1352.0.55.124

Australian Bureau of Statistics

Research Paper

Automated ARIMA Model Selection for Aiding Filter-Based Seasonal Adjustment

www.abs.gov.au

New Issue

Research Paper

Automated ARIMA Model Selection for Aiding Filter-Based Seasonal Adjustment

Alex Stuckey and Jonathan Campbell

Analytical Services Branch

Methodology Advisory Committee

1 June 2012, Canberra

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30 AM (CANBERRA TIME) TUES 22 OCT 2013

ABS Catalogue no. 1352.0.55.124

Produced by the Australian Bureau of Statistics

INQUIRIES

The ABS welcomes comments on the research presented in this paper. For further information, please contact Mr Philip Carruthers, Analytical Services Branch on Canberra (02) 6252 5307 or email <analytical.services@abs.gov.au>.

AUTOMATED ARIMA MODEL SELECTION FOR AIDING FILTER-BASED SEASONAL ADJUSTMENT

Alex Stuckey and Jonathan Campbell
Analytical Services Branch

QUESTIONS FOR THE COMMITTEE

1. Is the BPG (Best Practice Guidelines) algorithm a reasonable approach to ARIMA model selection for the purpose of aiding seasonal adjustment?
2. What further work should be undertaken to support the implementation of the BPG or other method into production at the ABS?

CONTENTS

APPENDIX

AUTOMATED ARIMA MODEL SELECTION FOR AIDING FILTER-BASED SEASONAL ADJUSTMENT

Alex Stuckey and Jonathan Campbell
Analytical Services Branch

ABSTRACT

Filter based methods of seasonal adjustment of time series data can be enhanced by forecasting. Time series can be extended via seasonal ARIMA model forecasts to reduce the reliance on asymmetric filters at the end of the series and thus reduce future revisions to seasonal factor estimates. Manual selection of appropriate models can impose a large time burden on analysts who must periodically re-assess these models for a large number of time series. Therefore, automatic procedures of model selection are preferred. In this paper we present an empirical study to evaluate several methods of selecting seasonal ARIMA models for the specific purpose of aiding seasonal adjustment. Our aim is to identify a model selection method that gives forecasts that are most effective in minimising revisions to publish seasonally adjusted and trend estimates, as produced with the X12-ARIMA seasonal adjustment procedure. In so doing, we compare the efficacy of three methods of model selection, including those contained in the packages TRAMO (Gómez and Maravall, 1997) and X12-ARIMA (U.S. Census Bureau) and the `R` package `forecast` (Hyndman, 2012). We also trial an *ad hoc* methodology based on past experience, as developed within the Australian Bureau of Statistics. The alternative procedures are compared via a simulation study and are ultimately evaluated on a number of real world data sets.

1. INTRODUCTION

The objective of this study is to assess methods of automated ARIMA model selection for the purpose of seasonal adjustment ameliorated by forecast estimates. Having approximately 1500 time series that employ ARIMA forecasting in their seasonal adjustment and are annually reanalysed, an automated method will reduce the analytical burden within the ABS. However, such a methodology must also create forecasts that aid high-quality seasonal adjustment.

Our first step towards an automated procedure is to create an algorithm that, as closely as possible, mimics what an ABS analyst would currently do. We compare the resulting models and seasonal adjustments from this approach to those resulting from some other existing ARIMA model selection algorithms.

1.1 Current ABS practice

The ABS uses a filter-based methodology for seasonal adjustment. This methodology is assisted by the use of forecasting at the current end of the series to reduce the reliance on asymmetric trend and seasonal filters in the process of seasonal adjustment. This technique has been shown, in most cases, to reduce the revisions to seasonally adjusted, and hence also trend, estimates. *Reduction in revisions to published, seasonally adjusted and trend estimates is the primary motivation for ARIMA forecasting for seasonal adjustment in the ABS.*

In practice, forecasting is done by first selecting a seasonal ARIMA model and fitting such a model using X12-ARIMA. Forecasts are made within SEASABS (the in-house software developed at the ABS which implements an X12-ARIMA variant) using this fitted model and the filter-based decomposition is done with these forecast-augmented data. Once the seasonal adjustment is complete the forecast values are then removed and the publication trend estimates are based only on the published seasonally adjusted estimates.

The choice of ARIMA model is currently made by trained analysts according to steps formulated at the ABS and documented in the ABS best practice guideline (BPG). The steps of the BPG are based on existing theory and have evolved with practical experience. The aim of the BPG is to arrive at a chosen model that is stable, adequate and effective. More details of the BPG method will follow.

2. LITERATURE REVIEW

2.1 Automated Box–Jenkins methodology

Numerous attempts have been made to automate the Box–Jenkins methodology to specify an ARIMA model for a time series since its inception in the early 1970's. One of the first such attempts was implemented by Hill and Woodworth (1980), which combines pattern recognition and an order testing criterion based on Akaike's FPE criterion to supply an appropriate model. Although this work is now some three decades old, the problems faced by the authors remain unchanged. These include the issue of which (if any) transform should be applied to data, how to identify and deal with short term transient effects in time series (in ABS terminology 'prior corrections') and how to parsimoniously identify an appropriate model for forecasting or describing a time series. The authors demonstrated that a large percentage of time series could be reliably forecast with significant time savings using the SIFT algorithm. Although this algorithm is now considered redundant, it did occupy a place in the literature for some time, as it represented one of the first attempts to supply an automated procedure for ARIMA model identification.

Hill and Fildes (1984) illustrated that the aforementioned algorithm compared favourably with alternative procedures considered in Makridakis *et al.* (1982). However, this is just one example of a host of competing algorithms that emerged at this time and in the intervening years. A further example of an early automated approach is CAPRI, as described by Libert (1984). A significant emphasis of this procedure is given to choosing the appropriate number of first-differences and seasonal-differences needed to form a stationary time series from a non-stationary one. This further illustrates that the difficulties faced by researchers when this problem was initially considered are much the same as those still existing today.

The problems associated with applying the differencing technique to make a series stationary were highlighted by Makridakis and Hibon (1997). They concluded that this approach often resulted in poorer accuracy of post-sample forecasts in ARIMA time series modelling than those of forecasts involving the removal and extrapolation of trend followed by ARMA modelling. Despite such criticism, the application of automated procedures to fit ARIMA models to time series has flourished.

Mélard and Pasteels (2000) outline a three step procedure (implemented in the software TSE-AX) for identifying an ARIMA model. The algorithm is composed of the following: (i) a choice of a difference, a seasonal difference and a transformation, (ii) specification of an ARIMA model, and (iii) model checking. Whilst such an approach is relatively standard, the technical details proposed by numerous researchers in each step vary greatly.

The approach to model specification suggested by Mélard and Pasteels (2000) involves the initial fitting of AR(p) models to 'stationarised' time series until the autocorrelations are approximately truncated above lag q and then representing the residuals by an MA(q) process. Such an approach is very similar to the earlier work of Hannan and Rissanen (1982) and has also been the subject of research by other authors. Alternatives (among many) include the so-called "Corner Method" described by Béguin, Gouriéroux and Monfort (1980), the TRAMO/SEATS methodology proposed by Gómez and Maravall (1997) and various implementations of the AIC and BIC (see Hyndman and Khandakar, 2008, for example). Additionally, much more complex algorithms are being developed for the specification of an appropriate ARIMA model for an observed time series. One such example is given by Valenzuela *et al.* (2004). However, such systems based on fuzzy logic and genetic algorithms are yet to play a significant role in the literature and are beyond the scope of this paper.

Further to the model specification problem, an important consideration is that of parameter estimation for small sample properties. One can consider a multitude of approaches to the estimation of parameters once a model has been chosen. Notwithstanding, this is not the focus of this paper and we refer the interested reader to De Gooijer and Hyndman (2006) who discuss this area more thoroughly.

A common approach to comparing the competing methodologies has been their performance in the M, M2 and M3 competitions. Although such comparisons have received criticism in the literature (see Ord, Hibon and Makridakis, 2000) there is a widespread belief that they provide a reasonable benchmark with which to judge the various algorithms. Perhaps a closer study to that which we undertake here is provided by McDonald–Johnson *et al.* (2007). However, this work concentrates on the identification of the regression parameters for trading day and Easter effect in the X12 algorithm rather than revisions which is the focus of our attention here.

2.2 ARIMA model specification by AIC

The popular `forecast` package in `R` includes the function `auto.arima`. It is well documented in Hyndman and Khandakar (2008) so that here we only give a brief outline of the approach. It can be executed with a choice of criteria, namely, it allows model choice based on use of the AIC, AICC or BIC. In our trial we restrict our attention to the AIC as previous studies have indicated that it has better performance over a wide range of time series. In the case that we assume a seasonal ARIMA model is appropriate, we define the criterion

$$AIC = 2\big(p + q + P + Q + k - \ln(L)\big).$$

where p, q, P and Q are, as usual, the number of AR, MA, SAR and SMA parameters with the value of $k = 1$ if the model contains a drift term and $k = 0$ otherwise.

Further, L is the maximised likelihood of the model when fitted to the differenced data.

The AIC was first developed by Akaike (1974) in the context of measuring information loss when a particular model is adopted to describe a data set. It has proven to be a very versatile and useful tool for measuring the relative likelihood of a competing set of models. In essence, the criterion balances the variance and bias of the estimation procedure by penalising the number of parameters in the model. For a more thorough discussion of this and the other information criteria the reader is referred to Bierens (2006).

A key feature of the auto.arima function is the initial application of a test for unit roots and the differencing of the time series if needed. An important consideration of the AIC is that it does not readily allow the comparison of models that undergo differencing with models that have not been differenced. Thus, a pre-modeling step of testing for stationarity is an important one. We implement the function with the default testing procedures of KPSS for first order differencing and the OCSB test for seasonal differencing. The default settings of the function have been chosen by the authors based on empirical performance over a large study set. A rigorous discussion of other available tests is available in Lopes (2001).

The primary difference between our study here and similar research that compares ARIMA model specification via the AIC is that our ultimate goal is the choice of a method that minimises revisions after the application of the X12-ARIMA procedure, whereas other studies to date have largely concentrated on forecasting accuracy. A secondary point of difference is that we will investigate the applicability of this procedure to a number of time series data sets that have heretofore not been investigated and for which there may be no true underlying ARIMA model.

2.3 ARIMA model specification in TRAMO and X12-ARIMA

The automatic ARIMA model specification contained within the X12-ARIMA program is based largely on the work of Gómez and Maravall (1997) and implemented in the software suite SEATS/TRAMO. We give a brief outline of the procedure here, as it is well documented elsewhere.

It is assumed that v_t follows an ARIMA process given by

$$\phi(B)\,\delta(B)\,v_t \;=\; \theta(B)\,a_t\,.$$

As a first step, the non-stationary polynomial $\delta(B)$ is obtained by an iterative approach based on results of Tsay and Tiao (1984) and Tsay (1984). The first order differences and seasonal differences are then obtained up to order $\delta^2\delta_s$ and checks for unit roots are made.

The second stage of the TRAMO identification is a modified form of the Hannan–Rissanen procedure. Hannan and Rissanen (1982) showed that p and q could be estimated by minimising the criterion

$$\log\left(\hat{\sigma}^2\right) + (p+q)\log(T)\Big/ T\,.$$

However, they implement an algorithm whereby the usual maximum likelihood estimation of σ^2 is replaced by a recursive procedure that involves constructing a sequence of regressions of $y(t)$ on $y(t-1), y(t-2), \ldots$ and $\varepsilon(t-1), \varepsilon(t-2), \ldots$.

The implementation (with its modifications) of this procedure in TRAMO then searches for the values of p, q, P and Q which minimises this criterion within the bounds $0 \le p,q \le 3$ and $0 \le P,Q \le 2$.

The automatic model selection specification in X12-ARIMA is based on TRAMO with modifications made by the U.S. Census Bureau.

3. ADDITIONAL METHOD

3.1 ABS best practice guidelines algorithm

The Best Practice Guidelines (BPG) used by the time series analysis section of the ABS is an evolving set of documents outlining the agreed procedures for seasonal adjustment. The documents cover many aspects of the seasonal adjustment process as diverse as defining aggregation structures, the annual seasonal reanalysis process, documenting analytical work and the presentation of time series estimates in publications.

One aspect of the BPG is a set of instructions that guide ABS analysts when selecting a seasonal ARIMA model for use in seasonal adjustment. The three broad criteria that an appropriate model should meet are that it is:

Adequate: The applied model should describe the underlying structure in the series and that what remains should be white noise. This may be evaluated using a range of model diagnostics on the residuals.

Stable: A stable model is one that describes the underlying structure of the data with a minimum number of parameters. More complex models may end up over-fitting some of the noise rather than describing the underlying nature of the series. Stable models are preferred because they are more likely to describe the signal in the data and hence produce more reliable forecasts.

Effective: By effectiveness we mean how good the model is at reducing revisions compared to the asymmetric filter. This must be evaluated on historical data which is only a good indication that the model will continue to produce good revisions into the future.

The task of finding such a model in the first instance and checking to see whether a previously selected model still meets these criteria at the time of subsequent reanalyses is a labour intensive and time consuming one. To increase efficiency an automatic procedure is preferred. To this end, the instructions in the BPG have been translated into an algorithm that, when available in production, can prompt an analyst with the preferred model and associated diagnostics. The details of this algorithm are included in Appendix A.

4. REAL DATA STUDY

4.1 Data

For the study based on real data we use ABS time series in the prior-corrected form. That is, corrections for large extremes, trend break, seasonal breaks and calendar related effects such as trading day and Easter proximity have already been made.

We take series from a variety of ABS publications. The publication areas, frequencies and numbers of series are given in table 4.1.

4.1 Real time series data summary

Group	*Frequency*	*Number of series*
Hours Worked stock series	Monthly	105
Labour Force	Monthly	294
New Motor Vehicles	Monthly	36
Retail Trade	Monthly	198
Household consumption volumes	Quarterly	174
Balance of Payments	Quarterly	92
Average Weekly Earnings	Quarterly	99
Livestock	Quarterly	49
Compensation of Employees	Quarterly	31
		1,078

The maximum span of data used in the modelling was 15 years of monthly data or 20 years of quarterly data. Series of less than 10 years length of either frequency were not included.

4.2 Methods used

In both the real data and simulated data studies we applied several methods and compared the results. The methods, abbreviated names and descriptions are as follows:

- TRAMO
  This is version 197 of TRAMO (at the time of writing) available on the Bank of Spain website.
- TRAMO+
  This is a recent, extended version of TRAMO not yet widely available.
- X12-ARIMA
  This refers to the automatic modelling procedure in version 0.3, build 188 of X12-ARIMA.
  This method is derived from TRAMO with some alterations made by the U.S. Census Bureau.

- `R:forecast`
  This is the *auto.arima* function from the `forecast` package in R, version 3.04.
- BPGA
  Best Practice Guidelines Algorithm is the automated procedure that mimics the instructions given to ABS time series analysts.
- Airline
  (0,1,1)(0,1,1) SARIMA model.
  This is a widely used model for seasonal time series.
- currentABS
  This is the model that is currently being used in the ABS environment.
  The model has been selected by an ABS analyst.
  For national accounts series (205 out of 1078) there is currently no ARIMA model specified.

Standard default options were used in the methods tested.

4.2 Identification method settings

Method	*Settings*	*Comments*
X12-ARIMA	automdl=TRUE max (p,d,q)(P,D,Q)=(3,2,3)(2,1,2)	
TRAMO	'RSA'=3 max (p,d,q)(P,D,Q)=(3,2,3)(1,1,1)	RSA is a range of settings RSA=3 is default settings P,Q≤1 in TRAMO
TRAMO+	'RSA'=3 max (p,d,q)(P,D,Q)=(3,2,3)(1,1,1)	RSA is a range of settings RSA=3 is default settings P,Q≤1 in TRAMO
R:forecast	allowdrift=FALSE stepwise=FALSE max (p,d,q)(P,D,Q)=(3,2,3)(2,1,2)	Default setting in auto.arima is for more lags. Maximum lags set here to be consistent with X12-ARIMA automodel.

For a given series, each method was employed to select an ARIMA model. This model was then specified within a call to X12-ARIMA (i.e. with automodeling turned off) which in turn, was used to carry out parameter estimation, forecasting, seasonal adjustment and trend estimation. X12-ARIMA resembles very closely the ABS production software SEASABS although it can be run by calling an executable from within R. This feature is not possible within SEASABS. Using X12-ARIMA allowed a larger scale study than would be feasible with SEASABS, while giving results that we expect to hold for seasonal adjustment in SEASABS also.

4.3 Results

Because the primary purpose of ARIMA forecasting in our context is that of reducing revisions to seasonally adjusted estimates, we wish to compare the level of revisions resulting from each method. We look at revisions from the very first published estimate (lag 0) to the benchmark estimate. The benchmark estimate being the final estimate reached once more data are available and asymmetric filters and forecast error no longer influence the seasonally adjusted estimate. We also look at revisions from lag 1 to benchmark.

Additionally, we include the corresponding revisions to the trend estimates produced by X12-ARIMA. These trend estimates are produced differently to the current ABS method however. Specifically, ABS published trends are based solely on the seasonally adjusted observed data and do not directly make use of forecast estimates, whereas the X12-ARIMA trend do.

As larger forecast errors are expected to lead to larger revisions, we include the mean squared forecast errors for each method also. These are divided by lowest mean square forecast error achieved for that series, over all model choices. This is to avoid results for series with overall high forecast errors to be overly influential in the results. The means of these measures are presented in the tables below for one-step-ahead (fce.1) and 12-steps-ahead (fce.12).

The results of the BPG algorithm tests are shown for each model selection method. This shows how often an ABS analyst would not have considered such a model for reasons of stability and/or adequacy.

We start by looking at the orders of differencing and the numbers of parameters chosen by each method.

4.3 Mean number of parameter or differences

	TRAMO	*TRAMO+*	*X12-ARIMA*	*R:forecast*	*BPGA*	*airline*	*current ABS*
p	0.5613	0.5897	0.4302	1.0256	0.3276	0.0000	0.0693
d	0.9639	0.9259	0.8917	0.6372	0.7474	1.0000	0.4311
q	0.7521	0.7493	0.7493	0.8357	0.3628	1.0000	0.3210
P	0.0912	0.1083	0.1111	1.0950	0.0684	0.0000	0.0057
D	0.8594	0.8053	0.8756	0.3875	0.6629	1.0000	0.4274
Q	0.8756	0.8357	0.9012	1.0180	0.6458	1.0000	0.4311

In terms of differencing, the automatic selection within TRAMO and X12-ARIMA chooses to difference more often than ABS analysts or `R:forecast` have for these series. `R:forecast` however has many more autoregressive terms (both seasonal and nonseasonal) than are being used at ABS.

The most striking characteristic of the models selected by the BPGA is that there are fewer AR and MA parameters (both seasonal and nonseasonal) selected than in the TRAMO methods, X12-ARIMA and particularly `R:forecast`.

In looking at the relative performance of each method with respect to revisions, we look at revisions from lag 0 (i.e. the very first published estimate) and lag 1 to the final estimate. Some series may be more prone to revision than others, so for each series we look at the difference between the revisions under a given method and revisions with no forecasting applied. The table below shows the means of these differences for each method. That is, a negative value means an overall reduction in revisions compared to having no forecasting applied.

As the only method that explicitly chooses a model based on low observed revisions, the BPGA shows revisions that are overall the lowest.

4.4 Revisions for all series (mean revisions minus revisions without forecasting)

	TRAMO	*TRAMO+*	*X12-ARIMA*	*R:forecast*	*BPGA*	*airline*	*current ABS*
rev.sa.0	–0.0270	–0.0725	–0.0762	–0.0813	–0.1609	–0.0707	–0.0502
rev.sa.1	–0.0092	–0.0400	–0.0437	–0.0366	–0.1506	–0.0386	–0.0564
rev.tr.0	–0.5541	–0.6079	–0.6211	–0.5892	–0.2904	–0.6084	–0.1449
rev.tr.1	–0.1142	–0.1313	–0.1294	–0.1219	–0.1405	–0.1255	–0.0805

The results for those series currently selected in ABS series ('current ABS') are slightly distorted by the inclusion of national accounts series that do not currently have ARIMA forecasting applied. Removing these series shows that the models currently used do not outperform exiting automatic methods at the lag 0 but are an improvement after the first extra data point (lag 1).

4.5 Revisions for series excluding national accounts (mean revisions minus revisions without forecasting)

	TRAMO	*TRAMO+*	*X12-ARIMA*	*R:forecast*	*BPGA*	*airline*	*current ABS*
rev.sa.0	–0.0227	–0.0767	–0.0783	–0.0855	–0.1651	–0.0700	–0.0613
rev.sa.1	–0.0027	–0.0389	–0.0405	–0.0337	–0.1565	–0.0348	–0.0689
rev.tr.0	–0.6249	–0.6894	–0.6996	–0.6614	–0.2985	–0.6869	–0.1770
rev.tr.1	–0.1360	–0.1568	–0.1526	–0.1433	–0.1598	–0.1491	–0.0984

The `R:forecast` method performs the best in terms of one-step-ahead forecasting. On average the mean squared forecast error is only 8% higher than that of the lowest for the series.

4.6 Forecast error for all series (mean ratio of forecast error relative to lowest forecast error)

	TRAMO	*TRAMO+*	*X12-ARIMA*	*R:forecast*	*BPGA*	*airline*	*current ABS*
fce.1	1.1842	1.2110	1.1379	1.0859	1.1291	1.1373	1.1157
fce.12	3.9742	5.6775	3.4479	2.1248	2.5481	1.4748	1.7100

The proportion of series in which the model selected by a given method fails the ABS-defined criteria of adequacy and stability are presented in table 4.7. As would be expected, the ABS analysts (currentABS) mostly select models that meet these criteria. Those that do not pass often fail the Ljung–Box test for autocorrelated residuals. The models automatically selected by TRAMO and X12-ARIMA are more likely to fail the BPG stability tests.

4.7 Proportion of model not meeting BPG criteria

	TRAMO	*TRAMO+*	*X12-ARIMA*	*R:forecast*	*BPGA*	*airline*	*current ABS*
overall	0.3371	0.3713	0.2669	0.9212	0.0000	0.2422	0.0617
adequacy	0.1548	0.1776	0.1500	0.7768	0.0000	0.1235	0.0332
stability	0.2650	0.2934	0.2023	0.8632	0.0000	0.1396	0.0342

4.4 Discussion of real data results

The results show that the models chosen by the BPG algorithm are different to the models currently used for many series. This is understandable as there are many detailed checks in the algorithm to which an analyst may not always strictly adhere. Also, the final selection criterion of the BPGA method is to compare a list of accepted candidate models against revisions performance. A similar manual comparison of many candidate models is not feasible and is therefore not performed currently. The fact that the final choice of the BPG algorithm explicitly chooses a model based on observed revisions performance explains the superior performance of the BPG algorithm on this measure. The BPGA method gave revisions of –0.16 and –0.15 respectively at these lags, with the next best results being obtained by the `R:forecast` package whose mean revisions were –0.08 and –0.04. The distribution of revisions for each of the methods is illustrated graphically at the end of this section.

In contrast, the mean of the standardised MSE of the forecasts is (not surprisingly) better for the Forecast package than the BPGA method. A very clear corollary of this simple fact is that better forecasts do not necessarily imply lower revisions (at least for seasonally adjusted estimates).

Of noticeable interest however, is that revisions to trend estimates at lag 0 are considerably better for all other tested methods than for the BPGA method. This could lend weight to the suggestion that the BPGA method is somewhat over-fitting models to enhance seasonal adjustment revisions at the cost of poorer results in other

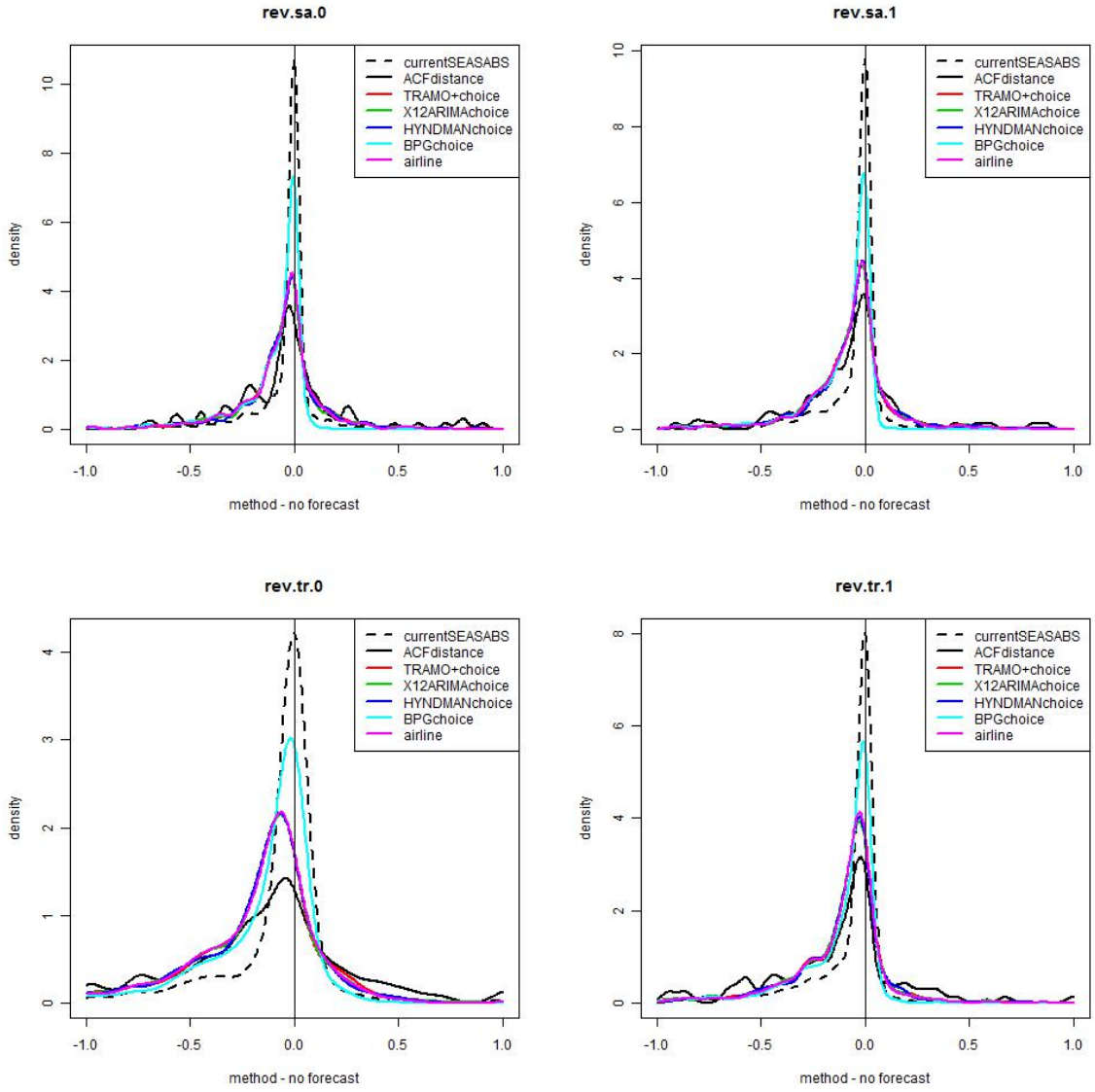
possible measures of goodness-of-fit. Additionally, it turns out here too, that the method with the best forecasts (`R:forecast`) do not produce the lowest average absolute percentage change to trend estimates. In this case, the TRAMO+ procedure produces lower revisions to trend estimates at all measured lags, whilst having 'worse' forecasts at all measured lags.

A further point which should be alluded to is that simply adopting the airline model for all series produces results which are only very slightly poorer than its counterparts in all measures and in some cases gives better results. Hence, in terms of general performance, the primary conclusion that one should draw from this analysis is that each method fares well in the measure that it was designed to achieve results in, yet no method can be considered the best overall.

An important aspect of our study was also the consideration of the rate at which non-BPGA methods 'failed' or 'passed' each of the various tests set out in the best practice guidelines and outlined in Appendix A. Most startlingly, this analysis reveals that arguably the 'best' two algorithms (The BPGA method and the `R:forecast` method) almost never identify the same model. This is most evident as, on more than 92% of the series analysed, the `R:forecast` method suggests an ARIMA model which fails at least one of the BPGA tests. Further, this method failed the BPGA tests quite uniformly on most fronts, including for example, an adequacy test on 77% of occasions (mostly having AR polynomial roots close to unity) and a stability test on 86% of occasions (mostly having correlated ARMA parameter estimates). This is to be expected as the authors Hyndman and Khandakar (2008) allow AR roots to approach much closer to unity (1.001) than the BPGA (a sliding scale beginning at 1.2).

Lastly, we consider the average number of parameters selected by each of the methods. The most obvious aspect of these results is the very low degree of differencing (both seasonal and at lag 1) of the `R:forecast` method. Clearly, this is a result of applying a hypothesis testing regime which is designed to avoid over-differencing. This fact is most evident in the mean number of seasonal differences. This quantity is 0.39 per series for `R:forecast`, 0.66 per series for the BPGA method and more than 0.8 for algorithms of TRAMO, TRAMO+ and X12-ARIMA. There is another very clear consequence of this distinct difference. Namely, the number of seasonal AR parameters is vastly greater for the Forecast package (being 1.09 on average per series) compared to the alternatives. Indeed, the BPGA method accepts a seasonal AR parameter on less than 7% of the series analysed. Noteworthy, also, is the discovery that the BPGA method (with one small exception) has on average the least number of parameters for each of the components p, q, P and Q despite only having a moderate number of average differences. We assume that this is due to the stringent tests on each model that must be passed for the model to be adopted.

4.8 Density of revisions for each method relative to revisions without forecasting



With respect to the figure above, note that:

(i) rev.sa.0 represents the revisions to the seasonally adjusted estimates at lag zero,

(ii) rev.sa.1 represents the revisions to the seasonally adjusted estimates at lag one,

(iii) rev.tr.0 represents the revisions to the trend estimates at lag zero, and

(iv) rev.tr.1 represents the revisions to the trend estimates at lag one.

5. SIMULATION STUDY

5.1 Data

A number of simulated data sets were constructed from a broad range of ARIMA models following the same structure as Maravall (2012). Our choice to use the same model types and parameters was largely driven by a desire to provide a comparable set of results to an alternative and recent study of ARIMA model selection. Moreover, the extensive class of ARIMA models considered in Maravall suited the purposes of our research well.

5.1 Models considered

Type A: Airline type models	*Type B: Non-seasonal models*		*Type C: Other seasonal models*	
	(a) Stationary	*(b) Non-stationary*	*(a) Stationary*	*(b) Non-stationary*
(0 1 1)(0 1 1)	(0 0 0)	(0 1 1)	(1 0 0)(1 0 0)	(0 0 0)(0 1 1)
(0 1 0)(0 1 1)	(1 0 0)	(0 1 0)	(0 0 0)(1 0 1)	(2 0 0)(0 1 1)
(0 1 1)(0 1 0)	(0 0 2)	(1 1 0)	(1 0 1)(1 0 0)	(0 0 0)(1 1 1)
	(1 0 1)	(1 1 2)	(0 1 2)(1 0 0)	(0 1 2)(0 1 1)
	(2 0 0)	(2 1 0)		(1 1 0)(0 1 1)
	(2 0 1)	(0 1 0)(1 0 0)		(0 1 1)(1 1 0)
	(3 0 0)	(0 2 1)		(1 1 2)(0 1 1)
		(0 2 2)		(1 1 1)(0 1 1)
				(0 1 1)(1 1 1)
				(2 1 0)(0 1 1)
				(3 1 0)(0 1 1)
				(0 2 1)(1 1 0)

The models considered fall into five broad categories and are summarised in table 5.1. For a full description of the models including parameter values, the reader is referred to Maravall. We note here however, that a great variety of parameters were used, including those giving awkward roots and near roots.

All simulations were created with a 'burn in period' to allow them to wander from zero. In the case of simulating from stationary models an additive constant was introduced. Once a series was made positive, it was appropriately scaled and the exponential function applied to make the series log-additive. The models which included a seasonal ARIMA component were generated with a 12 period frequency. Furthermore, we generated each series with 180 observations which corresponds to the minimum length for which the ABS best practice guidelines requires for an ARIMA model to be fitted for purposes of forecasting.

5.2 Results

For the simulation study we have elected to analyse the median (as opposed mean) of the revisions to seasonally adjusted estimates, trend estimates and forecast errors. This was done as an examination of our results indicated that a small number of series were being very poorly forecast by the various methodologies and rendered the mean of little use.

There are a number of distinct differences between the results obtained in the real data study and the simulated one. The most obvious difference is that the revisions to seasonally adjusted data for the BPGA method are worse than all alternative methods when dealing with simulated data, as opposed to the case where they were the best when dealing with real data. Indeed, the median absolute percentage change to revisions of seasonally adjusted estimates was –0.106 whilst the second worst methodology (being TRAMO+) had a much better value of –0.168. One may not argue that this difference is entirely due to the change from an examination of the median as opposed the mean of the results. Indeed, if we were to examine the mean results, the BPGA method is still not the best for the simulated data, being outperformed by the Airline model. We hypothesise that a second factor involved in this anomaly is due to the existence of prior corrections in the real data sets. Since the magnitude of the prior corrections were largely estimated using the BPG model, this methodology was at a distinct advantage. Hence, we are not seeing here a decrease in the performance of the BPGA algorithm, rather that its results were inflated when applied to the real data set.

A further distinction between the results for the two studies is that in the case of simulated data, the method which achieves the best forecasts (`R:forecast`) also achieves the best revisions to seasonally adjusted and trend estimates, whilst the worst forecasting method (the BPGA algorithm) also produces the worst revisions for all measures at all lags.

A seeming anomaly in the results is the rate at which each method recovers the true model for the simulations. The Forecast package has the worst recovery rate (10 %) of the true model yet still achieves the best results in terms of forecasts and revisions. The best recovery rate of the true model was provided by TRAMO+ (54.7 %).

An examination of the pass and failure rate of BPGA tests for the model selection algorithms in the setting of the simulated data are largely comparable to that observed in the case of the real data. Again, the Forecast package algorithm fails at least one test on more than 90% of occasions whilst all other methodologies produce models that pass the tests on approximately half of the series.

Likewise, an analysis of the mean number of parameters selected by each of the methodologies yields very similar results to that observed in the case of the real data. The BPGA method selected the least number of parameters on average and the `R:forecast` method selected the greatest number of parameters. Furthermore, `R:forecast` again invoked the least number of differences at both lag 1 and lag 12. A few general comments comparing the estimated number of parameters and differences with the number of true parameters and differences can be made. Discounting the Airline model (which is fixed), all methodologies under-differenced at lag 1 (although some only marginally), yet all over-differenced at lag 12. Additionally, all algorithms except the BPGA method estimated a greater number of parameters for all components (p , q , P and Q) than the true model.

5.2 Revisions for all series (mean revisions minus revisions without forecasting)

	TRAMO	*TRAMO+*	*X12-ARIMA*	*R:forecast*	*BPGA*	*airline*	*True*
rev.sa.0	–0.1877	–0.1678	–0.2068	–0.2099	–0.1063	–0.1875	–0.1621
rev.sa.1	–0.1644	–0.1492	–0.1824	–0.1843	–0.1010	–0.1439	–0.1397
rev.tr.0	–0.5555	–0.4773	–0.6774	–0.7045	–0.2591	–0.7278	–0.4712
rev.tr.1	–0.3084	–0.2811	–0.3742	–0.3766	–0.1394	–0.3980	–0.2784

5.3 Forecast error for all series (mean ratio of forecast error relative to lowest forecast error)

	TRAMO	*TRAMO+*	*X12-ARIMA*	*R:forecast*	*BPGA*	*airline*	*True*
fce.1	1.0561	1.0527	1.0379	1.0272	1.0624	1.1698	1.0553
fce.12	1.1539	1.1584	1.1034	1.0756	1.1632	1.2787	1.1466

5.4 Mean orders for different methods

	TRAMO	*TRAMO+*	*X12-ARIMA*	*R:forecast*	*BPG*	*airline*	*True*
p	0.8031	0.8256	0.8921	1.2586	0.3796	0.0000	0.7102
d	0.6667	0.6299	0.6940	0.4579	0.6999	1.0000	0.6448
q	0.6441	0.5967	0.7616	1.0451	0.4164	1.0000	0.7137
P	0.1554	0.2040	0.1803	0.9585	0.1329	0.0000	0.2304
D	0.5504	0.4484	0.5504	0.4116	0.5255	1.0000	0.4346
Q	0.4009	0.3511	0.4365	0.5516	0.3867	1.0000	0.3776

6. CONCLUSIONS

6.1 Summary of results

The study has provided good evidence that the BPG algorithm performs to a reasonable standard and would be a more than adequate replacement of an analyst's choice when identifying an ARIMA model to forecast time series for the purpose of reducing revisions to seasonally adjusted estimates. Notwithstanding, it remains unclear as to whether the BPG algorithm should be the preferred choice for an automated model selection method, as there are a number of competing criteria that could be used to rank the methodologies. Additionally, we recommend that the performance of each method should be trialled in a cross validation setting, to examine their out-of-sample performance. Furthermore, one should also consider the role of the ARIMA model in the context of detection and estimation of prior corrections to fully gauge the ultimate performance of a model selection algorithm.

One concern we have is that, in selecting the model based on observed revisions performance, the BPGA may be *over-fitting* to the data in a sense. The rather stringent stability and adequacy tests guard against this somewhat as the final model choice is only made from a set of candidate models that have passed these already.

This study also tested the models selected by some existing methods that have been developed for a variety of purposes. TRAMO and TRAMO+ methods have been developed specifically for choosing an ARIMA model to be a basis for signal extraction. X12-ARIMA automodel specification uses a modified version of TRAMO to aid filter-based seasonal adjustment. The `auto.arima` function from the forecast package in `R` was specifically developed to choose models that provide good forecasting.

The Forecast package algorithm has performed well against all criteria that it was set against. However, the models chosen often do not meet ABS defined criteria (BPG tests) and as such would be need to be implemented with a degree of caution.

It was demonstrated that the Airline model performs at least comparably with all other methodology over a large range of simulated and real data sets. Hence, this can be viewed as some validation that a large percentage of ABS time series are currently forecast with this model.

The TRAMO, TRAMO+ and X12-ARIMA methodologies performed adequately well in all measures, yet also had some difficulties in passing the BPG tests for a large percentage of the time series considered.

6.2 Recommendations

It is natural to suggest that the best criteria for identifying the ARIMA model which minimises revisions is simply to examine all possible models and use the magnitude (or some weighted sum of the magnitudes) of revisions at various lags as the ultimate criteria in model selection. In response to this, we could supply a number of arguments. If the gains to be had by choosing an unstable, complicated model over a 'nice' parsimonious model were minimal, many analysts would prefer to accept the slight losses. Moreover, there is a strong belief that the out-of-sample forecast properties of the simpler model will outperform that of the possibly over-parameterised model. After all, it is well understood that neither model is likely the 'true' model and that the estimation of the more complex model has captured transient effects in the data that are not likely to continue into the future. Additionally, as more data points are observed, the likelihood is that the estimation of parameters of a more complicated model are going to change to a much greater extent than those of a simpler model or even worse, the model itself will change. This process itself brings further revisions, as the forecasts made with the newly estimated model affect the estimation of the seasonally adjusted figure at the current end and through the filtering process, will also have an effect on past estimations.

6.3 Further work

There are several limitations to the current study that could be addresses in future work.

In the current study we used seasonally adjusted and trend estimates as produced by X12-ARIMA rather than the ABS production software SEASABS. This was done to enable large numbers of series and methods to be used in the R environment with an R function used to write X12-ARIMA specification files, run X12-ARIMA and read the output back into R for analysis. The differences between the final estimates produced by X12-ARIMA and SEASABS are no properly known although the core methods used by each are so similar that we expect methods that perform well under X12-ARIMA will also perform well under SEASABS.

Relatively simple, default settings were favoured in all methods without trialling multiple options. It is possible that different settings for some methods could results in model selections that further improve seasonal adjustment performance.

In particular, the weighting assigned to each test in the BPG method and the thresholds used in these tests could be considered somewhat arbitrary. Although they are based on experience there may be other settings more appropriate for optimal model choice for seasonal adjustment.

An important consideration when choosing an ARIMA model for forecasts in the X12 algorithm is that of regression for prior corrections. As the chosen ARIMA model plays a dual role in both forecasting the time series under analysis and estimating the location and magnitude of prior corrections, a thorough study of a methodology for model choice within the seasonal adjustment procedure should also incorporate some measure of a model's ability to estimate the occurrence of extreme values, trend breaks and seasonal breaks. To date however, we have not implemented any analysis of this issue and recognise its importance in future work. Nevertheless, it is not unreasonable to believe that the model which provides the best forecasts for minimising revisions should not also be a natural choice for estimating prior corrections.

There is an additional but very subtle point which could also be raised with respect to prior corrections. Namely, the time series which we analysed were the prior corrected series and for which, the estimation of the prior corrections were undertaken with a particular ARIMA model in use. In large part, the model in use was the BPG model and as such, the revision results will tend to appear more favourable for that particular model. In other words, if the prior corrections of the original series had of been estimated within one of the alternative model identification frameworks the results could have been somewhat different. However, the scope of this paper has not extended to include a measure of this effect.

ACKNOWLEDGEMENTS

The authors would to like thank those ABS staff responsible for developing the ABS best practice guidelines for ARIMA model selection. Responsibility for any errors or omissions remains solely with the authors. Thanks to Agustin Maravall for providing the prototype version of TRAMO+.

The authors are also indebted to the Methodology Advisory Committee (especially Rob Hyndman and Alistair Gray) for their helpful comments and advice, which we hope to address in a future extension to this work.

REFERENCES

Akaike, H. (1974) "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control*, 19(6), pp. 716–723.

Béguin, J.M.; Gouriéroux,C . and Monfort,A . (1980) "Identification of Mixed Autoregressive-Moving Average Process: The Corner Method", in O.D. Anderson (ed.), *Time Series*, pp. 423–436, North–Holland, Amsterdam.

Bierens, H.J. (2006) *Information Criteria and Model Selection*, unpublished manuscript, Pennsylvania State University.

De Gooijer, J.G. and Hyndman, R.J. (2006) "25 years of Time Series Forecasting, *International Journal of Forecasting*, 22(3), pp. 443–473.

Gómez, V. and Maravall, A. (1997) *Programs TRAMO and SEATS, Instructions for the User*, Working paper no. 97001, Ministerio de Economía y Hacienda, Dirección General de Análisis y Programación Presupuestaria.

Hannan, E.J. and Rissanen, J. (1982) "Recursive Estimation of Mixed Autoregressive-Moving AverageOrder", *Biometrika*, 69(1), pp. 81–94.

Hill, G. and Fildes, R. (1984) "The Accuracy of Extrapolation Methods: An Automatic Box–Jenkins Package, SIFT", *Journal of Forecasting*, 3(3), pp. 319–323.

Hill, G.W. and Woodworth, D. (1980) "Automatic Box–Jenkins Forecasting", *Journal of the Operational Research Society*, 31(5), pp. 413–422.

Hyndman, R.J. and Khandakar, Y. (2008) "Automatic Time Series Forecasting: The `forecast` Package for `R`", *Journal of Statistical Software*, 27(3).

Hyndman, R.J. (2012) "`forecast`: Forecasting Functions for Time Series and Linear Models", R package version 3.04. <http://CRAN.R-project.org/package=forecast>

Libert, G. (1984). "The M-Competition with a Fully Automatic Box–Jenkins Procedure", *Journal of Forecasting*, 3(3), pp. 325–328.

Lopes, A.C.B. da S. (2001) "The Robustness of Tests for Seasonal Differencing to Structural Breaks", *Economics Letters*, 71, pp. 173–179.

Makridakis, S.; Andersen, A.; Carbone, R.; Fildes, R.; Hibon, M.; Lewandowski, R.; Newton, J.; Parzen, E. and Winkler, R. (1982) "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition", *Journal of Forecasting*, 1(2), pp. 111–153.

Makridakis, S. and Hibon, M. (1997) "ARMA models and the Box–Jenkins Methodology", *Journal of Forecasting*, 16(3), pp. 147–163.

Maravall, A. (2012) *Is Reg-ARIMA Modelling Reliable in Large-Scale Seasonal Adjustment?* Keynote Speech to the 2012 Workshop on Methodological Issues in Seasonal Adjustment, Eurostat, Luxembourg. (Last viewed 25 July 2013) < http://www.cros-portal.eu/sites/default/files/WSSA%20Paper%20Luxemburgo%2003%2012.pdf >

McDonald–Johnson, K.M.; Hood, C.C.H.; Monsell, B.C. and Li, C. (2007) *Comparing Automatic Modeling Procedures of TRAMO and X-12-ARIMA, An Update*, Working paper, U.S. Census Bureau. (Last viewed 25 July 2013) <http://www.census.gov/ts/papers/ices2007kmj.pdf>

Mélard, G. and Pasteels, J.–M. (2000) "Automatic ARIMA Modeling including Interventions, Using Time Series Expert Software", *International Journal of Forecasting*, 16(4), pp. 497–508.

Ord, K.; Hibon, M. and Makridakis, S. (2000) "The M3–Competition", *International Journal of Forecasting*, 16(4), pp. 433–436.

Piccolo, D. (1990) "A Distance Measure for Classifying ARIMA Models", *Journal of Time Series Analysis*, 11(2), pp. 153–164.

R Development Core Team (2011) R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna. <http://www.R-project.org/>

Tsay, R.S. (1984) "Order Selection in Nonstationary Autoregressive Models", *Annals of Statistics*, 12(4), pp. 1425–1433.

Tsay, R.S. and Tiao, G.C. (1984) "Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models", *Journal of the American Statistical Association*, 79(385), pp. 84–96.

Valenzuela, O.; Márquez, M.; Pasadas, M. and Rojas, I. (2004) "Automatic Identification of ARIMA Time Series by Expert Systems Using Paradigms of Artificial Intelligence", *Monografías del Seminario Matemático García de Galdeano*, 31, pp. 425–435.

APPENDIX

A. BEST PRACTICE GUIDELINES TESTS

A list of candidate models is first generated. Presently this comprises of the union of

- the best five models as selected by the X12-ARIMA automodel specification,
- a list of eight commonly occurring, simple models, and
- the incumbent model, as defined in SEASABS (if available).

For the purpose of this study we also include the models as selected by the `R:forecast` package, TRAMO and TRAMO+ as candidate models for the BPGA method.

Each model is given an adequacy weight initialised to zero. If the model fails a particular test its weights will increase at a level depending on how badly it fails, i.e. depending on the p-value. A weight of 100 for any particular test will result in a failure of that test.

A.1 Model adequacy tests

The model adequacy weight is set to zero and if this weight reaches 100 after the adequacy tests below it is deemed inadequate and is removed from the candidate model list.

A.1.1 Unit root test

An AR polynomial root close to unity indicates possible under-differencing. That is, the series has not been differenced enough to achieve stationarity, leading to the model assumptions being invalid, and an inappropriate model fit. As a rule of thumb, roots under 1.2 should be investigated further.

An MA polynomial root close to one, which occurs frequently in ABS monthly time series (in the seasonal component of the model), will generally have a very low associated standard error. This is often partially due to difficulties with estimating SEs for parameter estimates which lie close to boundary values. If all other diagnostics indicate the model is adequate, then we do not consider this to be a problem.

The rules below are meant to indicate when there may be a problem with over-differencing, which can then be explored by the analyst in further detail. Another approach to this may be to use unit root testing during the model compilation phase – this will clarify the amount of differencing which should be specified for the model. Currently this is done by X12v0.3 only when the *automodel* specification is used.

Rules

a. If root on AR parameter is between 1.05 and 1.1 the weight of the candidate model is increased by 50.

b. If root on AR parameter is between 1.1 and 1.2 the weight of the candidate model is increased by 20.

c. If root on AR parameter < 1.05 then this is considered a serious breach of adequacy and the weight of the candidate model is increased by 100 (essentially discarding the model).

d. If root on AR parameter is >1.2 the weight of the candidate model is unchanged. If sum of non-seasonal AR coefficients is between 0.975 and 1.025 the weight is increased by 40.

e. Else, if the sum of non-seasonal AR coefficients is between 0.95 and 1.05 the weight is increased by 20.

f. If sum of seasonal AR coefficients is between 0.975 and 1.025 the weight is increased by 40.

g. Else, if sum of seasonal AR coefficients is between 0.95 and 1.05 the weight is increased by 20.
If root on seasonal MA parameter is between 1.0 and 1.1 the weight of the candidate model is increased by 40.

h. If root on non-seasonal MA parameter is between 1.0 and 1.1 the weight of the candidate model is increased by 40.

A.1.2 Test for normality and remaining autocorrelation in residuals

The idea here is that the assumptions on which ARIMA models are determined specify that the residuals should be normally distributed and contain no significant, unexplained autocorrelation. If there is remaining, unexplained structure in the residuals, this may indicate a poorly specified model and the possibility of inappropriate results through using this model. There are a number of tests which can be applied to examine the normality of the residuals and X12 v0.3 includes the following as a default, Skewness coefficient, Geary's a, Kurtosis and Histogram of the Standardized/Mean-Centred Residuals.

X12 v0.3 summarises these tests into a simple yes/no test for normality of residuals, stating (*No indication of lack of normality*). Note that the Ljung–Box Q-stat is also available when *pickmodelling* is run and this can be applied as an objective test of remaining autocorrelation of the residuals. The Ljung–Box test-statistic and p-value are also available from lag 24 and lag 12 of the residual ACF for monthly and quarterly series respectively.

Rules

a. If skewness test fails at the 1% level then increase the weight by 50.

b. If at least one of the kurtosis tests (Geary's a or Kurtosis coefficient) fails at the 1% level then increase the weight by 50. Failure of both a. and b. is a serious breach of adequacy and resulting in a cumulative weight of 100, essentially discarding this model.

c. Failure of the Ljung–Box Q-stat test at the 1% level indicates a serious breach of adequacy and increases the weight by 100, essentially discarding this model. A Ljung–Box result between 1 and 5% increases the weight by 50 while a Ljung–Box result between 5 and 10% indicates the weight should be increased by 20.

At this point, all models with an adequacy weight of at least 100 should be discarded.

A.2 Model stability tests

Each candidate model at this stage has a stability weight that is initialised to zero.

After the following stability tests:

1. Check significance of parameters,

2. Check correlation of ARMA parameters,

a model is removed from the candidate model list if the stability weight reaches 100.

A.2.1 Checking significance of parameters

If the parameters are all significant this is an indication that the model may be adequate. If the parameters are all insignificant this model can be effectively removed from the candidate list by giving it a high weight at this point. For models with marginally/almost significant parameter estimates the idea here is that standard errors and t-stats give an indication of how certain we are that a parameter is significantly different from 0. A general rule of thumb is that if the t-value is greater than 2, then the parameter is significant.

Rules

a. If a parameter estimate has a t-stat < 1 this is a serious breach of model stability and will increase the weight by 100, effectively discarding the model.

b. If a parameter estimate has a t-stat between 1. and 1.5 the candidate model is given a weight of +50.

c. If a parameter estimate has a t-stat between 1.5 and 2 the candidate model is given a weight of +20.

d. If all parameter estimates have a t-stat > 2 the candidate model is given a weight of $+0$.

e. Include test of the ARMA coefficient correlation matrices here. This has been included in the SR for X12 v0.3 in SEASABS, so it will now be automatically available. Suggest if a coefficient has correlation above 0.25, increase the weight by 50, above 0.5 increase the weight by 100, which is a serious breach.

Any model with a stability weight of at least 100 should be discarded at this stage.

A.2.2 Checking correlation of ARMA parameters

An assumption of the ARIMA model fitting is that the parameters are independent. Here we check for correlation between ARMA parameter estimates.

Taking the upper triangle of the ARMA coefficient correlation matrix, we penalise a model 50 points for a correlation above 0.25 and another 50 points for a correlation above 0.5. This means that if any pairs of parameter estimates have a correlation above 0.50 the model will not be considered.

A.3 Model effectiveness tests

The idea of this test is to check the candidate ARIMA model is fit for purpose. In this case, that the application of the ARIMA model appears to have reduced the required amount of revision on average before the benchmark estimate is achieved. We sum the difference in average percenatge revisions at lags 1 and 0. Let r^f be the average percentage revision to the benchmark for the forecasting estimates, let r^c be the average percentage revision to the benchmark for the concurrent estimates, then calculate the total difference in average percentage revisions ($TAPR$) across each lag.

$$TAPR = \sum_{l=0}^{m} \left(r_l^c - r_l^f \right).$$

Here $m = 1$, since we are focussing on lag 1 and 0 only.

Rule

a. Rank all candidate models by their $TAPR$ weight with the most efficient $TAPR$ models receiving the lowest weight of 1, with low weights being perceived as good.

$$TAPR_{\text{weight}} = 100\left(1 - \frac{TAPR_{\text{candidate}}}{TAPR_{\text{best}}}\right) + 1.$$

At this step we discard ineffective models, that is, any models for which $TAPR_{\text{weight}} > 100$ fail this test. If it is found that no models meet this criterion, then forecasting should not be applied. The model that best reduces revisions at these lags is then selected.

FOR MORE INFORMATION . . .

INTERNET **www.abs.gov.au** The ABS website is the best place for data from our publications and information about the ABS.

LIBRARY A range of ABS publications are available from public and tertiary libraries Australia wide. Contact your nearest library to determine whether it has the ABS statistics you require, or visit our website for a list of libraries.

INFORMATION AND REFERRAL SERVICE

Our consultants can help you access the full range of information published by the ABS that is available free of charge from our website, or purchase a hard copy publication. Information tailored to your needs can also be requested as a 'user pays' service. Specialists are on hand to help you with analytical or methodological advice.

PHONE 1300 135 070

EMAIL client.services@abs.gov.au

FAX 1300 135 211

POST Client Services, ABS, GPO Box 796, Sydney NSW 2001

FREE ACCESS TO STATISTICS

All statistics on the ABS website can be downloaded free of charge.

WEB ADDRESS www.abs.gov.au