

A Test for Model Choice in Seasonal Adjustment

Un Test per la Scelta del Modello per la Destagionalizzazione

Fabio Bacchini, Roberto Iannaccone, Edoardo Otranto ¹

Dise

Istat

bacchini@istat.it - iannacco@istat.it - otranto@istat.it

Riassunto: Nella pratica quotidiana gli istituti di statistica si trovano spesso nelle condizioni di dover scegliere quale strategia adottare per la produzione di dati destagionalizzati. Il presente lavoro ha come obiettivo l'identificazione di una procedura che renda possibile la scelta tra due differenti strategie utilizzando l'approccio di tipo model-based implementato da TRAMO-SEATS: una prima in base alla quale il modello è mantenuto fisso ogni volta che, durante l'anno, si rende disponibile una nuova osservazione e la seconda che considera una nuova identificazione del modello in presenza di nuova informazione. La procedura identificata utilizza un test basato sulla distanza tra filtri ed una appropriata funzione di perdita.

Keywords: distance, revisions, bootstrap.

1. Introduction

In this paper we investigate the characteristics of the procedure proposed in Bacchini *et al.* (2001) to help researchers for the choice between two different strategies in model-based seasonal adjustment methods implemented in TRAMO-SEATS Gomez and Maravall (1996). The first approach uses a fixed model with recursively estimated model parameters (hereafter concurrent estimator) and the second one updates both model specification and model parameter estimates when a new observation is available. Istat, for example, uses the concurrent estimator. This practice was recommended in the context of X11-ARIMA by Dagum (1982). However it can not be blindly adopted in a model-based context where the role of the identified model is crucial.

To compare the strategies we evaluate the distance measure discussed in Depoutot and Planas (1998) using the seasonal adjusted series estimated by the two approaches. Because of the lack of an analytical distribution for this distance, for each month we derive it empirically using bootstrap techniques to obtain a statistical test. Moreover we combine in a loss function the previous results with the costs due to the change of the model.

Given a precise value for the change of the model, the loss function is an operative tool for the choice between the two strategies.

As example for the procedure we consider the series of the total orders from January 1990 to December 1998.

¹We would like to thank Anna Ciammola for the helpful suggestion.

2. Methodological tools

The model based approach implemented by TRAMO-SEATS uses the Wiener-Kolmogorov filter to estimate the unknown components of a time series; under the hypothesis of orthogonal components this filter has properties of optimality.

To compare two different seasonal adjusted series we utilize the distance discussed in Depoutot and Planas (1998). In general, let $f_1(B)$ and $f_2(B)$ two symmetric filters with the same length r which extract two alternative seasonal adjusted series; the distance proposed can be defined as:

$$d(f_1, f_2) = (f_{10} - f_{20})^2 + 2 \sum_{k=1}^r (f_{1k} - f_{2k}) \quad (1)$$

where f_{1k} and f_{2k} are the coefficients of the Wiener-Kolmogorov filters for the two approaches at the lag k . Because of the lack of an analytical distribution for the distance in the previous equation, for each month we built up a simulation study to test the null hypothesis that the distance between the two filters would be not significantly different from 0. To do this we operate simulating the model with the ARIMA parameters estimated with the second approach bootstrapping the residuals.

The value of the distance is a first important indicator to consider for the decision to change the seasonal adjustment strategy during the year. The cost of the change of the model and the persistence of a significative distance have to be considered in a general decisional framework. With this in mind we introduce the following general loss function:

$$L = f(C_0, d, p) \quad (2)$$

where C_0 is the cost for the strategy change, d is the distance in equation 1 and p is the number of consecutive months with significative distance value. A simple specification for the function in (2) could be:

$$L = C_0 - \alpha p d_t \quad (3)$$

with $0 < \alpha < 1$ and C_0 as a constant that expresses, for example, the mean of the distance in the previous year. In this case the model change is performed when $L < 0$.

3. Empirical application

To illustrate our procedure we consider the series of total orders from January 1990 to December 1998. The model for this series up to observation corresponding to December 1997 is an ARIMA(3,1,1)(0,1,1). Considering the second approach described in the introduction, the model changes to an ARIMA(3,1,0)(0,1,1) for February and March, back to an ARIMA(3,1,1)(0,1,1) for April and finally, is an Airline model from May through the end of the year (see the second column in table 1).

In Table 1 are also reported, for each month, the AR and the MA parameters of the derived seasonal adjusted models for the first (first row) and second strategies (second row).

The distance calculated using equation (1) and the critical value resulting from bootstrap are reported in the last column.

Simulations are built up generating 1000 couples of series from the model identified using the full information for each month. For each couples we have estimated the seasonal adjusted series and then we have calculated the distance. Finally, we have derived the

empirical distributions for each one of the 12 months.

As expected, in the 10 cases with distance approximately zero, the null hypothesis is largely accepted. In February and May the distance values are similar but the empirical distribution are different: only in May we reject the null hypothesis.

Two points are important to stress. First it is possible to calculate the distance also when the model is not invertible (for example in February for the second approach); second the deterministic effects estimated are the same in both approaches so the distance is not dependent from different outliers or calendar effects.

In this case the evaluation of the loss function will be equal to C_0 for each month while in May the value of equation (3) will be $C_0 - 0.24\alpha$ ($p = 1$). If the mean of the distance for the current year is the same as the distance for the previous year then $C_0 = 0.041458$. So if $\alpha \geq 0.1727$ the best strategy will be the change of the model.

Table 1: Parameters of the models for seasonal adjusted components and relative distance

	ARIMA Model for original series	AR polynomials for seasonal adjusted series					MA polynomials for seasonal adjusted series					Distance (Critical values 95%)
jan	(3,1,1)(0,1,1)	-2.57	2.13	-0.57	0.00	0.00	-1.54	0.63	-0.09	0.00	0.00	0
	(3,1,1)(0,1,1)	-2.57	2.13	-0.57	0.00	0.00	-1.54	0.63	-0.09	0.00	0.00	(-)
feb	(3,1,1)(0,1,1)	-2.55	2.11	-0.55	0.00	0.00	-1.54	0.63	-0.09	0.00	0.00	0.2425
	(3,1,0)(0,1,1)	-1.64	0.36	-0.07	0.61	-0.26	-0.99	0.00	0.00	0.00	0.00	(0.2434)
mar	(3,1,1)(0,1,1)	-1.70	0.45	-0.05	0.56	-0.26	-1.06	0.07	0.00	0.01	0.00	0.0029
	(3,1,0)(0,1,1)	-1.18	-0.28	0.10	0.36	0.00	-0.57	-0.42	0.00	0.00	0.00	(0.2351)
apr	(3,1,1)(0,1,1)	-2.55	2.10	-0.55	0.00	0.00	-1.57	0.67	-0.10	0.00	0.00	0
	(3,1,1)(0,1,1)	-2.55	2.10	-0.55	0.00	0.00	-1.57	0.67	-0.10	0.00	0.00	(-)
may	(3,1,1)(0,1,1)	-2.56	2.11	-0.56	0.00	0.00	-1.57	0.67	-0.10	0.00	0.00	0.2431
	(0,1,1)(0,1,1)	-2	1	0.00	0.00	0.00	-1.36	0.36	0.00	0.00	0.00	(0.0698)
jun	(3,1,1)(0,1,1)	-1.75	0.51	-0.02	0.51	-0.25	-1.15	0.15	0.00	0.00	0.00	0.0068
	(0,1,1)(0,1,1)	-2	1	0.00	0.00	0.00	-1.36	0.36	0.00	0.00	0.00	(0.0567)
jul	(3,1,1)(0,1,1)	-1.75	0.51	-0.01	0.49	-0.24	-1.16	0.16	0.00	0.00	0.00	0.0012
	(0,1,1)(0,1,1)	-2	1	0.00	0.00	0.00	-1.35	0.36	0.00	0.00	0.00	(0.0606)
aug	(3,1,1)(0,1,1)	-1.68	0.40	-0.01	0.25	-0.24	-1.08	0.09	0.00	0.00	0.00	0.0005
	(0,1,1)(0,1,1)	-2	1	0.00	0.00	0.00	-1.36	0.37	0.00	0.00	0.00	(0.0657)
set	(3,1,1)(0,1,1)	-1.16	0.37	0.01	0.54	-0.25	-1.08	0.10	0.00	0.00	0.00	0.0003
	(0,1,1)(0,1,1)	-2	1	0.00	0.00	0.00	-1.36	0.37	0.00	0.00	0.00	(0.0698)
oct	(3,1,1)(0,1,1)	-1.69	0.42	-0.02	0.54	-0.25	-1.10	0.11	0.00	0.00	0.00	0.0000
	(0,1,1)(0,1,1)	-2.00	1	0.00	0.00	0.00	-1.36	0.37	0.00	0.00	0.00	(0.064)
nov	(3,1,1)(0,1,1)	-1.69	0.40	0.03	0.50	-0.24	-1.08	0.09	0.00	0.00	0.00	0.0000
	(0,1,1)(0,1,1)	-2	1	0.00	0.00	0.00	-1.35	0.36	0.00	0.00	0.00	(0.0629)
dec	(3,1,1)(0,1,1)	-1.66	0.33	0.06	0.54	-0.26	-1.03	0.05	0.00	0.00	0.00	0.0002
	(0,1,1)(0,1,1)	-2	1	0.00	0.00	0.00	-1.32	0.34	0.00	0.00	0.00	(0.0716)

4. Concluding remarks

In this paper we propose a method based on a distance measure to decide when identify a new model for the seasonal adjustment during the year leaving the concurrent approach. Moreover we introduce a loss function to take in consideration also the cost for model change and the persistence of a significative distance for consecutive months. This method was applied to the Italian order series that shows many changes in the model in the 1998. To use distance as a test procedure we have estimated empirical distribution via bootstrap. Further steps will deal with the study of analytical distribution and an extensive application of the loss function to identify plausible values of the constants.

References

- Bacchini F., Iannaccone R. and Otranto E. (2001) Model stability and model based seasonal adjustment, *Modelli Stocastici e Metodi Computazionali Intensivi per la Stima e la Previsione, Bressanone, 24- 26 settembre 2001*, 57–62.
- Dagum E.B. (1982) Revisions of seasonally adjusted data due to filter changes, *Proceedings of Business and Economic section of the American Statistical Association*, 39–45.
- Depoutot R. and Planas C. (1998) *Comparing Seasonal Adjustment Methods and Trend Extraction Filters with application to a model-based selection of X11 linear filters*, Working paper 9/1998/A9, Luxembourg, Eurostat.
- Gomez V. and Maravall A. (1996) *Program SEATS and TRAMO: Instruction for the User*, Working paper 9628, Bank of Spain.