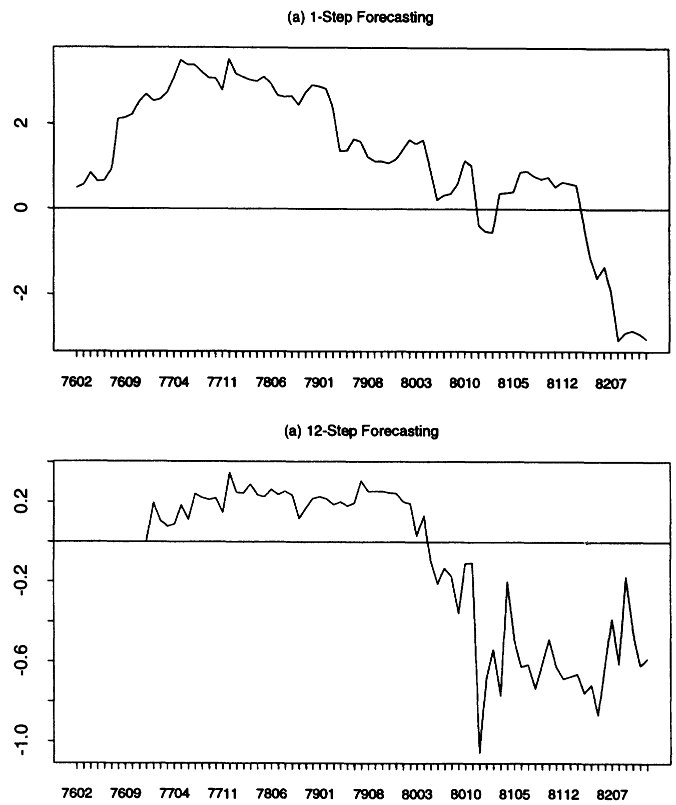

Figure 10. Comparison of Two Trading-Day-Effect Estimation Procedures for Values of Shipments of Communications Equipment. The graphs are analogous to Figure 9. Here, only for one-step forecasting are there indications of a recurring difference in performance: In an average sense, the squared forecast errors of Model 1 are smaller than those of Model 2 after 1976.

trading-day estimates are persistently better after 1977, and the twelve-month forecasts are at least as good, on average, as those of the regARIMA trading-day model.

There were no series for which the irregulars-regression model had persistently better twelve-month forecasting and five series for which its performance at lead 12 was persistently worse. Because $h = 12$ is usually the most important forecast lead for forecast extension, we have concluded that the regARIMA model approach of (14) should be the first approach considered, instead of that of (2), when a reasonably well-fitting regARIMA model is available. (This approach also provides the advantages AIC has over the irregular regression F statistic that we described in Section 1.4.3.)

Similar comparisons of Easter-holiday-effect models estimated from the irregulars and from the observed series were given by Chen and Findley (1996b). For all calendar-effect model comparisons, including comparison of a model with such a regressor to one without, each of the diagnostics (18), sliding spans, revision histories, and AIC histories can provide useful information.

4.4 Using AIC Histories to Decide Between Preadjustments

Findley and Monsell (1989) considered the problem of comparing a set of "subjective" preadjustments with a set of "objective" preadjustments for the series of numbers of units of autos sold multiplied by an average price for each type of car. The values of this series from 1979 on are graphed in Figure 11. The preadjustments were intended to remove the effects of special, short-duration sales programs involving cash rebates to purchasers. These programs were used by the automobile manufacturers to reduce their inventories of unsold cars. Such programs cause a large increase in sales in the month or two in which they occur, followed by a substantial decrease in the subsequent month or two. If such programs occur in the same month several years in a row, then seasonal-adjustment procedures incorporate much of their effects into the seasonal factors. This is not correct when it is known that the programs did not recur in later years. To prevent this distortion of the seasonal factors, it is necessary to estimate the effects of these sales programs and remove them from the series prior to seasonal factor calculation.

An expert analyst used her knowledge of the dates of sales programs to select values of the irregular component of an X-11 seasonal decomposition that she averaged to obtain the preadjustment factors (divisors) for sales-program effects that are graphed in Figure 11b. When she asked us about this approach, we were concerned that the irregulars series would be an unreliable source of information about these effects because of distortions in the seasonal component induced by the sales programs. As an alternative, we constructed five user-defined regressors to estimate sales-program effects in the years 1985–1987—one regressor each for the months of August, September, October, and November and a single regressor for December 1986 and January 1987. To conform to the analyst's specification of identical effects for the same calendar month in successive years in which the month is affected, the regressors for August, October, and November each had the value 1 in their month for 1985–1987 and the value 0 in all other months. The September regressor deviated from the analyst's pattern by having the value 1 in September of 1985 and 1986 but 0 in September of 1987 (and elsewhere), a deviation strongly preferred by AIC. The fifth regressor had the values 1 in December 1986, -1 in January 1987, and 0 elsewhere. We naively assumed that the automatic outlier-identification procedure described in Section 3.1 and Appendix C would deal effectively with any important sales-program effects in months prior to 1985, where the analyst had made numerous smaller preadjustments (the later data were of greater interest). Our objectively obtained divisors, estimated from a regARIMA model with the regressors just described, are graphed in Figure 11(c) along with automatically identified AO adjustments.

Findley and Monsell (1989) reported that the AIC value of this model was smaller by 17.6 than the AIC of the regARIMA model found for the series with the subjective adjustments, indicating a strong preference for the objective approach. A subsequent analysis of AIC's preferences over time using the diagnostic (20), however, showed that prior to early 1985 the subjective adjustments were preferred.

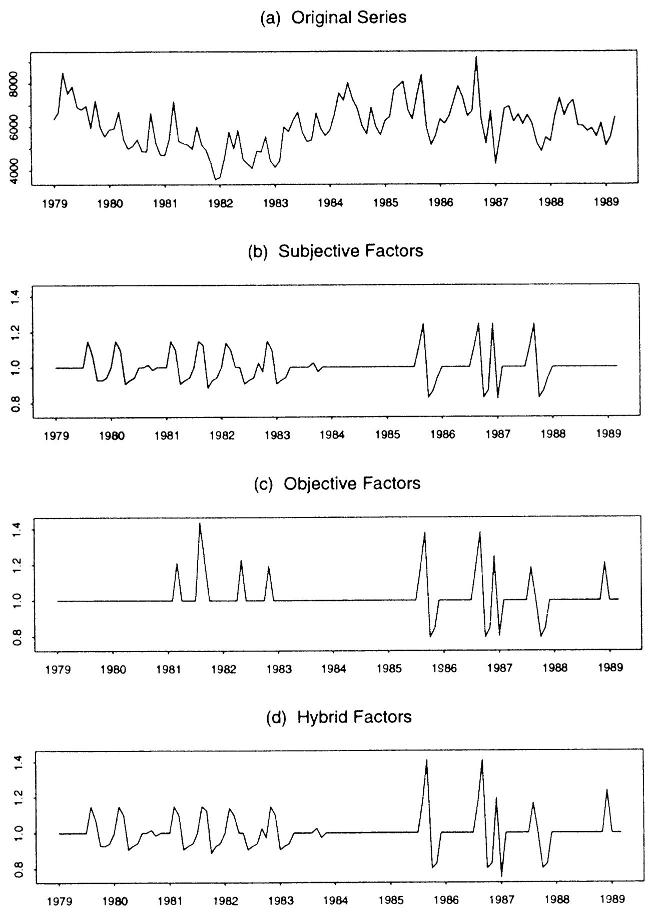

Figure 11. Monthly Auto Sales by Units (a) and Three Competing Series of Adjustment Factors Estimating the Effects of Manufacturers' Sales Campaigns. The "subjective" factors (b) were informally derived by a subject-matter expert from an irregular component. The objective factors (c) came from a regARIMA model. The hybrid factors (c) are subjective factors up to 1985 and objective factors thereafter.

Then W. P. Cleveland, who also had the analyst's adjustments, kindly pointed out to us that there were errors in our divisor set. (The divisors in Fig. 11(b) are the correct ones.) Thus, it was appropriate to redo our analysis. Labeling as Model 1 the regARIMA model that produced the objective factors of Figure 11(c) and as Model 2 the regARIMA model using only subjective preadjustments (the automatic outlier-identification procedure found no outliers), the graph of the history of the AIC differences (20) given in Figure 12(a) leads to a conclusion similar to the earlier one: Overall, the objective adjustments are favored (the final AIC difference is -13.7), but for a several-year period beginning in 1983 the subjective adjustments are better.

Therefore, we decided to try to replace Model 1 with a better model. We did not want to add a large number of regressors to imitate the analyst's adjustments prior to 1985. So we fit a hybrid model, in which the subjective adjustments were applied prior to 1985, and the user-defined regressors were used thereafter, together with any automatically identified AO's. (There were two such AO's, one at February 1975 and the other at December 1988.) The resulting divisors are graphed in Figure 11(d). The AIC difference history favored the hybrid model throughout. Knowing that AO regressors can have a large impact on AIC values, however, we wanted to determine if this conclusion depended substantially on the AO's included in the hybrid model but not in the subjective model. To investigate this, we augmented the latter model with the same two outlier regressors. In the augmented model, the t statistics of these two AO regressors were below the critical value used in the automatic outlier procedure but above 2.0. The AIC difference graph comparing the hybrid model and the augmented subjective model is given as Figure 12(b). For the calculation of (20), the hybrid model is labeled Model 1 and the augmented subjective model is Model 2. The graph shows that the hybrid model is still consistently preferred. The final AIC difference is -11.4 (about half of what its value had been before the subjective model was augmented).

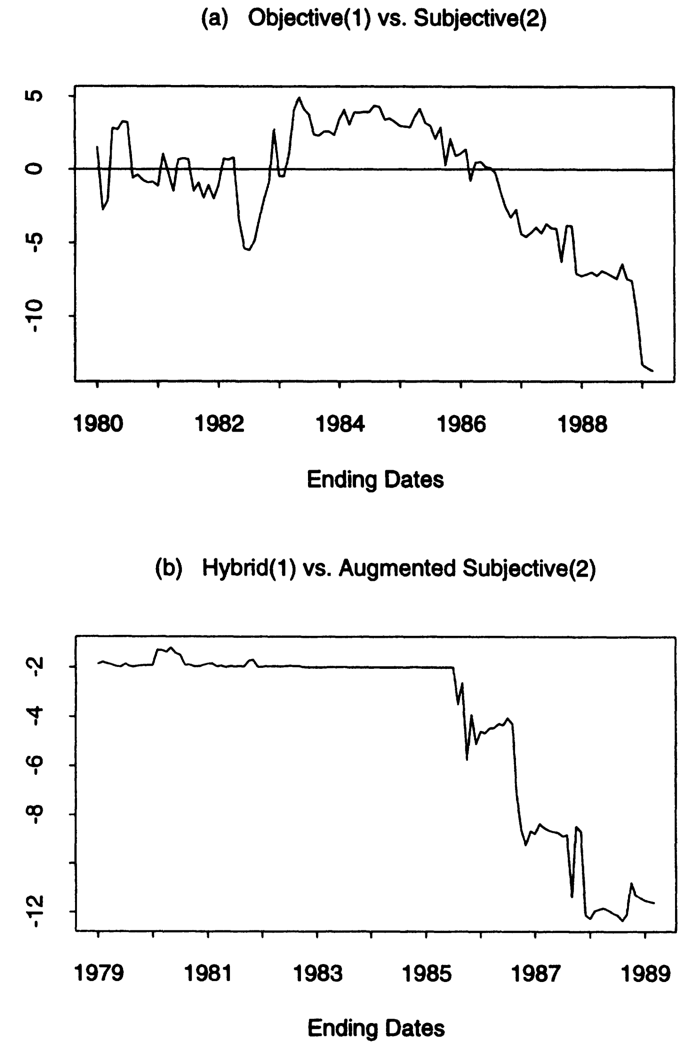

Figure 12. AIC Difference Histories (20) Comparing Two Pairs of Models That Use Different Adjustment Factors of Figure 12. In (a), Model 1 uses the objective factors and Model 2 the subjective factors. In (b), Model 1 uses the hybrid factors and Model 2 the subjective factors and also the two AO variables used by Model 1. Only in (b) do the AIC differences have a consistent sign, indicating a consistent preference for Model 1. In this sense, the hybrid factors are preferred.

These analyses demonstrate the power of the AIC history diagnostic to enable difficult model comparisons and to identify ways in which models under consideration can be improved. Note that the forecast performance diagnostics used in Section 4.3 are not applicable to the model comparisons of this subsection because the models cannot forecast the effects of interest.

5. THE USER INTERFACE: THREE EXAMPLES

Because the X-12-ARIMA program is designed for use with a broad variety of operating systems, its interface uses command files rather than windows and menus. We made substantial efforts to design a command structure that is largely self-descriptive and easy to do standard runs with. The latter is especially important because the program has very many adjustment and input/output options, yet its users should be able to deal with most series knowing just a few options. We now present some examples to give the reader a feeling for the interface.

5.1 Processing a Single Series

The simplest situation is that in which a single series is to be adjusted using default options. Suppose the series named `myseries` is stored in free format in a file named `Xfile.dat` in the same directory as the X-12-ARIMA program, along with the command file. The command file will be named `myseries.spc` and must have an extension `.spc`, chosen to connote "specifications."

The commands for the basic situation, in which regARIMA models are not used, and the program acts like the X-11 program in its default setting [except that the seasonal-filter length-selection criterion of Lothian (1984) is used] are as follows, assuming `myseries` begins in March of 1984:

```
series {start = 1984.3 file = "Xfile.dat"}
x11 { }
```

To execute this `.spc` file, the command `x12a myseries` is entered. When the execution is finished, the program writes the default output to a file named `myseries.out` in the same directory.

Suppose we wish instead to have what is essentially a default X-11-ARIMA run, with 12-month forecast and backcast extension from an ARIMA model selected from the default list of models. Suppose also, though, that we want to let AO's and LS's be automatically identified (using the default t statistic critical value 3.3 as described in Section 3.1 and Appendix C). The regression estimates of all identified AO's and LS's are to be adjusted out of the series before seasonal factor calculations begin. Because the default seasonal-adjustment decomposition is multiplicative, the log transform is chosen for the regARIMA models. For the series of the previous example, the commands in the specification file become

```
series {start = 1984.3 file = "Xfile.dat"}
  transform {function = log}
  automdl { }
  outlier { }
  x11 { }
```

5.2 Processing Many Series

There are features that facilitate running the program on many series with many `.spc` files, but we shall illustrate only the simple situation in which the same options, stored in a single `.spc` file, are used for many series, all of which have the same starting date and are stored in files with the same format. This can occur when a group of related series is examined for seasonality for the first time. It is also the natural situation when simulation experiments are done to investigate properties of seasonal adjustments or adjustment procedures. Examples include the irregular-component resampling approach to obtain standard errors for seasonal adjustments described by Findley and Monsell (1990) and the use of simulations to obtain confidence intervals for the estimated duration of the Easter effect, described by Chen and Findley (1996b). Studies using simulations to analyze sources of nonlinearity in the X-11 procedure were performed by Ghysels et al. (1996) and Findley (1996). With one `.spc` file for many series, the names of the input files are listed in a file that, in X-12-ARIMA terminology, is called a *data metafile.*

Assume that we have 500 monthly series, `sim1,...,` `sim500`, of the same length, all starting in January 1970, all stored with FORTRAN format (12F6.0) in the files named `sim1.dat,...,sim500.dat`. Suppose we wish to seasonally adjust them with 3×9 seasonal moving averages and 17-term Henderson trend filters, after fitting the ARIMA (0, 1, 3)(0, 1, 1)$_{12}$ model without a lag 2 moving average term (via exact Gaussian likelihood maximization, the default estimation method) to extend each series with 60 forecasts. To accomplish this, we create a data metafile whose name has the extension `dta`, say `sim.dta`, containing the data file names,

```
sim1.dat
   ⋮
sim500.dat.
```

In a `.spc` file, that we shall name `series.spc`, we place the commands

```
series {start = 1970.jan format = "(6f12.0)"}
transform {function = log}
arima {model = (0 1 [1 3])(0 1 1)}
forecast {maxlead = 60}
x11 {seasonalma = s3 × 9 trendma = 17}
```

In this context, with a data metafile named `sim.dta`, the command to execute the program is `x12a series -d sim`. The `-d` flag informs the program of the data metafile.

6. CONCLUDING REMARKS

The X-12-ARIMA program and its user's manual can be downloaded via `ftp` from the Internet address `ftp.census.gov`. The FORTRAN source code is available, as are executable versions for five platforms, DOS PC's, and SUN, Hewlett Packard, DEC Alpha, and DEC VAX workstations, in individual subdirectories of the directory `pub/ts/x12a`. This `ftp` site also has a version customized by Margaret Keating of the Federal Reserve Board for the FAME time series database system. We hope that the easy availability of a versatile program for seasonal adjustment, regARIMA modeling, and model selection will stimulate many statisticians, economic modelers, and economic analysts to undertake refined analyses of seasonal and calendar effects in their data. This would have important indirect benefits: A substantial increase in the number of economic data users having expertise in seasonal adjustment would lead to a more sophisticated use of adjusted data and would stimulate the development of improved adjustment diagnostics, methods, and practices. (The SEATS and TRAMO programs are available from `http://www.bde.es`.)

The most obvious and important feature lacking in X-12-ARIMA is high-resolution graphical diagnostics. Graphical diagnostics for seasonal adjustment are an area that is ripe for further research, 15 years after the pioneering work done by the authors of SABL. We expect to begin work soon on the development of a separate program to produce such graphics from X-12-ARIMA output, one that will be usable on a variety of computer platforms.

ACKNOWLEDGMENTS

We are very grateful to the editors of *JBES,* Ruey Tsay and Mark Watson, for giving us this opportunity to have the innovations of X-12-ARIMA presented and discussed. We also thank Matthew Kramer for his many helpful comments on several drafts of this article.

APPENDIX A: THE PROTOTYPE X-11 DEFAULT CALCULATIONS

Calculations are shown with X-11's default seasonal filter choices in Step (c) of Stages 1 and 2. Calculations used to reduce the influence of "extreme" values on seasonal factors are omitted. Forecasts and backcasts are required to enable the symmetric filters shown to be used near the ends of the series.

Y_t denotes a monthly series with no extreme values; extended by forecasts and backcasts so that modified formulas are not needed at the series' ends.

Three types of decomposition of Y_t into trend (T_t), seasonal (S_t), and irregular (I_t) components are presented:

$$\begin{array}{ll} \text{Multiplicative (M):} & Y_t = T_t S_t I_t \\ \text{Additive (A):} & Y_t = T_t + S_t + I_t \\ \text{Pseudo-Additive (PA):} & Y_t = T_t(S_t + I_t - 1). \end{array}$$

Stage 1. Initial Estimates

(a) Initial trend estimate via "centered 12-term" (13-term) moving averages:

$$T_t^{(1)} = \frac{1}{24}Y_{t-6} + \frac{1}{12}Y_{t-5} + \cdots + \frac{1}{12}Y_t + \cdots + \frac{1}{12}Y_{t+5} + \frac{1}{24}Y_{t+6}.$$

(b) Initial "SI ratio":

$$\begin{array}{ll} \text{(M, PA):} & \mathrm{SI}_t^{(1)} = Y_t/T_t^{(1)}. \\ \text{(A):} & \mathrm{SI}_t^{(1)} = Y_t - T_t^{(1)}. \end{array}$$

(c) Initial preliminary seasonal factor via "3×3" seasonal moving average:

$$\hat{S}_t^{(1)} = \frac{1}{9}\,\mathrm{SI}_{t-24}^{(1)} + \frac{2}{9}\,\mathrm{SI}_{t-12}^{(1)} + \frac{3}{9}\,\mathrm{SI}_t^{(1)} + \frac{2}{9}\,\mathrm{SI}_{t+12}^{(1)} + \frac{1}{9}\,\mathrm{SI}_{t+24}^{(1)}.$$

(d) Initial seasonal factor:

$$\text{(M, PA): } S_t^{(1)} = \frac{\hat{S}_t^{(1)}}{\frac{1}{24}\hat{S}_{t-6}^{(1)} + \frac{1}{12}\hat{S}_{t-5}^{(1)} + \cdots + \frac{1}{12}\hat{S}_{t+5}^{(1)} + \frac{1}{24}\hat{S}_{t+6}^{(1)}}$$

$$\text{(A): } S_t^{(1)} = \hat{S}_t^{(1)} - \left(\frac{\hat{S}_{t-6}^{(1)}}{24} + \frac{\hat{S}_{t-5}^{(1)}}{12} + \cdots + \frac{\hat{S}_{t+5}^{(1)}}{12} + \frac{\hat{S}_{t+6}^{(1)}}{24}\right).$$

(e) Initial seasonal adjustment:

$$\begin{array}{ll} \text{(M):} & A_t^{(1)} = \frac{Y_t}{S_t^{(1)}} \\ \text{(A):} & A_t^{(1)} = Y_t - S_t^{(1)} \\ \text{(PA):} & A_t^{(1)} = Y_t - T_t^{(1)}(S_t^{(1)} - 1). \end{array}$$

Stage 2. Seasonal Factors and Seasonal Adjustment

The $(2H+1)$-term Henderson coefficients (see Appendix B) are designated $h_j^{(2H+1)}, -H \le j \le H$.

(a) Intermediate trend: For data-determined H (see Appendix B),

$$T_t^{(2)} = \sum_{j=-H}^{H} h_j^{(2H+1)} A_{t+j}^{(1)}.$$

(b)

$$\begin{array}{ll} \text{(M, PA):} & \mathrm{SI}_t^{(2)} = Y_t/T_t^{(2)} \\ \text{(A):} & \mathrm{SI}_t^{(2)} = Y_t - T_t^{(2)}. \end{array}$$

(c) Preliminary seasonal factor via "3×5" seasonal moving average:

$$\hat{S}_t^{(2)} = \frac{1}{15}\,\mathrm{SI}_{t-36}^{(2)} + \frac{2}{15}\,\mathrm{SI}_{t-24}^{(2)} + \frac{3}{15}\,\mathrm{SI}_{t-12}^{(2)} + \frac{3}{15}\,\mathrm{SI}_t^{(2)} + \frac{3}{15}\,\mathrm{SI}_{t+12}^{(2)} + \frac{2}{15}\,\mathrm{SI}_{t+24}^{(2)} + \frac{1}{15}\,\mathrm{SI}_{t+36}^{(2)}.$$

(d) Seasonal factor:

$$\text{(M, PA): } S_t^{(2)} = \frac{\hat{S}_t^{(2)}}{\frac{1}{24}\hat{S}_{t-6}^{(2)} + \frac{1}{12}\hat{S}_{t-5}^{(2)} + \cdots + \frac{1}{12}\hat{S}_{t+5}^{(2)} + \frac{1}{24}\hat{S}_{t+6}^{(2)}}$$

$$\text{(A): } S_t^{(2)} = \hat{S}_t^{(2)} - \left(\frac{\hat{S}_{t-6}^{(2)}}{24} + \frac{\hat{S}_{t-5}^{(2)}}{12} + \cdots + \frac{\hat{S}_{t+5}^{(2)}}{12} + \frac{\hat{S}_{t+6}^{(2)}}{24}\right).$$

(e) Seasonal adjustment:

$$\begin{array}{ll} \text{(M):} & A_t^{(2)} = \frac{Y_t}{S_t^{(2)}} \\ \text{(A):} & A_t^{(2)} = Y_t - S_t^{(2)} \\ \text{(PA):} & A_t^{(2)} = Y_t - T_t^{(2)}(S_t^{(2)} - 1). \end{array}$$

Stage 3. Final Henderson Trend and Final Irregular

(a) Final trend: For data-determined H, possibly different from Stage 2(a):

$$T_t^{(3)} = \sum_{j=-H}^{H} h_j^{2H+1} A_{t+j}^{(2)}.$$

(b) Final irregular:

$$\begin{array}{ll} \text{(M, PA):} & I_t^{(3)} = \frac{A_t^{(2)}}{T_t^{(3)}} \\ \text{(A):} & I_t^{(3)} = A_t^{(2)} - T_t^{(3)}. \end{array}$$

Estimated decomposition:

$$\begin{array}{ll} \text{(M):} & Y_t = T_t^{(3)} S_t^{(2)} I_t^{(3)} \\ \text{(A):} & Y_t = T_t^{(3)} + S_t^{(2)} + I_t^{(3)} \\ \text{(PA):} & Y_t = T_t^{(2)}(S_t^{(2)} - 1) + T_t^{(3)} I_t^{(3)}. \end{array}$$

APPENDIX B: HENDERSON FILTERS AND MUSGRAVE SURROGATES

To provide a larger context for our discussion of the criterion used to obtain many of the asymmetric filters and to complete the description of the default X-11 procedure of Appendix A, we start by outlining how the symmetric Henderson filter coefficients $h_j^{(2H+1)}$ are derived.

B.1 The Symmetric Henderson Filters

In the appendix of Kenny and Durbin (1982), an insightful derivation was given of the coefficients $h_j^{(2H+1)}$ of the symmetric Henderson filter and of the equivalence of Henderson's alternative criteria for determining them. It was observed by Gray and Thomson (1996) that a slight modification of Kenny and Durbin's argument yields two improvements—One need not assume a priori that the coefficients are symmetric, and it is enough to require the filters to reproduce quadratic trends instead of cubic trends. We summarize Gray and Thomson's approach. Suppose that, for $-(H+3) \leq j \leq H$,

$$A_{t+j} = \alpha + \beta(t+j) + \gamma(t+j)^2 + I_{t+j}, \qquad \text{(B.1)}$$

where the I_{t+j} are Gaussian variates with mean 0 and variance σ^2, which, for different j, are independent. We only consider filters $h_j, -H \leq j \leq H$, that provide unbiased estimates of the value of the quadratic trend at time t. Equivalently, when the I_{t+j} in (B.1) are 0 for all j, we require

$$\sum_{j=-H}^{H} h_j A_{t+j} = \alpha + \beta t + \gamma t^2 \qquad \text{(B.2)}$$

for any values of α, β, and γ. Let Δ denote the differencing operator so that $\Delta A_t = A_t - A_{t-1}$ and $\Delta h_j = h_j - h_{j-1}$. Let E denote expectation. Then, among filters satisfying (B.2), the Henderson filter is the minimizer of the smoothness measure $E(\Delta^3 \sum_{j=-H}^{H} h_j A_{t+j})^2$. This can be reduced to a smoothness measure on the filter coefficients,

$$E\left(\Delta^3 \sum_{j=-H}^{H} h_j A_{t+j}\right)^2 = \sigma^2 \sum_{j=-H}^{H+3} (\Delta^3 h_j)^2,$$

if we define $h_j = 0$ for $j = \pm(H+1), \pm(H+2), \pm(H+3)$. (On the left, Δ^3 is applied to the A_{t+j}; on the right, to the h_j.) With $q_j(H) = \{(H+1)^2 - j^2\}\{(H+2)^2 - j^2\}\{(H+3)^2 - j^2\}$, and with a and b determined by

$$a \sum_{j=-H}^{H} q_j(H) + b \sum_{j=-H}^{H} q_j(H) j^2 = 1$$

and

$$a \sum_{j=-H}^{H} q_j(H) j^2 + b \sum_{j=-H}^{H} q_j(H) j^4 = 0,$$

the Henderson coefficients are given by $h_j^{(2H+1)} = q_j(H)(a + bj^2), -H \leq j \leq H$. This formula shows they are symmetric, $h_j^{(2H+1)} = h_{-j}^{(2H+1)}$. They can be < 0.

B.2 Musgrave's Criterion for Asymmetric Surrogates

So that the following discussion can encompass both trend and seasonal filters, we change to a neutral notation, W_j, for the original filter coefficients and X_{t+j} for the variates to which they are applied. For given $X_1, \ldots, X_T$ and positive integer J such that $2J + 1 \leq T$, we can calculate $\sum_{j=-J}^{J} W_j X_{t+j}$ only for indices t satisfying $J + 1 \leq t \leq T - J$. To obtain coefficients for calculating filtered values at the remaining times t, Musgrave (1964) applied a minimum mean squared revision criterion to the case in which the X_t are independent Gaussian variates with variance σ^2 and with a linear mean function, $EX_t = \alpha + \beta t$ [in contrast to the quadratic mean in (B.1)]. More precisely, if $t = T - J + d$ with $1 \leq d \leq J$, Musgrave's strategy [independently deduced by Laniel (1986)] is to find the asymmetric filter whose coefficients $V_j, -J \leq j \leq J - d$, sum to 1 and also minimize

$$E\left(\sum_{j=-J}^{J} W_j X_{t+j} - \sum_{j=-J}^{J-d} V_j X_{t+j}\right)^2.$$

In an unpublished report, Doherty (1992) derived an explicit formula for the coefficients of these abbreviated filters that has been implemented in X-12-ARIMA. To have a convenient form, we change notation. Set $N = 2J + 1, M = N - d$, and, for $1 \leq j \leq N$, define $w_j = W_{j-1-J}$ and $x_j = X_{t-J+(j-1)}$. We are assuming that $x_j = \gamma + \beta j + I_j$, where the I_j are independent Gaussian variates with mean 0 and variance σ^2. Define $\bar{\Delta} = |Ex_j - Ex_{j-1}| (= |\beta|), \bar{I} = E|I_j - I_{j-1}| (= 2\sigma/\sqrt{\pi})$, and $R = \bar{\Delta}/\bar{I}$. Then the coefficients $v_j, 1 \leq j \leq M$, satisfying $\sum_{j=1}^{M} v_j = 1$ that minimize

$$E\left(\sum_{j=1}^{N} w_j x_j - \sum_{j=1}^{M} v_j x_j\right)^2$$

are given by

$$\begin{aligned} v_j &= w_j + \frac{1}{M} \sum_{i=M+1}^{N} w_i \\ &\quad + \frac{\left(j - \left(\frac{M+1}{2}\right)\right) \frac{4}{\pi} R^2}{1 + \frac{M(M-1)(M+1)}{12} \frac{4}{\pi} R^2} \\ &\quad \times \sum_{i=M+1}^{N} \left(i - \left(\frac{M+1}{2}\right)\right) w_i. \end{aligned} \qquad \text{(B.3)}$$

Doherty (1992) also derived a formula for the v_j when no assumption is made about the form of the mean function of the x_j. With symmetric filters, $w_{N+1-j} = w_j$, and from this property it follows that the time-reversed filter coefficients $v_j^R = v_{N+1-j}, d+1 \leq j \leq N$ minimize $E(\sum_{j=1}^{N} w_j x_j - \sum_{j=d+1}^{N} v_j^R x_j)^2$ and therefore provide the surrogates for symmetric filters near the beginning of the time series.

In X-12-ARIMA, to obtain the default asymmetric surrogates for the 9-term and 13-term monthly Henderson filters, R^{-1} is set equal to .99 and 3.5, respectively. For longer Henderson filters, $R^{-1} = 4.5$ is used. For the 5-term and 7-term quarterly filters, .001 and 4.5, respectively, are used. For the 3×9 seasonal filter, with the time index j having units of years, $R^{-1} = 9.5$ is used in (B.3) to determine asymmetric surrogates.

Finally, we explain how the lengths are determined for the Henderson filters used in (a) of Stages 2 and 3 of Appendix A. In each of these stages, an estimate $\hat{R}^{-1}$ of R^{-1} is calculated as follows. Let $\hat{T}_t$ denote the 13-term symmetric Henderson trend of the available seasonally adjusted series ($A_t^{(1)}$ in Stage 2, $A_t^{(2)}$ in Stage 3), and let $\hat{I}_t$ denote the irregular component resulting from removing this trend estimate from the seasonally adjusted series. With $\bar{C}$ denoting the sample mean of the available values of the absolute trend changes $|\hat{T}_t - \hat{T}_{t-1}|$ and $\bar{I}$ the sample mean of the $|\hat{I}_t - \hat{I}_{t-1}|$, the value of the *noise-to-signal ratio*, $\hat{R}^{-1} = \bar{I}/C$, called the I-bar C-bar ratio, determines the value of $2H + 1$ as follows. If $\hat{R}^{-1} < 1.0$, the 9-term Henderson filter is used. Otherwise, in Stage 2, the 13-term filter is used. In Stage 3, the 13-term filter is used when $1 < \hat{R}^{-1} < 3.5$, but the 23-term Henderson filter is used when $\hat{R}^{-1} \geq 3.5$. This procedure is called the *X-11 variable trend cycle routine*.

APPENDIX C: THE PROCEDURE FOR AO AND LS DETECTION

We describe the AO and LS detection procedure mentioned in Section 3.2. The algorithm proceeds from critical values $\gamma^{\text{ao}}, \gamma^{\text{ls}}$ specified separately for the AO- and LS-regressor t statistics, denoted $T_t^{\text{ao}}, T_t^{\text{ls}}$, that are calculated at each time point t in each iteration of the forward addition cycle to be described later. (The default values are $\gamma^{\text{ao}} = \gamma^{\text{ls}} = 3.8$.) Let T_t denote one of these statistics, γ the corresponding critical value, β and ψ the vectors of regression coefficients and ARMA model coefficients, and $a_t(\beta, \psi)$ the estimates of the innovations a_t determined by these coefficients in the model (10). AO regressors are available for all observation times $1 \leq t \leq N$, LS regressors for all but the first two and the last of these times.

Initialization: Estimate the coefficients (β, ψ) of the model (10) specified by the user. If the model includes prespecified AO and LS regressors, these will always be kept in the regressor set.

Forward Addition:

1. Calculate the robust standard error, $\sigma_a^R = 1.49\times$ $\text{median}_t|a_t(\beta, \psi)|$, for the current β, ψ.
2. Using the current ψ and σ_a^R, calculate the values of T_t for all AO and LS regressors not currently in the model—that is, excluding those prespecified or already identified in forward addition. (To calculate T_t for any given outlier regressor, the generalized least squares estimation determined by ψ is carried out for the regressors in the model plus the given outlier regressor.)
3. Determine the outlier regressor with maximum $|T_t|$.
4. If $\max|T_t| \geq \gamma$, add this regressor to the model and reestimate β and ψ. Otherwise, stop.

Repeat Steps 1–4 until there are no additional outliers satisfying $|T_t| \geq \gamma$.

Backward Deletion: Start with the model including all outlier regressors added in the forward addition stage.

1. Calculate maximum likelihood estimates of (β, ψ, σ_a).
2. Using the estimated (β, ψ, σ_a), calculate T_t for all AO and LS regressors identified in forward addition that remain in the model. Determine which of these regressors has $\min|T_t|$.
3. If $\min|T_t| < \gamma$, delete this regressor from the model and go to 1. Otherwise, stop.

An alternative procedure is available that, at Step 4 of forward addition, adds to the model *all* outlier regressors with $|T_t| \geq \gamma$.

[Received May 1996. Revised November 1997.]

REFERENCES

Abraham, B., and Ledolter, J. (1983), *Statistical Methods for Forecasting*, New York: Wiley.

Akaike, H. (1973), "Information Theory and an Extension of the Likelihood Principle," in *Second International Symposium on Information Theory*, eds. B. N. Petrov and F. Czaki, Budapest: Akademia Kiado, pp. 267–287. (Reprinted in *Breakthroughs in Statistics, Vol. I*, eds. S. Kotz and N. L. Johnson, New York: Springer-Verlag, 1991, pp. 599–624.)

——— (1980), "Seasonal Adjustment by a Bayesian Modeling," *Journal of Time Series Analysis*, 1, 1–13.

Akaike, H., and Ishiguro, M. (1983), "Comparative Study of the X-11 and BAYSEA Procedures of Seasonal Adjustment," in *Applied Time Series Analysis of Economic Data*, ed. A. Zellner, Washington, DC: U.S. Department of Commerce, Bureau of the Census, pp. 17–23.

Ansley, C. F., Spivey, W. A., and Wrobleski, W. J. (1977), "A Class of Transformations for Box–Jenkins Seasonal Models," *Applied Statistics*, 26, 173–178.

Battipaglia, P., and Focarelli, D. (1995), "A Comparison of Indicators for Evaluating X-11-ARIMA Seasonal Adjustments," technical report, Bank of Italy, Rome, Research Dept.

Bell, W. R. (1983), "A Computer Program for Detecting Outliers in Time Series," in *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, pp. 634–639.

——— (1984), "Seasonal Decomposition of Deterministic Effects," Research Report 84/01, Bureau of the Census, Washington, DC, Statistical Research Division.

——— (1996), "An Overview of REGARIMA Modeling," lecture notes, Bureau of the Census, Washington, DC.

Bell, W. R., and Hillmer, S. C. (1983), "Modeling Time Series With Calendar Variation," *Journal of the American Statistical Association*, 78, 526–534.

——— (1984), "Issues Involved With the Seasonal Adjustment of Economic Time Series" (with discussion), *Journal of Business & Economic Statistics*, 2, 291–349.

Bell, W. R., and Monsell, B. C. (1992), "X-11's Symmetric Linear Filters and Their Transfer Functions," Research Report 92/15, Bureau of the Census, Washington, DC: Statistical Research Division.

Bobbitt, L., and Otto, M. C. (1990), "Effects of Forecasts on the Revisions of Seasonally Adjusted Values Using the X-11 Seasonal Adjustment Procedure," in *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, pp. 449–453.

Box, G. E. P., and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control* (rev. ed.), San Francisco: Holden-Day.

Breiman, L. (1997), "Heuristics of Instability and Stabilization in Model Selection," *The Annals of Statistics*, 24, 250–283.

Brockwell, P. J., and Davis, R. A. (1987), *Time Series: Theory and Methods*, New York: Springer-Verlag.

Bruce, A. G., and Jurke, S. R. (1996), "Non-Gaussian Seasonal Adjustment: X-12-ARIMA Versus Robust Structural Models," *Journal of Forecasting*, 15, 305–328.

Bruce, A. G., and Martin, R. D. (1989), "Leave-k-out Diagnostics for Time Series" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 51, 363–424.

Burman, J. P. (1983), Comment on "Modeling Considerations in the Seasonal Adjustment of Economic Time Series," by S. C. Hillmer, W. R. Bell, and G. C. Tiao, in *Applied Time Series Analysis of Economic Data*, ed. A. Zellner, Washington, DC: U.S. Department of Commerce, Bureau of the Census, pp. 101–105.

Burman, J. P., and Otto, M. C. (1988), "Outliers in Time Series," Research Report 88/14, Bureau of the Census, Washington, DC, Statistical Research Division.

Chang, I., and Tiao, G. C. (1983), "Estimation of Time Series Parameters in the Presence of Outliers," Technical Report 8, University of Chicago, Statistics Research Center.

Chang, I., Tiao, G. C., and Chen, C. (1988), "Estimation of Time Series Parameters in the Presence of Outliers," *Technometrics*, 30, 193–204.

Chen, B.-C., and Findley, D. F. (1993), "Multiplicative Trading Day Adjustments: X-11 and REGARIMA Compared," in *Proceedings of the Business and Economics Statistics Section, American Statistical Association*, pp. 220–225.

——— (1996a), "Models for Trading Day Adjustment of Economic Time

Series, Both Flow and Stock: An Empirical Study," technical report, Bureau of the Census, Washington, DC.

——— (1996b), "Comparison of X-11 and REGARIMA Easter Holiday Adjustments," technical report, Bureau of the Census, Washington, DC.

Cleveland, W. P., and Grupe, M. R. (1983), "Modeling Time Series When Calendar Effects Are Present," in *Applied Time Series Analysis of Economic Data*, ed. A. Zellner, Washington, DC: U.S. Department of Commerce, Bureau of the Census, pp. 57–67.

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

Cleveland, W. S. (1983), "Seasonal and Calendar Adjustment," in *Handbook of Statistics 3: Time Series Analysis in the Frequency Domain*, eds. D. R. Brillinger and P. R. Krishnaiah, Amsterdam: Elsevier, pp. 39–72.

Cleveland, W. S., and Devlin, S. J. (1982), "Calendar Effects in Monthly Time Series: Modeling and Adjustment," *Journal of the American Statistical Association*, 77, 520–528.

Cleveland, W. S., Devlin, S. J., Schapira, D. R., and Terpenning, I. J. (1981), *The SABL Computer Package: Users Manual*, Murray Hill, NJ: Bell Laboratories.

Cook, R. D., and Weisberg, S. (1982), *Residuals and Influence in Regression*, New York: Chapman and Hall.

Dagum, E. B. (1980), *The X-11-ARIMA Seasonal Adjustment Method* (No. 12-564E), Ottawa: Statistics Canada.

——— (1987), "Monthly Versus Annual Revisions of Concurrent Seasonally Adjusted Series," in *Time Series and Econometric Modeling*, eds. I. B. MacNeill and G. J. Umphrey, Dordrecht: D. Reidel, pp. 131–146.

——— (1988), *The X-11-ARIMA/88 Seasonal Adjustment Method—Foundations and User's Manual*, Ottawa: Statistics Canada.

Dagum, E. B., Huot, G., and Morry, M. (1988), "Seasonal Adjustment in the Eighties: Some Problems and Solutions," *The Canadian Journal of Statistics*, 16 (Supplement), 109–126.

Doherty, M. (1992), "The Surrogate Henderson Filters in X-11," technical report, Statistics New Zealand, Wellington.

Findley, D. F. (1985), "On the Unbiasedness Property of AIC for Exact or Approximating Linear Stochastic Time Series Models," *Journal of Time Series Analysis*, 6, 229–252.

——— (1996), Comment on "Is Seasonal Adjustment a Linear or Nonlinear Data Filtering Process?" by E. Ghysels, C. Granger, and P. Siklos, *Journal of Business & Economic Statistics*, 14, 389–393.

Findley, D. F., and Monsell, B. C. (1986), "New Techniques for Determining if a Time Series Can be Seasonally Adjusted Reliably," in *Regional Economic Modeling*, eds. M. R. Perryman and J. R. Schmidt, Amsterdam: Kluwer-Nijhoff, pp. 195–228.

——— (1989), "Reg-ARIMA Based Preprocessing for Seasonal Adjustment," in *Analysis of Data in Time*, eds. A. C. Singh and P. Whitridge, Ottawa: Statistics Canada, pp. 117–123.

——— (1990), "Standard Errors for Seasonal Adjustment: A Resampling Approach," in *Proceedings of the Business and Economics Statistics Section, American Statistical Association*, pp. 455–460.

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

Findley, D. F., and Parzen, E. (1995), "A Conversation With Hirotugu Akaike," *Statistical Science*, 10, 104–117.

Ghysels, E., Granger, C. W., and Siklos, P. (1996), "Is Seasonal Adjustment a Linear or Nonlinear Data Filtering Process?" *Journal of Business & Economic Statistics*, 14, 374–386.

Gomez, V., and Maravall, A. (1994a), "Program SEATS Signal Extraction in ARIMA Time Series: Instructions for the User," Working Paper ECO 94/28, European University Institute, Florence.

——— (1994b), "Program TRAMO 'Time Series Observations With ARIMA Noise, Missing Observations and Outliers': Instructions for the User," Working Paper ECO 94/31, European University Institute, Florence.

Gray, A., and Thomson, P. (1996), "Design of Moving-Average Trend Filters Using Fidelity and Smoothness Criteria," in *Athens Conference in Applied Probability and Time Series, Vol. II: Time Series Analysis in Memory of E. J. Hannan*, eds. P. M. Robinson and M. Rosenblatt, New York: Springer-Verlag, pp. 205–219.

Hannan, E. J., and Quinn, B. G. (1979), "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society*, Ser. B, 41, 190–195.

Hillmer, S. C., and Tiao, G. C. (1979), "Likelihood Function of Stationary Multiple Autoregressive Moving Average Models," *Journal of the American Statistical Association*, 74, 652–660.

Hosoya, Y., and Taniguchi, M. (1982), "A Central Limit Theorem for Stationary Processes and the Parameter Estimation of Linear Processes," *The Annals of Statistics*, 10, 132–153.

Huot, G. J., Chiu, K., Higginson, J., and Gait, N. (1986), "Analysis of Revisions in the Seasonal Adjustment of Data Using X-11-ARIMA Model-Based Filters," *International Journal of Forecasting*, 2, 217–229.

Hurvich, C. M., and Tsai, C. L. (1989), "Regression and Time Series Model Selection in Small Samples," *Biometrika*, 76, 297–307.

Hylleberg, S. (1986), *Seasonality in Regression*, New York: Academic Press.

Jones, R. H. (1980), "Maximum Likelihood Fitting of ARMA Models to Time Series With Missing Observations," *Technometrics*, 22, 389–395.

Kenny, P. B., and Durbin, J. (1982), "Local Trend Estimation and Seasonal Adjustment of Economic and Social Time Series," *Journal of the Royal Statistical Society*, Ser. A, 145, 1–41.

Kohn, R., and Ansley, C. F. (1986), "Estimation, Prediction, and Interpolation for ARIMA Models With Missing Observations," *Journal of the American Statistical Association*, 81, 751–761.

Koopmans, L. H. (1974), *The Spectral Analysis of Time Series*, New York: Academic Press.

Laniel, N. (1986), "Design Criteria for the 13-Term Henderson End-Weights," Working Paper TSRA-86-011, Statistics Canada, Methodology Branch.

Ledolter, J. (1989), "The Effect of Additive Outliers on the Forecasts From ARIMA Models," *International Journal of Forecasting*, 5, 231–240.

Ljung, G. M. (1993), "On Outlier Detection in Time Series," *Journal of the Royal Statistical Society*, Ser. B, 55, 559–667.

Lothian, J. (1984), "The Identification and Treatment of Moving Seasonality in X-11-ARIMA," technical report, Statistics Canada, Time Series Analysis and Research Division.

Macaulay, F. R. (1931), *The Smoothing of Time Series*, New York: National Bureau of Economic Research.

McNulty, M. S., and Huffman, W. E. (1989), "The Sample Spectrum of Time Series With Trading Day Variation," *Economics Letters*, 31, 367–370.

Musgrave, J. C. (1964), "A Set of End Weights to End all End Weights," working paper, Bureau of the Census, Washington, DC.

Oppenheim, A. V., and Schafer, R. W. (1975), *Digital Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall.

Priestley, M. B. (1981), *Spectral Analysis and Time Series* (vol. 1), London: Academic Press.

Schips, B., and Stier, W. (1995), "The CENSUS-X-11 Procedure: Theory, Assessment and Alternatives," Technical Report 156-0, Swiss Federal Statistical Office, Bern.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464.

Shiskin, J. (1978), "Seasonal Adjustment of Sensitive Indicators," in *Seasonal Analysis of Economic Time Series*, ed. A. Zellner, Washington, DC: U.S. Department of Commerce, Bureau of the Census, pp. 97–103.

Shiskin, J., Young, A. H., and Musgrave, J. C. (1967), "The X-11 Variant of the Census Method II Seasonal Adjustment Program," Technical Paper 15, Bureau of the Census, U.S. Department of Commerce, Washington, DC.

Shulman, H. B., and McKenzie, S. K. (1984), "A Study of Pre-Adjustment Transformations," Research Report RR 84/34, Bureau of the Census, Statistical Research Division, Washington, DC.

Thomson, P., and Ozaki, T. (1992), "Transformation and Seasonal Adjustment," technical report, Victoria University, Wellington, New Zealand. Institute of Statistics and Operations Research.

Vandaele, W. (1983), *Applied Time Series and Box–Jenkins Models*, New York: Academic Press.

Wallis, K. F. (1982), "Seasonal Adjustment and Revision of Current Data: Linear Filters for the X-11 Method," *Journal of the Royal Statistical Society*, Ser. A, 145, 74–85.

Young, A. H. (1965), "Estimating Trading Day Variation in Monthly Economic Time Series," Technical Paper 12, U.S. Department of Commerce, Bureau of the Census, Washington, DC.