Paper 264-25

SAS Programs to Get the Most from X-12-ARIMA's Modeling and Seasonal Adjustment Diagnostics

Catherine C. Hood, U.S. Census Bureau, Washington, DC

ABSTRACT

SAS®'s PROC X11 is based on the Census Bureau's X-11 seasonal adjustment program written in the early 1960's. Future SAS/ETS® Versions will contain PROC X12, based on the Census Bureau's new X-12-ARIMA program, a publicly-available program used worldwide for time series modeling and seasonal adjustment. Major advances in X-12-ARIMA include 1) the ability to use regARIMA models—regression models with ARIMA errors, 2) a new user interface, and 3) new modeling and seasonal adjustment diagnostics, including graphics.

This paper will briefly review some of the basic issues of seasonal adjustment and advances in X-12-ARIMA over X-11. I will focus on the diagnostics available in X-12-ARIMA, particularly the graphical diagnostics including spectral graphs, forecast error history graphs, and revision graphs. The Census Bureau has developed two SAS programs to provide graphical diagnostics and easier access to the diagnostics: X-12-Graph, the companion graphics package to X-12-ARIMA and a SAS interface to X-12-ARIMA. X-12-Graph uses SAS/AF®, SAS/GRAPH®, and SAS Macros and is available for the PC and Unix. The Bureau's SAS interface uses SAS/AF. This paper is intended for anyone interested in time series analysis.

INTRODUCTION

Time Series and Seasonal Adjustment

A *time series* is any sequence of repeated measurements of the same concept over regular, consecutive time intervals, such as every month or every quarter. The data must be comparable over time, so they must be consistent in the concept being measured and the way the concept is measured.

Seasonal adjustment is the process of estimating and removing the seasonal effects from a time series. Seasonal adjustment procedures usually decompose a time series into a trend or trend-cycle, a seasonal component, and the residual or irregular component.

Why would someone want to seasonally adjust a time series? For economic time series, some important features include direction, turning points, and consistency or movement between related indicators. Seasonal movements can make these features difficult or impossible to see. So economists often want seasonally adjusted time series for econometric models.

Other calendar effects that are often estimated are trading day effects and moving holiday effects. When they are persistent and predictable, such calendar-related effects are removed from the original series with the seasonal effects.

A *trading day effect* relates to the number of days of the week in a given month. For example, for a series with higher activity on the weekend such as grocery sales, the number of weekends in the month will have an effect on the series for the month. So we expect a month with five weekends to have more activity than a month with only four weekends.

Inventory or stock series, where the measurement is made on an amount left in inventory on a given day of the month, have a kind of trading day that is different from a *sales or flow series*. Most economic series are flow series.

Another calendar effect is *moving holidays*. Easter is probably the most important moving holiday for seasonal adjustment in the United States. Although we know when Easter will occur every year, it can alternate between March and April. This movement between months makes it different from a fixed holiday, such as Christmas, that can be included in the seasonal effect for the month.

If a time series is a sum (or other composite) of component series that are seasonally adjusted, we can sum the seasonally adjusted component series to get a seasonally adjusted aggregate series. This procedure is called the *indirect adjustment* of the aggregate series. We can also get a direct adjustment of the aggregate series. When the component series have quite distinct seasonal patterns and have adjustments of good quality, indirect seasonal adjustment is usually of better quality than the direct adjustment. For example, US Total Housing Starts is one of the seasonally adjusted series published by the Census Bureau. The US Total is made up of housing starts data from the Northeast, Midwest, South, and West. We seasonally adjust the series at the regional level, because seasonal patterns are different in the different regions of the country. We can estimate the seasonality better by adjusting at the regional level and summing the results to obtain the seasonal adjustment for the U.S.

X-12-ARIMA and Companion SAS Programs

X-12-ARIMA is the Census Bureau's new seasonal adjustment program. X-12-ARIMA is based on the well-known X-11 program (Shiskin, 1967) and Statistics Canada's X-11-ARIMA and X-11-ARIMA/88 (Dagum, 1988). The X-11 family of programs estimates seasonality by using signal-to-noise ratios to choose between a fixed set of moving-average filters, often called X-11-type filters.

The major improvements in X-12-ARIMA fall into three general categories: 1) new modeling capabilities using regARIMA models—regression models with ARIMA errors—for estimating calendar effects or outliers with built-in or user-defined regressors; 2) a new user interface, including new input syntax; and 3) new diagnostics for modeling, model selection, adjustment stability, and for the quality of indirect as well as direct seasonal adjustment. The article by Findley, Monsell, Bell, Otto, and Chen (1998) gives a detailed overview of the improvements. See also U.S. Census Bureau (2000).

X-12-ARIMA uses regARIMA models to preadjust a series before the seasonal adjustment by removing effects such as trading day, moving holidays, and outliers. Often the model is also used to forecast or perhaps backcast the series. Forecast extension allows X-12-ARIMA to use symmetric seasonal and trend filters, and generally results in smaller revisions to the initial seasonal estimates. See Dagum (1988) and Bobbitt and Otto (1990). X-12-ARIMA has automatic procedures for model, outlier, and calendar effect identification. X-12-ARIMA also has built-in regression variables for several different kinds of trading day variables, moving holiday variables, and outliers.

User-defined regression variables can also easily be read in and included in models. This allows regression models for effects not defined in the program, such as Chinese New Year.

The new user interface allows input specification files (also called spec files for short) written in "English." It also has convenient output files, including log files and diagnostics files. Users can customize log files and output files.

Additional capabilities in X-12-ARIMA make it easier to adjust large numbers of series and compute indirect adjustments for aggregate series, a feature inherited from X-11-ARIMA. X-12-ARIMA also has diagnostics for aggregate series to judge the quality of indirect as well as direct seasonal adjustment.

X-12-ARIMA does not have the capability for high resolution graphics, but part of the new user syntax is an option that creates ASCII files to help users import files into a graphics package. To take advantage of this feature, you can run X-12-ARIMA in graphics mode. When you run X-12-ARIMA in graphics mode, it creates graphics files that correspond to the X-12-ARIMA options you choose. X-12-ARIMA also creates a graphics metafile, a list of all the graphics output files produced during the X-12-ARIMA run.

New diagnostics in X-12-ARIMA include forecast error histories to aid in model comparisons (Findley, 1990) and two kinds of stability diagnostics: sliding spans (Findley, Monsell, Shulman, and Pugh, 1990) and revision histories (Findley, Monsell, Bell, Otto, and Chen, 1998). X-12-ARIMA also has some spectral diagnostics for the model residuals and for the seasonal adjustment that were developed from the spectral diagnostics of BAYSEA (Akaike and Ishiguro, 1980).

X-12-ARIMA also has many standard model diagnostics to aid in model selection, such as the Autocorrelation Function (ACF) values and plots, the Partial Autocorrelation Function (PACF) values and plots, and the Ljung-Box Q to test for uncorrelated residuals. X-12-ARIMA also includes the AIC and BIC to aid in model comparisons.

X-12-ARIMA also contains several different seasonal adjustment diagnostics, including the M and Q statistics from X-11-ARIMA (Lothian and Morry, 1972).

In addition to the new features in X-12-ARIMA, the Census Bureau has also written two companion SAS programs for X-12-ARIMA: X-12-Graph to produce graphical diagnostics and a SAS interface to X-12-ARIMA to help users run X-12-ARIMA and manage the output.

Graphs are an important part of seasonal adjustment diagnostics. Since X-12-ARIMA does not have high resolution graphical capabilities, we've developed a companion graphics package in SAS called X-12-Graph. X-12-Graph reads in graphics output files from X-12-ARIMA and gives the users a simple way to get SAS graphs, even if they are not familiar with the SAS System.

X-12-Graph comes in both a batch version and an interactive version. At the heart of both versions are the GPLOT and GREPLAY procedures in SAS/GRAPH, templates, and annotate data sets. The batch program uses SAS Macros and can send selected graphs for any number of series directly to a printer. The interactive program uses SAS/AF and Screen Control Language (SCL). We developed batch and interactive programs for both PC and UNIX platforms since we support X-12-ARIMA on both platforms. More information on X-12-Graph is given in the next section. See also Hood (2000a and 2000b).

Though X-12-ARIMA has several capabilities to help users run a large number of series, running X-12-ARIMA for a large number of series or repeated runs for the same series can be difficult to manage. So we have also developed a SAS interface to X-12-ARIMA that runs X-12-ARIMA and helps manage the various output files. The SAS interface also helps us write the input specification files, and uses X-12-ARIMA's diagnostic and output files to save selected diagnostics in a SAS data set. For a large number of series, the interface helps the user see which series need more attention. For repeated runs on the same series, the interface helps the users see which options give the best results.

The SAS interface uses SAS/AF and SCL, and is written only for the PC. More information on the SAS interface is given later in the paper. See also Hood (2000c).

X-12-GRAPH AND GRAPHICAL DIAGNOSTICS

Graphical diagnostics are very important to seasonal adjustment. Graphs help the user see changes in the seasonal pattern, changes in the level of the series, and outliers. Graphs can help the user look at the behavior of the seasonal factors. Some graphical diagnostics, such as spectral graphs and forecast error history graphs, help us with regARIMA model selection. Other graphical diagnostics help us see the size and location of the revisions in the seasonally adjusted series or trend over time.

Providing specific time series graphs is one way to help people who are using X-12-ARIMA for seasonal adjustment but are not familiar with the SAS System. There were several useful time series graphs in the literature that we wanted to program into X-12-Graph. So far, X-12-Graph contains 14 different types of graphs.

There are also several graphical options included in the program so users can customize the graphs, including the ability to change the types and colors of lines in the graphs, the ability to change the dates in the graphs, and the ability to change titles and footnotes.

Although space does not permit details for every type of graph and every type of option, examples of all the graphs and instructions for producing them in X-12-Graph are available in the User's Guide. Some examples of how we use the diagnostics on a set of real series can be found in the paper by Findley and Hood (1999).

How do we get a good general picture of the time series? We will look at some general purpose graphs: overlay graphs and seasonal factor graphs. How do we choose between regARIMA models? We will look at spectral graphs and forecast error history graphs. How do we judge the quality of the seasonal adjustment? We will look at revision graphs for the seasonally adjusted series.

General Purpose Graphs

Overlay Graphs

Overlay graphs allow the user to select from one to three different graphical elements to plot above a single axis. Examples of the kind of elements in an overlay graph are the original series, the trend, and the seasonally-adjusted series. Overlay graphs are useful for getting a good general feel for the series and the seasonal adjustment.

We will look an example of how overlay graphs can be useful to help find a good data span for modeling or adjustment. Notice in the graph of the original series there are troughs in July that are deeper after approximately 1992.

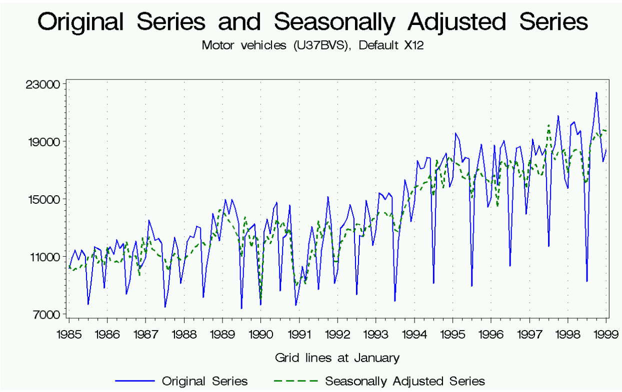

Figure 1. Graph of Original and Seasonally Adjusted Series with a Change in the Seasonal Pattern – Motor Vehicle Value of Shipments

Seasonal Factors by Month or Quarter

Another useful graph is the Seasonal Factor by Month graph, developed by Cleveland and Terpenning (1982). This graph allows you to see the seasonal pattern clearly and to see the stability of the seasonal factors. Each calendar period has a year axis drawn at the level of its factor mean. Notice in the graph below the decreased shipments in July relative to the other months.

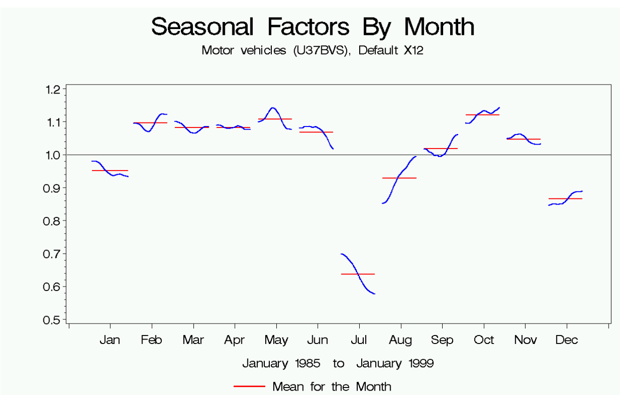

Figure 2. Seasonal Factors by Month – Motor Vehicle Value of Shipments

Graphical Modeling Diagnostics

Spectral Graphs

X-12-ARIMA and X-12-Graph provide an optional spectral diagnostic for the regARIMA model residuals. This is a useful tool to look for problems with the regARIMA model. Seasonal and trading day peaks may indicate a problem with the model. Common remedies include shortening the span of data for the seasonal adjustment or for the modeling, changing the type of trading day variable, and trying different ARIMA models.

Graphs of 10 times the $\log_{10}$ of the spectrum amplitudes are similar to those in the X-12-ARIMA output file. X-12-ARIMA automatically estimates three spectra whenever seasonal adjustment is requested: the spectrum of the differenced original series, the spectrum of the differenced seasonally adjusted series, and the spectrum of the final irregular component. Seasonal frequencies are marked by vertical lines at k/12 cycles per month for $1 \leq k \leq 5$. Cleveland and Devlin (1980) identified the trading day frequencies of this graph as the frequencies most likely to have spectral peaks if a flow series has a trading day component. Trading day frequencies are marked by vertical lines at 0.348 and 0.432 cycles per month.

Visually significant peaks at any of the seasonal or trading day frequencies could indicate a problem with the model. Notice a seasonal spectral peak at 4/12 in the graph below.

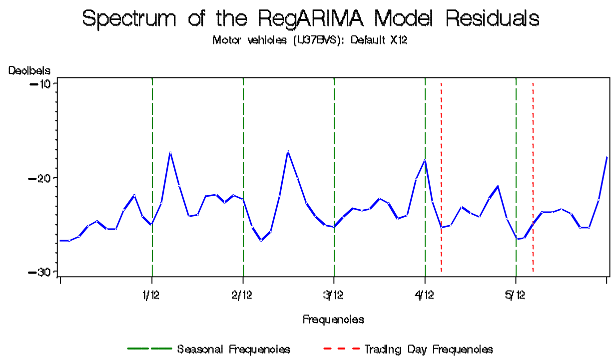

Figure 3. Spectral Graph of the RegARIMA Model Residuals – Motor Vehicle Value of Shipments

Forecast Error History Graphs

The history diagnostics in X-12-ARIMA can compute historical likelihood statistics, such as AIC, and historical forecast errors from the regARIMA model estimation on a sequence of runs from truncated sets of data. Then we can compare the AICs and forecast errors over time.

Since one reason to use regARIMA models is to forecast the series, it is logical to look at forecast performance for competing regARIMA models. The forecast error history diagnostic can be especially helpful when the AIC values are close. We can also use the forecast error history diagnostic to choose between an additive or multiplicative model, again, especially if the AICs are close. The forecast error history diagnostic is also useful in situations where the AIC is not appropriate, for example, when you are looking at series of different lengths. This can happen when you are trying to decide between two ARIMA models with different differencing operators, or when you want to look at possibly shortening the series.

X-12-ARIMA's history spec computes the sums of squared forecast errors at specified forecast leads, or forecast leads of 1 and 12 by default. We then use X-12-Graph to produce graphs of the accumulating differences between the two models. Let X_t be a time series defined for t=1,2,...,T. Let $\hat{X}_{t+h|t}$ be the *h*-step-ahead forecast of X for observation t, for $t_0 \le t \le T - h$. For points within the series, we know the correct value, so we can compute the within sample forecast error as $Y_{t+h} - \hat{Y}_{t+h|t}$ for $t_0 \le t \le T - h$. We can square the errors for both models and take the difference of the squared errors. Then we can sum the differences, and for given *h* and *M*, we plot

$$SSE_{h,N}^{(1,2)} = \sum_{t=M}^{N} [(Y_{t+h} - \hat{Y}_{t+h|t}^{(1)})^2 - (Y_{t+h} - \hat{Y}_{t+h|t}^{(2)})^2]$$

versus N for $M < N \le T - h$.

In the graph below, the direction of the accumulating differences is predominantly downward for Lag 1 and neither upward or downward for Lag 12. Thus, for Lag 1, the forecast errors are persistently smaller for the first model—the regARIMA model that begins with data in 1993. Therefore, we prefer the model that uses data beginning in 1993 instead of the model that uses data for the entire time series.

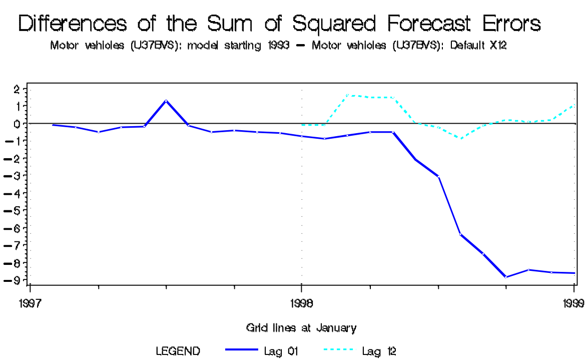

Figure 4. Differences of the Sum of Squared Forecast Errors History Graph – Motor Vehicle Value of Shipments

Graphical Diagnostics for the Quality of the Adjustment

Spectral Graphs

For an adjustment to be acceptable, there should be no residual seasonal or calendar effect present in the seasonally adjusted series or in the irregular component. We use spectral graphs to look for the presence of residual seasonal effects. Visually significant peaks at any of the seasonal or trading day frequencies for either the seasonally adjusted series or the irregular is a signal of possible residual seasonality or trading day effect.

Seasonal Adjustment History Graphs

Another desirable property of the seasonally adjusted series is small revisions. Revisions are the change or percent change from the initial estimate to the final estimate for any given point.

Let X_t be a time series defined for t=1,2,...,T. Let $A_{n|t}$ be the seasonal adjustment of X for observation n calculated using X_1, X_2, . . ., X_t, where $n \le t \le T$. Define $A_{n|n}$ to be the *initial* or concurrent seasonal adjustment—the first seasonal adjustment for observation n. Define $A_{n|T}$ to be the *final* or full-series seasonal adjustment—the seasonal adjustment for observation n including all the data up to observation N.

We then can calculate a mean and maximum absolute percent difference between the initial and final estimates for the seasonal adjustment and between the initial and final estimate of the percent change from the preceding month for the seasonal adjustment. But we can also look at graphs. Graphs of the initial and final estimates for the last years of the series allow us to look at individual months to see how various models and options affect the revisions to the final seasonally adjusted series and the month-to-month change in the series.

We can look at the initial and final seasonally adjusted series for the end of the series. The final estimates are graphed in a line and the initial estimates as dots. Notice some large discrepancies between the estimates for a default run of X-12-ARIMA, particularly in June 1996 and March 1997.

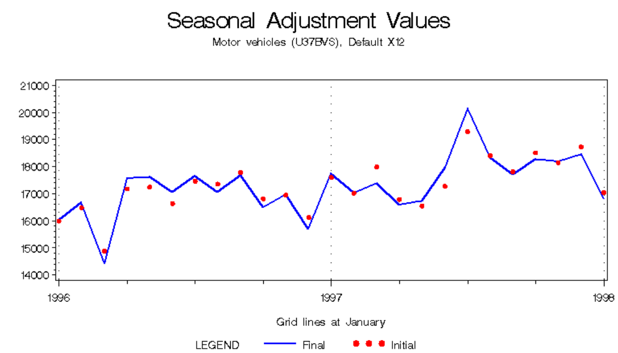

Figure 5. Graph of Initial and Final Estimates of the Seasonally Adjusted Series – Motor Vehicle Value of Shipments, Default X-12

We can also look at graphs of the month-to-month (or quarter-to-quarter) percent changes. We can compare the initial estimate of the percent change to the final percent change. We are looking for differences between the initial and final estimates. It can be especially troublesome when the initial and final estimates have a different sign.

It is useful to see the initial and final percent changes as well as the difference between the two. For example, suppose the difference in the percent changes is 2%. If the percent change for the initial estimate is 3% and that of the final is 1%, then the 2% difference is not very good. But if the percent change for the initial and final estimates are 20% and 18%, then the 2% difference is not dramatic.

A vertical line shows the change between the two estimates. A diamond marks the final estimate. A circle marks the initial estimate.

In the graph below, you will notice a few months with some rather large revisions to the month-to-month percent change.

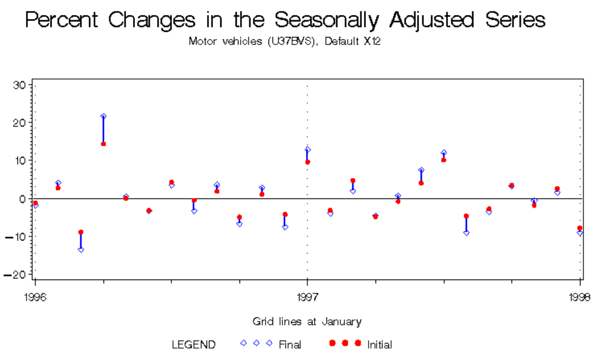

Figure 6. Graph of Initial and Final Estimates of the Seasonally Adjusted Series – Motor Vehicle Value of Shipments, Default X-12

Since the July values for this series have more variation than the other calendar months, it may be useful to use a shorter X-11 filter for July. To see if changing the filters gives us a better adjustment, we can compare the seasonal adjustment history graphs. The differences in the percent changes are smaller when we use a shorter filter for July.

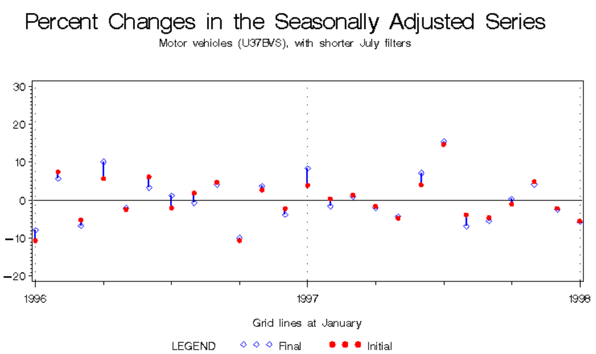

Figure 7. Graph of Initial and Final Estimates of the Seasonally Adjusted Series – Motor Vehicle Value of Shipments with shorter July filters

THE SAS INTERFACE TO X-12-ARIMA

One of X-12-ARIMA's new input options is the ability to use a metafile. This is a list of all the specification files you want to run. Each spec file contains all the commands to specify the adjustment and diagnostic options for a single run for one series. You can input the metafile to X-12-ARIMA and it will run all the series listed. The SAS interface takes advantage of this new capability in X-12-ARIMA.

One of the new output options in X-12-ARIMA is an option to write the seasonal adjustment diagnostics to a file separate from the main output file. This diagnostics file is an ASCII file that lists a code for the diagnostic followed by a colon, and then the value for the diagnostic. It is very easy to read this file into any package and see the information you need. Another new feature in X-12-ARIMA's output is the ability to save certain diagnostics to a log file. The SAS interface takes advantage of both these files.

The SAS interface allows you to select the metafile you want to use, select the series or group of series you want to run, and run X-12-ARIMA with or without graphics mode for the series you have selected.

When X-12-ARIMA finishes, the screen output from DOS appears in an FSLIST window. Once you close this window, you have the option to view a number of different outputs, including the log file, the main output file, a summary from the diagnostics file, or a summary from the main output file. The summary from the main output file comes from code within the interface that parses the main output file and prints out a one page summary.

Once you've seen the output, you then have the option to save some diagnostics from the diagnostics file into a SAS data set. The user can change the diagnostics that the program saves. The user also has the ability to go into an FSVIEW session with the data set where the program saves the diagnostics, select the series with the best diagnostics, and write out the spec file with the options that produced that set of diagnostics.

The SAS interface to X-12-ARIMA also links to X-12-Graph so that users can see the output from X-12-ARIMA and also see the graphical diagnostics available in X-12-Graph.

The SAS interface also has a section that helps you write spec files. It uses the capabilities in X-12-ARIMA to get preliminary regARIMA model options and preliminary seasonal adjustment options and writes the input specifications needed to a spec file. It can do this for every series in the metafile list, making it easy to write preliminary spec files for a large number of series.

PLANS FOR THE FUTURE

The Census Bureau is currently researching seasonal adjustment with ARIMA-model-based signal extraction techniques and how the diagnostics in X-12-ARIMA can be used by model-based programs. The SAS interface has been an important part of the research work since we use it to collect the diagnostic results. In the future, the Bureau might release a version of X-12-ARIMA that allows the use of model-based adjustment.

Changes planned for X-12-Graph include some changes in the code to allow maximum benefit from the new interactive features in SAS/GRAPH Version 8. We are also working on new graphs for the sliding spans diagnostic.

New features planned for the SAS interface to X-12-ARIMA include flexible report-writing capabilities. The current version of the SAS interface is serving as a prototype and starting point for a production system for seasonal adjustment and the annual seasonal review process at the Bureau.

CONCLUSION

With the SAS System, we have been able to develop a set of very useful programs for graphical seasonal adjustment diagnostics. X-12-Graph is easy to use and has a wide range of options for different types of graphs and options for those graphs. The SAS interface makes it easier to look at the wide range of diagnostics available in X-12-ARIMA.

REFERENCES

Akaike, H. and M. Ishiguro (1980), "BAYSEA, a Bayesian Seasonal Adjustment Program," *Computer Science Monographs,* No. 13, The Institute of Statistical Mathematics: Tokyo.

Cleveland, W.S. and I. Terpenning (1982), "Graphical Methods for Seasonal Adjustment," *Journal of the American Statistical Association*, 77: 52-72.

Dagum, E.B. (1988), "X-11-ARIMA/88 Seasonal Adjustment Method - Foundations and Users' Manual," Statistics Canada.

Findley, D.F. (1990), "Making Difficult Model Comparisons Graphically," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, Alexandria, VA, 622-627.

Findley, D.F. and C.C. Hood (1999). "X-12-ARIMA and Its Application to Some Italian Indicator Series," to appear in *Seasonal Adjustment Procedures—Experiences and Perspectives*, Istituto Nazionale di Statistica (ISTAT), Rome, 231-251.

Findley, D.F., B.C. Monsell, W.R. Bell, M.C. Otto and B.-C. Chen (1998), "New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program" (with discussion), *Journal of Business and Economic Statistics*, 16: 127-176.

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

Hood, C.C. (2000a), "X-12-Graph: A SAS/GRAPH® Program for X-12-ARIMA Output, User's Guide for X-12-Graph Interactive for PC/Windows, Version 1.2," U.S. Census Bureau: Washington, DC.

Hood, C.C. (2000b), "X-12-Graph: A SAS/GRAPH® Program for X-12-ARIMA Output, User's Guide for X-12-Graph Batch, Version 1.2," U.S. Census Bureau: Washington, DC.

Hood, C.C. (2000c), "The SAS Interface for X-12-ARIMA, User's Guide, Version 1.0," U.S. Census Bureau: Washington, DC.

Lothian, J. and M. Morry (1978), "A Test of Quality Control Statistics for the X-11-ARIMA Seasonal Adjustment Program," Research Paper, Seasonal Adjustment and Time Series Staff, Statistics Canada.

Shiskin, J., A.H. Young and J.C. Musgrave (1967), "The X-11 Variant of the Census Method II Seasonal Adjustment Program, " Technical Paper No. 15, Bureau of the Census, U.S. Department of Commerce.

U.S. Bureau of the Census (2000), *X-12-ARIMA Reference Manual, Version 1.0*, Washington, DC: U.S. Bureau of the Census.

ACKNOWLEDGMENTS

David Findley and Brian Monsell provided valuable suggestions for the design of the graphs in X-12-Graph, the layout of the program, and the documentation. Thanks also go to Kathleen McDonald-Johnson for her help in testing and design. Many thanks to everyone who gave us suggestions.

This paper reports the general results of research undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau publications. This report is released to inform interested parties of research and to encourage discussion.

CONTACT INFORMATION

For more information on X-12-ARIMA, X-12-Graph, or the SAS interface to X-12-ARIMA, please contact

Catherine Hood
US Census Bureau
Time Series Methods Staff, ESMPD
Washington, DC 20233
catherine.c.hood@ccmail.census.gov
(301) 457-4912

Or visit the X-12-ARIMA website at
www.census.gov/srd/www/x12a
