

Comparing Automatic Modeling Procedures of TRAMO and X-12-ARIMA, an Update

Kathleen M. McDonald-Johnson¹, Catherine C. Harvill Hood², Brian C. Monsell¹, Chak Li¹
U. S. Census Bureau¹
Catherine Hood Consulting²

Abstract

The U. S. Census Bureau's enhanced X-12-ARIMA seasonal adjustment program includes the automatic ARIMA (Autoregressive Integrated Moving Average) model selection procedure developed by Statistics Canada and a second procedure based on the automatic procedure of TRAMO (Time series Regression with ARIMA noise, Missing observations and Outliers), a modeling package developed by Victor Gómez and Agustín Maravall. Each program has automatic identification of key regressors, allowing for full automatic selection of a regARIMA model (regression with an underlying ARIMA process). X-12-ARIMA's procedure differs from TRAMO's in a number of ways. Our study updates previous work as we compared the procedures again using improved versions of the two programs. We applied the procedures to a set of Census Bureau time series and simulations. When model choices differed, we compared standard modeling diagnostics to look for a consistent preference for either procedure. As in the previous study, we found that X-12-ARIMA still seems to choose trading day effects more appropriately than TRAMO. However, we found that X-12-ARIMA inaccurately identifies Easter effects more often than TRAMO. Overall, we found that the diagnostics for the X-12-ARIMA models were at least as good as the diagnostics from TRAMO models.

Keywords: Time series, Seasonal adjustment, Trading day effect

1. Background

X-12-ARIMA Version 0.3 is the latest version in the X-11 line of seasonal adjustment programs. Although this version of the program does not involve ARIMA-model based seasonal adjustment, ARIMA models are used to extend the series with forecasts before applying the usual X-11 seasonal moving averages. The program includes two automatic ARIMA modeling procedures (U. S. Census Bureau 2007). The program retains the model-comparison method implemented in X-11-ARIMA by Statistics Canada (Dagum 1988) and adds a second method based on the procedure found in TRAMO, the companion automatic modeling program to SEATS

(Signal Extraction in ARIMA Time Series), an ARIMA-model based seasonal adjustment program (Gómez and Maravall 1997).

The ARIMA model fit in X-12-ARIMA, written in shorthand as (p d q)(P D Q), follows the form

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D z_t = \theta(B)\Theta(B^s)a_t$$

where z_t is the original time series or possibly a transformation of the original data, t indexes time, B is the backshift operator such that $B^k z_t = z_{t-k}$, s is the seasonal period (12 for monthly series and four for quarterly series), $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ is the nonseasonal autoregressive (AR) operator of order p , $\Phi(B^s) = (1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps})$ is the seasonal AR operator of order P , $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ is the nonseasonal moving average (MA) operator of order q , $\Theta(B^s) = (1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs})$ is the seasonal MA operator of order Q , a_t is a white noise series, that is, independent and identically distributed with mean zero and variance σ^2 , and $(1-B)^d(1-B^s)^D$ indicates nonseasonal differencing of order d and seasonal differencing of order D (U. S. Census Bureau 2007).

The new method of X-12-ARIMA has been under development for several years. Recent work by Dent, Hood, McDonald-Johnson, and Feldpausch (2005) compared the new method to the model-comparison method. (The paper referred to Statistics Canada's model-comparison method as 0.2 and the newer method as 0.3 although both methods are available in X-12-ARIMA Version 0.3.) Their work found that those methods produced models of similar quality, although the new method is more flexible when modeling data that may not be seasonal. The TRAMO method has this flexibility as well. The new method of X-12-ARIMA Version 0.3 and TRAMO have been compared as well; this work updates the study by Farooque, Hood, and Findley (2001) by using the same modeling approach but using updated and improved versions of the programs. Their work found that X-12-ARIMA was perhaps better at identifying trading day effects than TRAMO. Both X-12-ARIMA and TRAMO have undergone many updates since 2001, and the programs have added features and changes to the

previous methodology. Also, because of its new release status, users may want to know how well X-12-ARIMA's new procedure compares to that of TRAMO.

Besides the automatic ARIMA modeling methods, TRAMO and X-12-ARIMA have options that automatically can choose a full regARIMA model. The programs can determine (1) whether or not to use a log transformation, (2) whether trading day effects and Easter effects should be part of the model, and (3) what outliers are significant.

The programs use similar approaches to determine whether a log transformation is appropriate, basing the decision on likelihood statistics after fitting a default ARIMA model. By default, X-12-ARIMA has a slight bias toward the log transformation.

The usual trading day effect, available in both programs, has six parameters so that each day of the week may have a different effect. The full trading day effect is constrained so that the estimated effect for the seventh day is determined by the other days' estimates. There is a difference in the Easter effects available in the programs. The default X-12-ARIMA test for Easter effects checks for effects that affect one, eight, and fifteen days before Easter. The default settings in TRAMO check for an Easter effect of six days.

The approach to outlier detection also differs in the two programs. For each possible outlier, the programs calculate a t statistic and then compare that t statistic to a critical value determined by the series length. TRAMO and X-12-ARIMA have different critical values for this identification. Both programs can test for additive outliers (also known as point outliers), level shifts (abrupt changes in the series level that continue over time), and temporary changes (abrupt outlier effects that decay back to the original series level). During subsequent runs, users can hard-code outliers that the programs identify from their tests. Each outlier is represented by a specific regression variable name that indicates the type of outlier and the date when the outlier occurred.

Another feature of TRAMO is that it prefers balanced models during the model identification process. A balanced model is one whose AR (autoregressive) order plus the order of differencing is equal to the MA (moving average) order, that is, $p + d = q$ and $P + D = Q$. For example, the nonseasonal models (0 1 1) and (1 0 1) are balanced; (0 1 3) and (1 1 0) are not balanced. By default, X-12-ARIMA does not favor balanced models, but the user can change this setting.

Gómez and Maravall (2000) describe the TRAMO modeling procedure. Additional details are available from reading the FORTRAN code which Gómez and Maravall generously provided to the U. S. Census Bureau for use in developing the new automatic modeling procedure for X-12-ARIMA. Some of the most notable differences between the two modeling procedures, as mentioned by Findley (2005) are the criteria for using the log transformation, the outlier regressor critical values, the default models for trading day and Easter effects, and the criteria for determining when those effects should be included in the model.

In our comparisons we used two standard regARIMA-modeling diagnostics: (1) Ljung-Box Q statistics and (2) spectrum of the model residuals.

The Ljung-Box Q statistics are goodness-of-fit diagnostics based on the sample autocorrelations of the residuals (Ljung and Box 1978). A model that fits well should have residuals that behave like white noise. The Ljung-Box Q statistic is a measure of the significance of the lags of the autocorrelation function. For each Ljung-Box Q statistic of positive degrees of freedom, there is a corresponding p value. A lag is said to fail if the p value for the Q statistic for that lag is less than 0.05. Based on professional judgment, we decided that (a) if seven or more of the first 12 lags failed, or (b) if 13 or more of the first 24 lags failed, or © if lag 12 failed, then the model was not a good fit.

The spectrum diagnostic indicates if there are remaining seasonal or trading day effects in the model residuals (Cleveland and Devlin 1980). Figure 1 shows an example spectrum graph. The graph marks visually significant peaks at seasonal (with "S") and trading day frequencies (with "T") (Soukup and Findley 1999). For a peak to be significant, it must reach a height beyond the median height of all the frequency measures, and it must be taller than its nearest neighbors by a visually significant amount. In addition to marking significant peaks, the graph indicates the median level and the calculated visual significance height so users can evaluate the peaks relative to these measures. The example indicates one visually significant seasonal peak at three cycles per year (occurring every four months) and one visually significant trading day peak at the frequency between four and five cycles per year. These peaks mean that there are remaining seasonal and trading day effects in the model residuals, so this particular model has failed the spectrum diagnostic. Spectrum failures are shown on the screen when running X-12-ARIMA and users can save the spectrum information to the log or diagnostics file.

2. Program Information

The version of TRAMO that we used for this work is from March 2006 and is available from the Bank of Spain Internet site, www.bde.es/servicio/software/econome.htm. We used X-12-ARIMA Version 0.3 Build 174, compiled in February 2007. These were the most recent versions of the programs available.

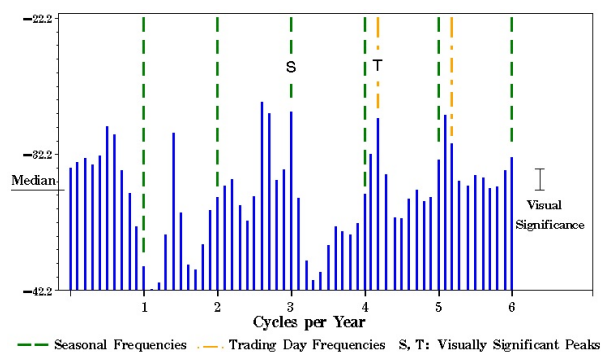


Figure 1: Example Spectrum of RegARIMA Model Residuals

3. Methods, Census Bureau Series

For our comparison of TRAMO and X-12-ARIMA, we started with 457 U. S. Census Bureau series, including U. S. Building Permits, Manufacturing, Retail Sales, and Import/Export data. For information about data collection methods and reliability of the estimates, see the Economic Indicators page on the Internet at www.census.gov/cgi-bin/briefroom/BriefRm. Program overviews and current data are available from links on that page.

For TRAMO, we had the program test for the need for a log transformation and perform automatic regARIMA model identification and outlier detection. The program tested for the usual trading day, leap year, and Easter effects.

We wanted the two procedures to choose models from a basic level that would not provide an unfair advantage to either, so for X-12-ARIMA's automatic modeling, we selected options that would be similar to what we chose for TRAMO. In addition, from our personal experience, we expected that some of our series would have quarterly effects, so we chose the maximum nonseasonal model order to be three (maximum for p and q) instead of the default maximum order of two. We also asked X-12-ARIMA to prefer balanced models to have an approach more like the TRAMO procedure.

If we had compared diagnostics from models estimated with TRAMO to diagnostics from models estimated with X-12-ARIMA, the differences could be indications of a difference in the programs' estimation methods rather than the adequacy of the models. Because we wanted to use model diagnostics available from X-12-ARIMA, we hard-coded the regARIMA model choices identified with TRAMO and X-12-ARIMA into input specification files for X-12-ARIMA. We then ran X-12-ARIMA, estimating those models, and we compared the resulting diagnostics. To clarify, we ran TRAMO to identify a regARIMA model for each series, but what we call the TRAMO model in our comparisons is the result of setting the X-12-ARIMA options to match the regARIMA model choice from our initial TRAMO run.

We did not want choices of outliers or Easter length to be a deciding factor for any of the models. To avoid these problems, we used the X-12-ARIMA outlier set for each series, and for each model that included an Easter effect of any length, we set the Easter effect length to eight days. After the initial outlier identification, we raised the outlier critical value to 5.5 to make it less likely to identify additional outliers. We then refit the models with these changes to the outliers and Easter effects. During the refit, even with the higher outlier critical value, X-12-ARIMA identified an additional two outliers for one TRAMO model (for this series, the original TRAMO outlier set had been larger than the X-12-ARIMA outlier set). For both the TRAMO and X-12-ARIMA models, we changed the regression to adjust for these additional two outliers and estimated those models again.

4. Results, Census Bureau Series

The programs agreed on transformation choice for 91% of the series. Of the 40 series where transformation choice differed, TRAMO chose a log transformation when X-12-ARIMA chose no transformation for 85% (34 series), and for the other 15% (six series), TRAMO chose no transformation when X-12-ARIMA chose a log transformation. The choice of transformation is fundamental to modeling, and we did not want to compare the models from data with the log transformation to those with no transformation. Not wanting to favor one program's transformation over the other, we dropped these 40 series from further analysis, leaving us with 417 series.

Of those 417 series, 30% (124) of the regARIMA models matched. As we describe below, we did not try to evaluate the length of the chosen Easter effect. If the two methods chose an Easter effect of any length, we considered those to be a match. We were left with 293 nonmatching models to compare.

An additional 24% (70 out of 293) of the ARIMA models matched, showing differences only in the chosen regression effects.

Interestingly, 9% (26) of the 293 series showed a difference in seasonal differencing. For 8% (22), TRAMO chose no seasonal difference but X-12-ARIMA did include a seasonal difference. A series that does not require a seasonal differencing is unlikely to have a seasonal effect that is stable enough for reliable seasonal adjustment, so seeing this kind of model switch could affect seasonal adjustment decisions.

For 13% (37) of the series TRAMO and X-12-ARIMA both chose an Easter effect. There were another 24% (71) of the series for which only one of the modeling procedures chose an Easter effect. X-12-ARIMA chose Easter when TRAMO did not for 24% (70 series). No Easter effect was chosen for the remaining 63% (185). We do not have an additional general check for whether Easter was an appropriate regression effect. TRAMO checks for an Easter effect of one length, but the default test in X-12-ARIMA checks for three different potential regressors, so perhaps having the additional tests is why X-12-ARIMA chooses an Easter effect more often than TRAMO. It is hard to evaluate how appropriate the Easter effect is for those additional 71 series. These economic series could indeed have Easter effects, but these results show Easter effects to be more prevalent than we would have expected.

The choice of a trading day effect is somewhat easier to compare because of the spectrum diagnostic. TRAMO and X-12-ARIMA each chose trading day for 24% (70) of the series, and neither chose trading day for 33% (96) of the series.

It can be difficult to make judgments of the appropriateness of a trading day effect. We did not have a general way to evaluate whether choosing to include a trading day effect was incorrect. However, we decided that for a specific circumstance, we could evaluate whether including a trading day effect was correct. If one procedure did not select the trading day regression and the spectrum of those model residuals showed a peak at either of the trading day frequencies, and the other procedure did select a trading day effect and showed no trading day spectral peak, then we considered the trading day effect to be appropriate and the omission to be incorrect. We saw this situation for 22% (64) of the series. This choice was more problematic for TRAMO: 20% (60 series). X-12-ARIMA's choice was problematic for 1% (4 series). Using a binomial distribution, we calculated the probability of seeing 60 out of 64 failures for one method

if the probability of a failure were equally 0.5 for each method. The probability is less than 0.01.

In checking the Ljung-Box Q results, we saw that 24% (69) of the series failed our criteria (listed in Background) for one of the methods while the other method passed. The failures happened more often for the TRAMO models (17%, 50 series) than for the X-12-ARIMA models (6%, 19 series). Again, we calculated the binomial probability that 50 of 69 failures would be from one method. The probability is less than 0.01.

We went on to check the seasonal spectrum results. Of the 293 series, 14% (41) had seasonal spectral failures for one of the methods (passing for the other method). The problem occurred for the TRAMO model 8% of the time (24 series) and for the X-12-ARIMA model 6% of the time (17 series). The binomial probability of 24 of 41 failures being from one method is 0.17, not significant at a 10% level, so there was not a true difference in the seasonal spectrum results.

Combining the results of the Ljung-Box Q and seasonal spectrum diagnostics, we saw that 30% (87) of the series had models that passed for one modeling procedure and failed for the other. Overall, the TRAMO model failed 21% of the time (61 series) and the X-12-ARIMA model failed 9% of the time (26 series). The probability of that result for the 87 failures is less than 0.01.

Tables 1 and 2 summarize our comparison results for the actual Census Bureau series.

5. Methods, Simulated Series

For our comparison of the two modeling procedures results with simulated data, we simulated series that followed the airline-model process, $(0\ 1\ 1)(0\ 1\ 1)$ (Box and Jenkins 1976). The series were simulated as additive processes; that is, they would not require a log transformation before modeling or seasonal adjustment. We looked at 3,500 monthly series, 15 years long, with nonseasonal moving average coefficient (θ) set at 0.6 and seasonal moving average coefficient (Θ) set at 0.9. Those coefficients are representative of typical economic time series. Arbitrarily we set the start date for the series to be January 1980. The series had no trading day, Easter, or intentional outlier effects. We ran the automatic modeling procedures of TRAMO and X-12-ARIMA and evaluated how often the programs chose the airline model. We ran X-12-ARIMA using the default settings (maximum nonseasonal order two and no preference for balanced models) and also with the settings we had used for the Census Bureau data (maximum nonseasonal order three and a preference for balanced models).

Table 1: Comparison of TRAMO and X-12-ARIMA Model Agreement, U. S. Census Bureau Series
(Because of rounding, not all cells sum to totals.)

	Combined	TRAMO	X-12-ARIMA
Same RegARIMA Model	30%		
Same ARIMA Model but Different Regressors	24%		
Easter Effect			
Agreement No Easter Effect	63%		
Agreement Yes Easter Effect	13%		
Disagreement (and Which Chose the Effect)	24%	0.3%	24%
Trading Day Effect			
Agreement No Trading Day Effect	33%		
Agreement Yes Trading Day Effect	24%		
Disagreement (and Which Chose the Effect)	43%	8%	35%

Table 2: Diagnostic Comparison of Models, Where One Model Passed and the Other Failed,
U. S. Census Bureau Series

(Because of rounding, not all cells sum to totals.)

	Combined	TRAMO	X-12-ARIMA
Ljung-Box Q Failure	24%	17%	6%
Seasonal Spectrum Peaks	14%	8%	6%
Either Ljung-Box Q or Seasonal Spectrum Failure	30%	21%	9%
Problematic Trading Day Omission	22%	20%	1%

6. Results, Simulated Series

For 0.6% (21) of the 3,500 series, X-12-ARIMA could not identify a model (same outcome for both option sets). The model did not converge during the identification process. TRAMO selected a model for all 3,500 series.

Of the 3,479 series for which we had model choices for both X-12-ARIMA and TRAMO, 24% (849) had negative or zero values, so the log transformation was not possible. Both methods correctly chose no data transformation for 66% (2,289); both incorrectly agreed on a log transformation for 3% (120). The two sets of X-12-ARIMA options agreed on transformation for all series. For the 221 series where TRAMO and X-12-ARIMA disagreed, TRAMO correctly chose no transformation and X-12-ARIMA incorrectly chose a log transformation for 74% (164) of the series, and X-12-ARIMA correctly chose no log transformation for 26% (57) of the series. As before, using a binomial approach to compare the 221 series, the probability is less than 0.01 that there would be such a large difference by chance. Our X-12-ARIMA settings were slightly biased toward log transformations, so we were not concerned by this result.

If we eliminate the 221 series of disagreement from the 3,479 series that had model choices, there are 3,258 series for which TRAMO and X-12-ARIMA agreed on transformation choice. Of those 3,258 series, TRAMO correctly chose the airline model with no constant term or trading day or Easter effects for 66% (2,152) of the series. X-12-ARIMA run with default settings chose the airline model with no regressors for 73% (2,363) of the series, and when run with the modified settings X-12-ARIMA chose the correct model with no regressors for 72% (2,351) of the series. The two sets of X-12-ARIMA options chose exactly the same model for 99% of the series (3,222 of 3,258).

If we disregard inclusion of trading day and Easter effects, the model identification accuracy was much improved. TRAMO correctly identified the airline model with or without regressors for 85% (2,781) of the series. The two sets of X-12-ARIMA options each chose the airline model for 91% of the series (2,978 for the default settings, 2,964 for the modified settings).

As we saw with the actual data, X-12-ARIMA chose an Easter effect more often than TRAMO. For these simulated series TRAMO identified an Easter effect for

Table 3: Comparison of TRAMO and X-12-ARIMA Model Identification, Simulated Series

	Combined	TRAMO	X-12-ARIMA Default
Airline Model With No Regressors		66%	73%
Airline Model Without Regard to Regressors		85%	91%
Easter Effect			
Agreement No Easter Effect (Correct)	88%		
Agreement Yes Easter Effect	4%		
Disagreement (and Which Chose the Effect)	8%	0.3%	8%
Trading Day Effect			
Agreement No Trading Day Effect (Correct)	86%		
Agreement Yes Trading Day Effect	3%		
Disagreement (and Which Chose the Effect)	11%	10%	1%

4% (138) of the series, and X-12-ARIMA chose an Easter effect for 12% (375) of the series under the default options and for 11% (374) under the modified options. In this instance, we know there was not an Easter effect present, so we are concerned by this level of selection from X-12-ARIMA. The Easter effect regressors are significant according to their t statistics, but we are not sure why they would be. For the default X-12-ARIMA settings only, we looked at the length of the Easter effect that we were identifying: 3% (108) of all 3,258 series were 15-day effects, 3% (101) were 8-day effects, and 5% (166) were one-day effects.

Comparing the default X-12-ARIMA settings and TRAMO, there were 259 series for which they disagreed on the Easter effect, and for 8% (248), X-12-ARIMA incorrectly identified an Easter effect. For 0.3% (11 series) TRAMO identified an Easter effect. Using the binomial approach as before, the probability is less than 0.01 that we would see such a difference assuming equal probabilities of selection.

After our experiences with the actual data, we thought perhaps X-12-ARIMA might select trading day effects more often than TRAMO, but we were surprised to see that TRAMO identified a trading day effect 13% of the time (431 series), and each X-12-ARIMA set of options identified a trading day effect for 4% of the series (133 series). The methods disagreed for 354 series. TRAMO misidentified a trading day effect for 10% (326) of the series, and X-12-ARIMA misidentified a trading day effect for 1% (28) of the series. Using the binomial approach, the probability is less than 0.01 that we would see such a difference assuming equal probabilities of selection.

Table 3 summarizes our comparison results for the simulated series.

7. Conclusions

The new automatic modeling procedure has been released with X-12-ARIMA Version 0.3 and is available to time series analysts across the world. This evaluation was necessary for those users to know the usefulness and the limitations of the automatic modeling software. We saw from the simulated data that TRAMO more accurately determined that no transformation was needed, but our X-12-ARIMA settings were biased toward log transformations. We were concerned that X-12-ARIMA mistakenly selected an Easter effect significantly more often than TRAMO did. We were able to see that TRAMO missed a necessary trading day effect more often than X-12-ARIMA, but also, TRAMO mistakenly selected trading day effects more often than X-12-ARIMA for the simulated data. For our set of actual series, we saw that Ljung-Box Q and seasonal spectrum diagnostics for the X-12-ARIMA models were at least as adequate as for the TRAMO models.

8. Future Work

We hope to expand our study of simulated series to perform a more thorough evaluation of X-12-ARIMA's new automatic modeling procedure by including multiplicative processes and using more varied models, model coefficients, regression effects, and series lengths. More evaluation could indicate possible ways to improve the automatic modeling procedure, especially with regard to the selection of the Easter effect.

Acknowledgements

We thank Rita Petroni and James Gomish for their valuable comments, suggestions, and time in reading multiple drafts of this paper. We thank David Findley for his ideas and suggestions regarding the methods we used and plans for future work.

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.

Authors' Note

Much of the data analysis for this paper was generated using Base SAS® software, SAS/AF® software, and SAS/GRAPH® software, Versions 8 and 9 of the SAS System for Windows. Copyright © 1999-2003 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA. We used R to simulate the airline model time series data. Additional analysis was performed using Microsoft® Excel 2000. Copyright © 1985-1999 Microsoft Corporation. We checked our own calculations of the binomial probabilities involving the actual data using the Binomial Calculator on the Internet at onlinestatbook.com/java/binomialProb.html (home page at onlinestatbook.com/), and we used it alone for the comparisons involving the simulated data.

References

- Box, G. E. P. and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control* (2nd ed.). San Francisco, CA: Holden-Day.
- Cleveland, W. and Devlin, S. (1980), "Calendar Effects in Monthly Time Series: Detection by Spectrum Analysis and Graphical Methods," *Journal of the American Statistical Association* 75, 487–496.
- Dagum, E. B. (1988), "X-11-ARIMA/88 Seasonal Adjustment Method - Foundations and Users' Manual," Ottawa: Statistics Canada.
- Dent, A. M., Hood, C. C. H., McDonald-Johnson, K. M., and Feldpausch, R. M. (2005), "Comparing the Automatic ARIMA Model Selection Procedures of X-12-ARIMA Versions 0.2 and 0.3," *Proceedings of the American Statistical Association, Business and Economic Statistics Section* [CD-ROM].
- Farooque, G. M., Hood, C. C. H., and Findley, D. F. (2001), "Comparing the Automatic ARIMA Model Selection Procedures of TRAMO and X-12-ARIMA 0.3," *Proceedings of the American Statistical Association, Business and Economic Statistics Section*.
- Findley, D. F. (2005), "Some Recent Developments and Directions in Seasonal Adjustment," *Journal of Official Statistics*, Vol. 21, No. 2, pp. 343–365.
- Gómez, V. and Maravall, A. (1997), "Programs TRAMO and SEATS: Instructions for the User, Beta Version," Banco de España.
- Gómez, V. and Maravall, A. (2000), "Automatic Modeling Methods for Univariate Series," Chapter 7 of *A Course in Time Series Methods* (Eds. D. Pena, G. C. Tiao, R. S. Tsay) New York: Wiley.
- Ljung, G. M. and Box, G. E. P. (1978), "On a Measure of Lack of Fit in Time Series Models," *Biometrika* 65, 297–304.
- Soukup, R. J. and Findley, D. F. (1999), "On the Spectrum Diagnostics Used by X-12-ARIMA to Indicate the Presence of Trading Day Effects After Modeling or Adjustment," *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 144–149. www.census.gov/ts/papers/rr9903s.pdf.
- U. S. Census Bureau (2007), *X-12-ARIMA Reference Manual*, Version 0.3, Beta, U. S. Census Bureau, U. S. Department of Commerce.