THE IMPACT OF SEASONAL ADJUSTMENT ON TIME SERIES PREDICTION

Adela SASU

"Transilvania" University of Braşov

Abstract: *In this article we intend to compare the performance in term of accuracy for two models ARIMA(p,d,q)(P,D,Q)$_s$, corresponding to two kinds of time series: seasonally adjusted and not seasonally adjusted . The time series used in this work are publicly available. The measure of accuracy used is mean absolute percentage error. Experimentally, we show that the prediction accuracy is better for seasonally adjusted time series.*

Key words: *forecasting, time series, seasonally adjusted, ARIMA.*

1. INTRODUCTION

A *time series* is a sequence of observations $y_1, \dots, y_{t-2}, y_{t-1}, y_t$ generated sequentially in time. The key properties of time series are: the data are not independently generated, their variance may vary in time, they are often governed by a trend and they may have cyclic components. Statistical procedures that suppose independent and identically distributed data are, therefore, excluded from the analysis of time series.

Time series analysis includes a broad spectrum of exploratory and hypothesis testing methods with two main goals: (a) identifying the nature of the phenomenon represented by the sequence of observations, and (b) forecasting , i.e. predicting future values of the time series variable.

Seasonal adjustment (SA) is the process of estimating and removing seasonal effects from a time series in order to better reveal certain non-seasonal features. The mechanics of seasonal adjustment involve breaking down a series into trend-cycle, seasonal, and irregular components:

- *Trend – cycle*: Level estimate for each month (quarter) derived from the surrounding years of observations.
- *Seasonal effects*, defined as effects that are reasonably stable in terms of annual timing, direction, and magnitude. Possible causes include natural factors (weather), administrative measures (starting and ending dates of the school year), and social/cultural/religious traditions (fixed holidays such as Christmas).
- *Irregular components*, that is anything not included in the trend-cycle or the seasonal effects (or in estimated trading day or holiday effects). Their values are unpredictable as regards timing, impact, and duration. They can arise from sampling error, non-sampling error, unseasonable weather, natural disasters, strikes, etc.

X-12-ARIMA is the seasonal adjustment software produced and maintained by the Census Bureau. It is used for all official seasonal adjustments at the U. S. Census Bureau. The original time series is not seasonal adjusted (NSA). Methods for seasonally adjusting time series are also described in [5], [6].

2. The ARIMA Model

Autoregressive Integrated Moving Average (ARIMA) processes are a class of stochastic processes used in the area of time series modeling. The application of the ARIMA methodology for the study of time series analysis is due to Box and Jenkins [1].

Let us consider $y_1 \ldots y_{t-1}, y_t, y_{t+1}, \ldots$ (shortly: $\{y_t\}_t$) the observations at equally spaced time moments and let $a_1 \ldots a_{t-1}, a_t, a_{t+1}, \ldots$ or $\{a_t\}_t$ be a white noise series consisting of independent and identically distributed random variables, whose distribution is approximately normal with mean zero and variance σ_a^2. Assume that $E(y_t) = \mu_y$ and we note $y_t - \mu_y = \tilde{y}_t$; therefore, $E(\tilde{y}_t) = 0$.

Let us consider the general ARMA(p, q) model as in [2]:

$$\begin{aligned}\tilde{y}_t &= \phi_1 \tilde{y}_{t-1} + \ldots + \phi_p \tilde{y}_{t-p} + a_t \\ &\quad - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}\end{aligned} \tag{1}$$

or, equivalently

$$\phi(B)\tilde{y}_t = \theta(B)a_t \tag{2}$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ is the autoregression operator of order p and $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q$ is the moving average operator of order q, with B being the backward shift operator $By_t = y_{t-1}$, $B^k y_t = y_{t-k}$.

The general model ARIMA(p, d, q) is defined as in [3]:

$$\phi(B)(1-B)^d \tilde{y}_t = \phi(B)\nabla^d \tilde{y}_t = \theta(B)a_t \tag{3}$$

where $\nabla = 1 - B$ is backward difference, and $\nabla^d = (1-B)^d$ is the backward difference of order d.

A common assumption for many time series techniques is that the data are stationary. A stationary process has the property that the mean, variance and autocorrelation structure do not change over time. Stationarity can be defined in precise mathematical terms, but for our purpose we mean a flat looking series, without trend, constant variance over time, a constant autocorrelation structure over time and no periodic fluctuations. If the roots of polynomial $\phi(B)$ lie outside the unit circle, it may be shown that an ARMA(p, q) is stationary.

Many time series are not stationary. It is often the case that the series of first differences $w_t = y_t - y_{t-1} = (1-B)y_t$ is stationary. If a series $\{y_t\}_t$ has to be differenced once to obtain stationarity, then the model corresponding to the original series is called an integrated ARMA model of order *p, 1, q* or ARIMA(*p, 1, q*). In practice, differencing on the first order is necessary, while differencing on the second order is rarely needed. If the original series $\{y_t\}_t$ is stationary, then it is not necessary to differentiate these series.

However, if time series manifest a periodic fluctuation (a seasonal pattern), then the general ARIMA model is defined following [3]:

$$\Phi(B^s)\nabla_s^D y_t = \Theta(B^s)a_t \tag{4}$$

where s is the number of periods in a season. Let us note the seasonal ARIMA model ARIMA(p,d,q)(P,D,Q)$_s$, where P=number of seasonal autoregressive terms, D=number of seasonal differences, Q=number of seasonal moving average terms.

In [1], Box and Jenkins suggested that the search for a good model could be based on the following:

(i) Model identification, i.e. deciding on (initial values for) the orders *p; d; q; P;D;Q*,

(ii) Estimation, i.e. fitting of the parameters in the ARIMA model,

(iii) Diagnostic checking and model criticism,

(iv) Iteration: modifying the model (i.e. the orders *p; d; q; P;D;Q*) in the light of (iii) and returning to (ii).

The implementation provided by IBM SPSS Statistics analysis package version 21 was used for ARIMA. As ARIMA is delivered as an SPSS procedure at least from version 13 of this product, we believe that the current version is error-free. The steps for producing the most appropriate ARIMA model are detailed in [6].

3. Experiments and results

The overall performance of a forecasting model is evaluated by an accuracy measure, *Mean Absolute Percentage Error (MAPE)* computed as:

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \cdot 100 \tag{7}$$

where y_t is the desired value, $\hat{y}_t$ is the predicted value for period t, and N is the number of forecasted values. *MAPE* is a common metric in forecasting applications and it measures the proportionality between the forecasting error and the actual value.

The five time series used in this section are:

- US Total New Privately Owned Housing, 01.1968 – 06.2014
- US Retail Sales and Food Services, 01.1992 – 06.2014
- Florida Construction; All Employees, 01.1990 – 06.2014
- Manufacturing Information Technology Industries Total Inventories Millions of Dollars , 01.1992 – 06.2014
- Manufacturing Wood Products Inventories to Shipments Ratio, 01.1992 – 06.2014

They are freely available at [8, 9, 10, 11, 12, 13, 14, 15, 16, 17]. Each of these time series are given both as *seasonally adjusted* (SA) and as *not seasonally adjusted* (NSA).

For the ARIMA model, the values for the coefficients p, d, q, P, D, and Q are automatically determined by the IBM SPSS software, through its internal routines.

After the forecast is induced, the MAPE values are computed and the values of MAPE are grouped in Table 1:

Table1. The MAPE values of the considered time series

SERIES	MAPE	
	NSA	SA
US Total New Privately Owned Housing	6.118693643	5.772784487
US Retail Sales and Food Services	1.289655212	0.702631293
Florida Construction;	3.606306025	2.046015739
Manufacturing Information Technology Industries	1.77913964	0.711533582
Manufacturing Wood Products Inventories to Shipments Ratio	4.135570639	2.929566385

CONCLUSION

As shown in Table 1, all the values of MAPE for SA time series are lower the ones for NSA. In summary, the MAPE values for the SA series are between 39.9% and 94.3% of the corresponding NSA MAPE scores. In four out of the five cases, the MAPE score for SA is less than 71% of the MAPE for NSA. We conclude that seasonal adjustment of time series improves the prediction performances and it is recommended to be performed as a preprocessing step before analyzing of time series, despite the supplemental computational resources needed.

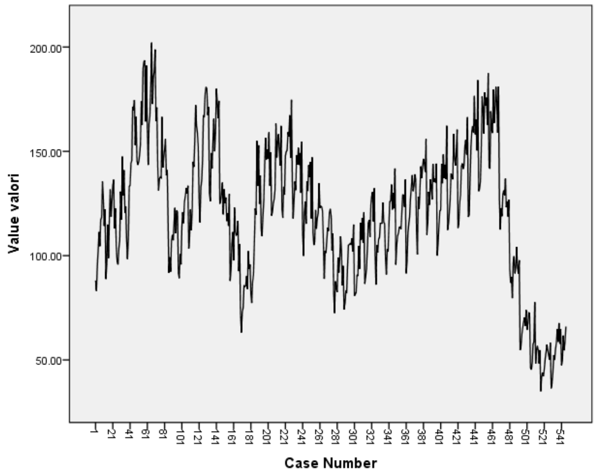

Fig. 1(a) US Total New Privately Owned Housing Units Completed; Thousands; NSA between 01.1968 and 06.2014 [8]

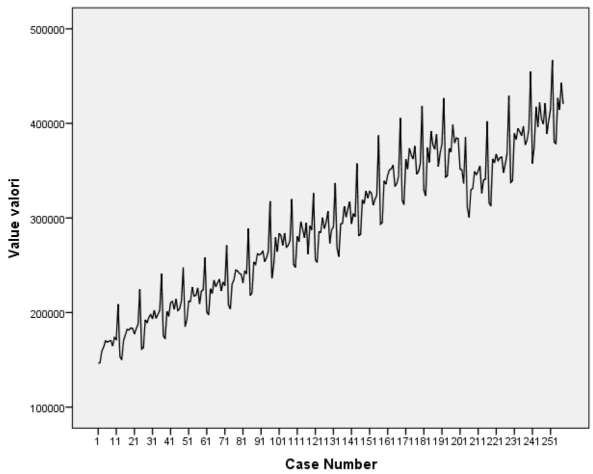

Fig. 2(a) US Retail Sales and Food Services, Total, NSA between 01.1992 and 06.2014 [10]

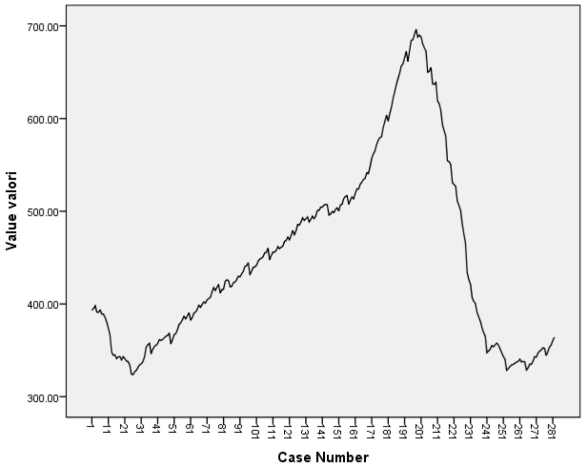

Fig. 3(a) Florida Construction; All Employees; Thousands; NSA between 01.1990 and 06.2014 [12]

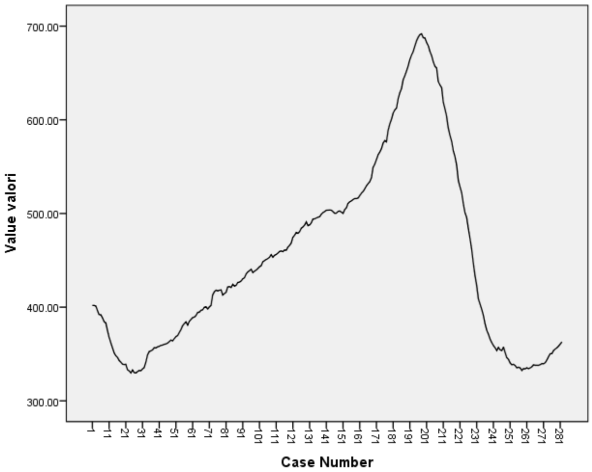

Fig. 3(b) Florida Construction; All Employees; Thousands; SA between 01.1990 and 06.2014 [13]

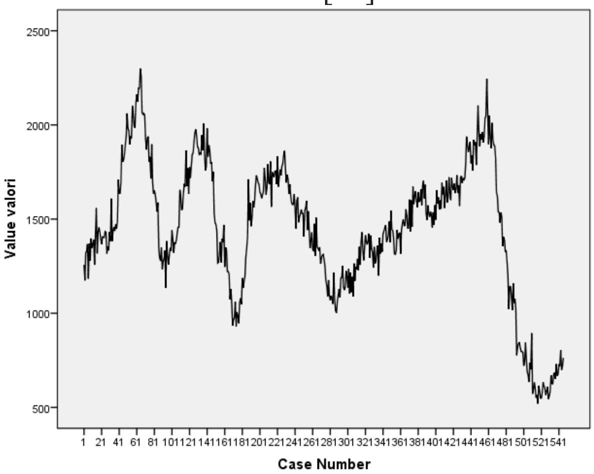

Fig. 1(b) US Total New Privately Owned Housing Units Completed; Thousands; SA between 01.1968 and 06.2014 [9]

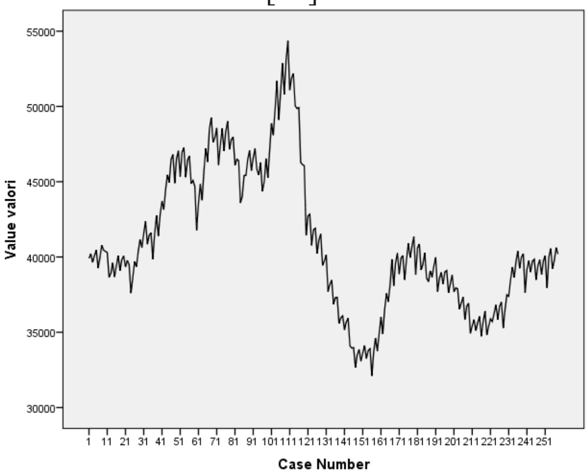

Fig. 4(a) Manufacturing Information Technology Industries Total Inventories Millions of Dollars NSA between 01.1992 and 06.2014 [14]

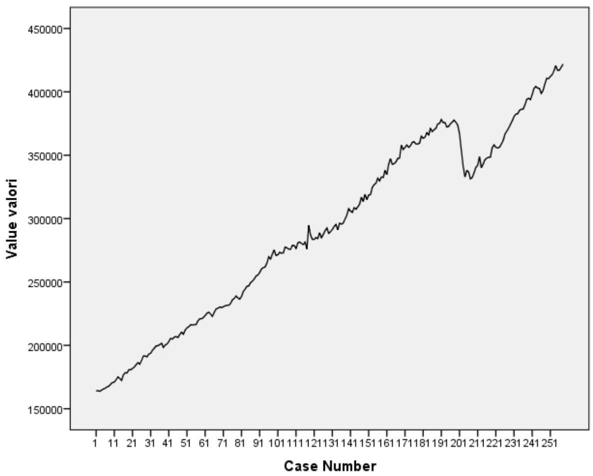

Fig. 2(b) US Retail Sales and Food Services, Total, SA between 01.1992 and 06.2014 [11]

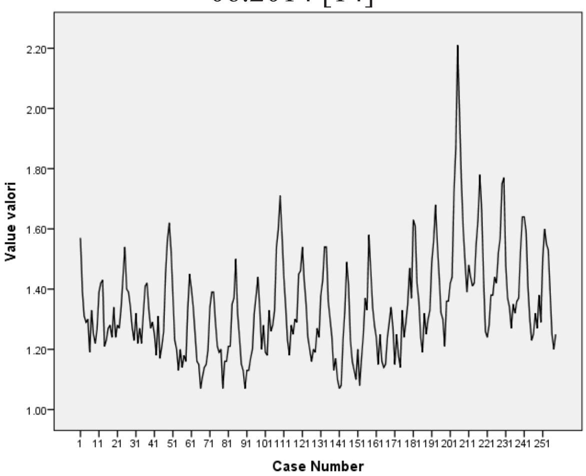

Fig. 5(a) Manufacturing Wood Products Inventories to Shipments Ratio NSA between 01.1992 and 06.2014 [16]

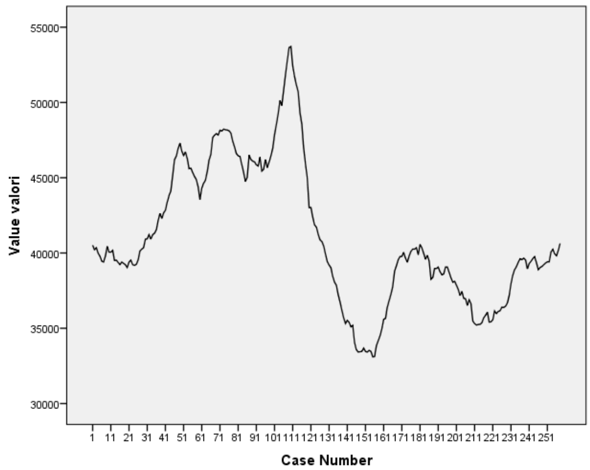

Fig. 4(b) Manufacturing Information Technology Industries Total Inventories Millions of Dollars SA between 01.1992 and 06.2014 [15]

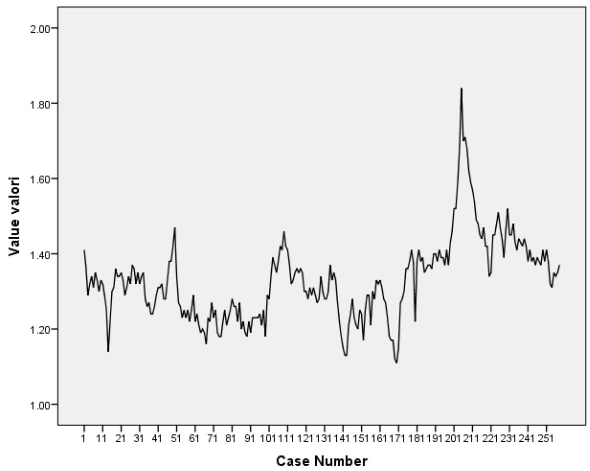

Fig. 5(b) Manufacturing Wood Products Inventories to Shipments Ratio SA between 01.1992 and 06.2014 [17]

BIBLIOGRAPHY

1. Box, G.E.P., Jenkins, G.M., *Time Series Analysis, Forecasting and control,* San Francisco: Holden-Day (1976).
2. Chatfield, C., *The Analysis of Time Series. An Introduction*, New York: Chapman & Hall/CRC(1995).
3. Popescu, T., *Time Series Application in System Analysis*, Technical Publishing House(2000) (in Romanian).
4. Sasu, A., An Application of ARIMA Models, *Bulletin of the Transilvania University of Braşov 11(**46**) - New Series, Series B1*, p. 61-68(2005).
5. Sasu A., Time series seasonal adjustment method, *Bulletin of the Transilvania University of Brasov 14(49), Series B1.* Transilvania Univ. Press-Brasov, 91-94, (2007).
6. Sasu, A., Time Series Forecasting using ARIMA Models, *8-th European Conference E-COMM-LINE* , 67e, (2007).
7. Yaffee, R.,A., McGee, M., *An Introduction to Time Series Analysis and Forecasting: With Applications of SAS and SPSS*, Academic Press(2000).
8. US Total New Privately Owned Housing Units Completed; Thousands; NSA: http://www.economagic.com/em-cgi/data.exe/cenc25/compua01
9. US Total New Privately Owned Housing Units Completed; Thousands; SA: http://www.economagic.com/em-cgi/data.exe/cenc25/compsa01
10. US Retail Sales and Food Services, Total, NSA: http://www.economagic.com/em-cgi/data.exe/cenret/nrt101nsa
11. US Retail Sales and Food Services, Total, SA: http://www.economagic.com/em-cgi/data.exe/cenret/nrt101sa
12. Florida Construction; All Employees; Thousands; NSA: http://www.economagic.com/em-cgi/data.exe/blssm/SMU1200000200000001
13. Florida Construction; All Employees; Thousands; SA: http://www.economagic.com/em-cgi/data.exe/blssm/SMS1200000200000001

14. Manufacturing Information Technology Industries Total Inventories Millions of Dollars NSA: http://www.economagic.com/em-cgi/data.exe/cenm3/uititi

15. Manufacturing Information Technology Industries Total Inventories Millions of Dollars SA: http://www.economagic.com/em-cgi/data.exe/cenm3/aititi

16. Manufacturing Wood Products Inventories to Shipments Ratio NSA: http://www.economagic.com/em-cgi/data.exe/cenm3/u21sis

17. Manufacturing Wood Products Inventories to Shipments Ratio SA: http://www.economagic.com/em-cgi/data.exe/cenm3/a21sis