X-13ARIMA-SEATS and iMetrica

Brian Monsell[1], Chris Blakely[1]
[1] US Census Bureau, Washington, DC

Abstract

This paper will give the latest developments in two ongoing software projects. One is a joint collaboration with the current developers of the SEATS seasonal adjustment program, X-13ARIMA-SEATS, officially released by the Census Bureau in 2012. This program combines the seasonal adjustment and modeling modules of the X-12-ARIMA program with the model–based seasonal adjustment module from SEATS. This program allows producers of seasonally adjusted series to generate X-11 and SEATS seasonal adjustments using the same interface, and compare these seasonal adjustments using a common set of diagnostics. The second is a software system named iMetrica, which is a unique GUI oriented software for both simulating and modeling from many different types of time series models. This software package focuses on speed, user interaction, visualization tools, and point-and-click simplicity for building models for time series data of all types and is written entirely in GNU C and Fortran with a rich interactive interface written in Java. One powerful feature that is unique to the iMetrica software is the capability of easily combining both model-based and non-model based methodologies for designing data forecasts, signal extraction filters, or strategies for comparison of nested or non-nested models. Furthermore, the model comparisons can be computed and tested both in-sample and out-of-sample using a built-in data partitioner that effectively partitions the data into an in-sample storage where models and filters are computed and then an out-of-sample storage where new data is applied to the in-sample strategy to test for robustness and many other desired properties. This gives the user complete liberty in creating a fast and efficient test-bed for implementing signal extractions and forecasting regimes. This paper will demonstrate how SEATS adjustments are integrated into the X-13ARIMA-SEATS procedure and give examples of new diagnostics and modeling options, including some designed for use in modeling the recent recession, as well as a description of the five interacting time series analysis modules that comprise iMetrica. These include modules for X-13ARIMA-SEATS, unobserved component modeling, and state space modeling.

Keywords: Signal extraction, likelihood statistics, empirical mode decompositions, model comparison.

Disclaimer

This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

1. Introduction

This paper will give the latest developments in two ongoing software projects conducted by the Time Series Research Staff of the U. S. Census Bureau. The first is the X-13ARIMA-SEATS (X-13A-S) program, used to generate official seasonal adjustments since its release in March of 2011. Monsell (2007, 2009) gave a description of an ongoing collaboration between the U. S. Census Bureau and the developers of the SEATS seasonal adjustment procedure to produce a program that would allow users to generate X-11 and SEATS seasonal adjustments using the same interface and to compare these seasonal adjustments using a common set of diagnostics. For more details on the SEATS seasonal adjustment method, see Maravall (1995), Gomez and Maravall (2000), and Gomez and Maravall (2001). This paper will document new features and developments for this software. The second is the iMetrica software project, which began in 2010 as a simple graphical user interface for the X-13ARIMA-SEATS Fortran program. One of the main motivations was to allow easy access to performing the most used X-13A-S features with the simple click of a mouse or keyboard button, and being able to visualize the results instantaneously. This initial program became uSimX13 (see Blakely 2013) and is featured as one of the four model-based time series analysis modules of the iMetrica software package.

2. X-13ARIMA-SEATS

X-13ARIMA-SEATS is a collaboration between the developers of X-12-ARIMA and SEATS and is a merged version of the two programs. Developed to assist in evaluating the effectiveness of model-based seasonal adjustment, it allows for the seamless integration of these adjustments into production processes that currently are designed to use only X-12-ARIMA, the previous version of the Census Bureau's seasonal adjustment software. Version 1.1 of X-13ARIMA-SEATS, to be released by the summer of 2013 (download at http://www.census.gov/srd/www/x13as/), will have updated SEATS source code from the Bank of Spain with several important revisions. In addition, there will be several new features, described briefly below.

2.1 QS diagnostic

An important diagnostic for determining if there is seasonality present in a given time series is the spectrum diagnostic. The X-13ARIMA-SEATS program provides two versions of this diagnostic, and the program flags seasonal peaks in several series components of interest. Starting with Version 1.0, these plots are provided only for monthly series; a diagnostic to detect residual seasonality in time series of other periodicity, including quarterly, is the QS diagnostic from the TRAMO and SEATS programs.

The QS test statistic checks for evidence of seasonality in the autocorrelation function of relevant series associated with a seasonal adjustment: original series, the seasonally adjusted series, the irregular component, (and each adjusted for extreme values in the case of an X-11 decomposition), and the regARIMA residuals. If the series is not monthly or quarterly, then the QS diagnostic is only produced for the regARIMA residuals. Two sets of these diagnostics are generated: a first set for the full span of the series and a second set for the span of data starting at the date given by the **start** argument of the **spectrum** spec (if it is different from the span of the series).

First, the program chooses the order of differencing (*ndif*) to be used on the series before the application of the QS test. For the irregular component, this is set to 0. For other series, if there is an ARIMA model specified for the series, *ndif* is initially set as $ndif = max(1,min(d+D,2))$, where d is the order of nonseasonal differencing, and D is the order of seasonal differencing. If there is no ARIMA model specified for the series, *ndif* is initially set to 1.

Let s be the seasonal period (12 for monthly data, 4 for quarterly data, for example). Autocorrelations are generated from the *ndif* times differenced mean corrected series. If $s > 4$, the program checks if the autocorrelations at lags 1, 2, 3, 4 and s are positive. If $s \leq 4$, the program checks if the autocorrelations from lags 1 to s are positive. If either of these conditions is satisfied and ndif=1, the autocorrelations will be generated from the twice-differenced mean corre cted series and *ndif* will be reset to 2. Otherwise, the autocorrelations of the *ndif* differenced mean corrected series are used to generate the QS statistic.

Let r_i be the *ith* autocorrelation of the differenced, mean corrected series, and *nz* be the length of the series. Set $n = nz - ndif$ and

$$R_i = \begin{cases} r_i \text{ if } r_i > 0 \\ 0 \text{ if } r_i \leq 0 \end{cases} .$$

If $R_s > 0$, then $QS = n(n+2) * \left({R_s^2}\Big/{(n-s)} + {R_{2*s}^2}\Big/{(n-(2*s))} \right)$; otherwise, $QS = 0$. The statistic is assumed to be distributed as a chi-squared with two degrees of freedom. Sample output for this diagnostic can be found in Census (2013).

2.2 New Regression Variables

Version 1.1 of X-13ARIMA-SEATS contains a set of new regressors added to the software in response to modeling issues related to recession effects that were evident in U. S. time series starting in late 2008. More information on modeling issues related to the 2008 recession can be found in Lytras and Bell (2013).

The first set of new regressors are sequence outlier regressors – a sequence of additive outlier (AO) or level shift (LS) regressors that can be specified from a starting and ending date. These regressors can be used to form complex models for economic activity that change the behavior of the series over a span of time, and allow the user a compact method of specifying a series of outlier regressors over a period of time. Specifying `variables = AOS2008.jul-2008.nov` in the **regression** spec will add AO outliers for each observation in the span of July 2008 to November 2008, inclusive. In addition, the **tlimit** argument of the regression spec allows the user to specify a critical value for the AO and LS sequence regressors. T-statistics are generated for each AO or LS regressors specified for a given sequence, and only those regressors with t-values above the value of **tlimit** will be kept in the model.

Another set of new regressors are **quadratic ramp variables**. These are ramp variables that change quadratically rather than linearly, since in some instances the change in the level of the series may not be modeled well with the standard ramp variable. There are two versions of this variable – increasing quadratic ramps (where the rate of change is increasing), and decreasing quadratic ramps (where the rate of change is decreasing).

2.3 Optional critical values for AIC tests

Version 1.1 of X-13ARIMA-SEATS adds another option to the **regression** spec (**pvaictest**) that allows users to specify an optional critical value for the **aictest** variable to convert AICC comparisons to standard likelihood ratio test. Simulation studies comparing the use of this critical value to simply picking the model with the lowest AICC value show that using the critical value reduces Type 1 testing errors in AIC testing, particularly for those regressors with multiple non-nested alternatives (Easter for example).

Let ΔL be the difference between the maximum Gaussian log-likelihood values for an unconstrained stationary time series model and for a model nested within it having ν fewer independent parameters. Taniguchi and Kakizawa (2000) show that $2\Delta L$ is distributed asymptotically as a chi-square with ν degrees of freedom when the smaller model is correct. Using this, we can construct critical values for the AICC differences by choosing CV(α,ν) such that Pr(X2(ν) > 2ν + CV(α,ν)) = α. For example, if α = 0.05 and ν=1, then CV(α,ν) = 3.84146 – 2 = 1.84146. A table of these values is given below.

ν	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0.95	1.8414	1.9915	1.8147	1.4977	1.0705	0.5915	0.0671
0.99	4.6394	5.2103	5.3449	5.2767	5.0863	4.8119	4.4753

Table 1 : Critical values CV(α,ν) for two values of α

3. iMetrica

The iMetrica software package is a unique system of econometric and data analysis tools that focuses on speed, user interaction, visualization tools, and point-and-click simplicity in building models for time series data of all types. Written entirely in C and Fortran with a rich interactive interface written in Java, the iMetrica software offers an abundance of econometric tools for signal extraction and forecasting in multivariate time series. These tools are both easily accessible with the click of a mouse button and fast with results computed and plotted instantaneously without the need for creating output data files or calling exterior plotting devices.

One feature that is unique to iMetrica is the capability of easily combining model-based and non-model based methodologies for designing data forecasts, signal extraction filters, or seasonal adjustment strategies. These can be computed and tested both in-sample and out-of-sample, using a built-in data partitioner that separates the data into an in-sample storage where models and filters are computed and an out-of-sample storage where new data is applied to the in-sample strategy to test for robustness and other desired properties. This gives the user complete liberty in creating a fast and efficient test-bed for implementing signal extractions, forecasting regimes, or financial trading strategies. The iMetrica software environment includes five interacting modules for building hybrid forecasts, signal extraction strategies.

uSimX13 – A computational environment for univariate seasonal auto-regressive integrated moving-average (SARIMA) modeling and simulation using X-13ARIMA-SEATS, described in Section 1. This module features an interactive approach to modeling seasonal economic time series with SARIMA models and automatic outlier detection, trading day, and holiday regressor effects and includes a suite of model comparison tools using a number of signal extraction diagnostics.

BayesCronos – An interactive time series module that performs signal extraction and forecasting of multivariate economic and financial time series and focuses on Bayesian computation and simulation. This module includes a multitude of model types, including ARIMA, GARCH, EGARCH, Stochastic Volatility, Multivariate Factor Stochastic Volatility, Dynamic Factor, and Multivariate High-Frequency-Based Volatility (HEAVY). For most of the models featured, one can compute the Bayesian and/or the Quasi-Maximum-Likelihood (QML) estimated model fits either using a Metropolis-Hastings Monte Carlo Markov Chain approach (Bayesian) or a QMLE formulation for computing the model parameters estimates. Using a model selection panel, complete access to model-type, model parameter dimensions, and prior distribution parameters are available. In the case of Bayesian estimation, the software offers complete control over the prior distributions of the model parameters along with interactive visualization of the Monte Carlo Markov Chain parameter samples. For each model, up to 10 sample forecasts up to 36-steps ahead can be produced and visualized along with other model features such as model residuals, computed volatility, forecasted volatility, factor models, and more. Results from this module can be exported to other iMetrica modules for additional filtering and/or modeling.

MDFA - A comprehensive multivariate real-time direct filter analysis and computation environment, which can build real-time filters using both I-MDFA and Zero-Pole Combination (ZPC), filter constructions. The module includes interactive access to timeliness, smoothing, and accuracy controls for filter customization along with parameters for filter regularization to control over fitting. More advanced features include an interface for building adaptive filters, and controls for filter optimization, customization, data forecasting, and target filter construction.

State Space Modeling – A module for building observed component ARIMA and regression models for univariate economic time series. Similar to the uSimX13 module, the State Space Modeling environment focuses on modeling and forecasting economic time series data, but with much more generality than SARIMA models (see Blakely 2013b). An aggregation of observed stochastic components in the form of ARIMA models are stipulated for the data (for example trend + seasonal + irregular) and then regression components to model outliers, holiday, and trading day effects are added to the stochastic components, giving flexibility in model building. The module uses regCMPNT, a suite of Fortran code written at the US Census Bureau, for the maximum likelihood and Kalman filter computational routines (see Bell 2011).

EMD - An EMD module (for empirical mode decomposition) that offers a time-frequency decomposition environment for the time-frequency analysis of time series data. The module offers both the original empirical mode decomposition technique of Huang et al. (1998) using cubic splines, along with an adaptive approach using reproducing kernels and direct-filtering. This empirical decomposition technique decomposes nonlinear and nonstationary time series into amplitude modulated and frequency modulated (AM-FM) components and then computes the intrinsic phase and instantaneous frequency components from the FM components. All plots of the components as well as the time-frequency heat maps are generated instantaneously.

Along with these modules, there is also a data control module that handles all aspects of time series data input and export. Within this main data control hub, one can import time series data from a multitude of file formats, as well as download them directly from an online time series database such as the World Bank, the Federal Reserve, or Yahoo! Finance, or a source such as Reuters for higher-frequency data. Once the dataset is loaded, the data can be normalized, scaled, demeaned, and/or log-transformed through the use of simple slider and button controls, with the effects being plotted on a graphic canvas.

Another feature of the iMetrica software is the ability to learn more about time series modeling through the using of data simulators. The data control module includes an array of panels for simulating data from both univariate and multivariate time series models. By specification of the number of observations, the random seed for the innovation process, the innovation process distribution, and the model parameters, simulated data can be constructed for any type of economic or financial time series, including (S)ARIMA models, GARCH models, correlated cycle models, trend models, multivariate factor stochastic volatility models, and HEAVY models. From simulating data and varying the parameters, one can visualize the effects of the each parameter on the simulated data. This data can be exported to any of the modules for practicing one’s skills in hybrid modeling, signal extraction, and forecasting.

4. Future Work

After the release of Version 1.1 of X-13ARIMA-SEATS, the automatic ARIMA model identification module within the software will be revised, after an evaluation comparing it to the model identification routine in TRAMO+.

References

Bell, W. (2011) “REGCMPNT: A Fortran Program for Regression Models with ARIMA Component Errors.” Journal of Statistical Software, Vol 41, Issue 7.

Blakely, C. (2013) "iMetrica-uSimX13: A Graphical User-Interfaced Time Series Modeling and Simulation Environment Featuring X-13ARIMA-SEATS", Center for Statistical Research & Methodology, Research and Methodology Directorate Research Report Series, U. S. Census Bureau

Blakely, C, (2013b) "iMetrica-SSM: A State Space Modeling Graphical User-Interfaced Environment Featuring regComponent", Center for Statistical Research & Methodology, Research and Methodology Directorate Research Report Series, U. S. Census Bureau

Census (2013), “Reference Manual for the X-13ARIMA-SEATS Program, Version 1.0”, U. S. Census Bureau, U. S. Departement of Commerce. Available: http://www.census.gov/ts/x13as/docX13AS.pdf

Gomez, V. and Maravall, A. (2001a), “Automatic Modeling Methods for Univariate Series,” in A Course in Time Series Analysis, eds. Pena, D., Tiao, G. C., and Tsay, R. S., New York, NY: J. Wiley and Sons.

Gomez, V. and Maravall, A. (2001b), “Seasonal Adjustment and Signal Extraction in Economic Time Series,” in A Course in Time Series Analysis, eds. Pena, D., Tiao, G. C., and Tsay, R. S., New York, NY: J. Wiley and Sons.

Huang, N. E.; Shen, Z.; Long, S. R.; Wu, M. C.; Shih, H. H.; Zheng, Q.; Yen, N. C.; Tung, C. C.; Liu, H. H. (1998). "The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis." Proceedings of the Royal Society of London A 454 (1971): 903–995.

Lytras, D. and Bell, W. (2013) “Modeling Recession Effects and the Consequences on Seasonal Adjustment”, Proceedings of the 2013 Joint Statistical Meetings, Business and Economics Section.

Maravall, A. (1995), “Unobserved Components in Economic Time Series,” in The Handbook of Applied Econometrics, eds. Pesaran, M. H. and Wickens, M., Basil Blackwell.

Monsell, B. (2007), “The X-13A-S Seasonal Adjustment Program,” Proceedings of the 2007 Federal Committee On Statistical Methodology Research Conference, [Online]. Available: http://www.fcsm.gov/07papers/Monsell.II-B.pdf [2009, February 2].

Monsell, B. (2009) “Update on the Development of X-13ARIMA-SEATS”, Proceedings of the 2009 Joint Statistical Meetings, Business and Economics Section.

Taniguchi, M. and Y. Kakizawa (2000). Asymptotic Theory of Statistical Inference for Time Series. New York: Springer-Verlag.