The 34th International Symposium on Forecasting (ISF 2014). International Institute of Forecasters, Rotterdam, 2014.

Time series models for different seasonal patterns.

Andreozzi, Lucía, Blaconá María Teresa y Luciana Magnano.

Time series models for different seasonal patterns

Blaconá, M.T, Andreozzi, L. and Magnano, L.
National University of Rosario

Abstract

In this paper Innovations Space State Models (SSM) are used to fit series with: 1) a single seasonal period and 2) multiple seasonal periods. Sales data of 1) axles and 2) suspensions of a metallurgical company from Alvear (Santa Fe, Argentina) are analyzed as series with a single seasonal pattern. To analyze series with complex seasonal patterns, the series of a) daily vehicles passing through the toll booth on the Rosario-Buenos Aires (Argentina) highway and b) Las Rosas (Santa Fe, Argentina) daily average gas consumption per customer measured in m^3. The main purpose of these comparisons is to obtain predicted values with an acceptable error and a controllable level of uncertainty. Another reason for these comparisons is that Argentinean series show more variability than those with more stable development countries.

In series with a single seasonal pattern, ETS models have a good post-sample forecasting performance. The out-of-sample average forecast error five-step-ahead are 9.4% and 6.9%, for axles and suspensions, respectively, with a controllable level of uncertainty. BATS (Exponential Smoothing State Space model with Box-Cox transformation, ARMA errors, Trend and Seasonal Components) and TBATS (Trigonometric Exponential Smoothing State Space model with Box-Cox transformation, ARMA errors, Trend and Seasonal Components) are introduced to forecast complex seasonal time series. The results show that both types of models are suitable to describe and predict the time series of daily vehicles. The TBATS model has some advantages over the BATS model such as: i) better goodness of fit (lower AIC), ii) lower out-sample forecast percentage for different horizons (measured by MAPE); reduction in computation time to estimate the model, given the smaller number of seed values.

However, for the gas demand data, the performance of the proposed models is not as good, the BATS model does not show a good fit, and although the TBATS model fits the data well, it provides forecasts with more error than a MEE with *Spline.* A possible explanation for the lower quality forecasts of the TBATS, is that in this application TBATS models do not include explanatory variables that are included in the SSM, and it is known that climatic variables have much influence on utilities demand series. However, given the simplicity these models, they cannot be completely discarded.

I. Introduction

There are several approaches to model series with a single seasonal pattern, such as exponential smoothing (Holt-Winters, 1960), seasonal ARIMA models (Box and Jenkins, 1970), the state-space models (SSM, Harvey, 1989) and the innovations ETS (Hyndman et al., 2008).

However, models for multiple seasonal patterns are not as frequent, to cite some examples: SSM with *spline* for daily series (Harvey and Koopman, 1993) exponential smoothing models for double seasonality (Taylor and Snyder, 2009, Taylor 2010) and innovations space state models (ETS) for complex seasonal patterns (De Livera et al 2011), BATS (Exponential Smoothing State Space model with Box-Cox transformation, ARMA errors, Trend and Seasonal Components) and TBATS (Trigonometric Exponential Smoothing State Space model with Box-Cox transformation, ARMA errors, Trend and Seasonal Components).

In this paper the innovations space state models (ETS) are used in series with:

1. A single seasonal pattern and
2. Multiple seasonal patterns

Sales data of 1) axles and 2) suspensions of a metalworking firm in Alvear (Santa Fe, Argentina), that spans from January 2009 to August 2013 are analyzed as series with a single pattern. For the analysis of series with complex seasonality, two series are considered: a) daily vehicles passing through the toll booth on the Rosario - Buenos Aires (Argentina) highway in the period April 22, 2010 to December 31, 2013 b) Las Rosas (Santa Fe, Argentina) daily average gas consumption per customer measured in m^3 in the period March 1, 2008 to August 31, 2011.

The main purpose of these comparisons is to obtain predicted values with an acceptable error and a controllable level of uncertainty. Another reason for these comparisons is that Argentinean series show more variability than those from more stable development countries.

ETS, BATS and TBATS models are presented briefly in section II. Applications over four time series are shown in section III and concluding remarks are stated in section IV.

II. Methodology

Holt-Winters (HW) exponential smoothing methods are widely used to forecast time series with a single seasonal pattern (additive or multiplicative) providing good results. However, this framework presents two weaknesses: models cannot be estimated by maximum likelihood and calculating prediction intervals are not allowed.

Two superior proposals are State Space Models with multiple sources of error (SSM, Harvey, 1989) and innovations State Space Models with a single source of error (ETS, Hyndman et al, 2008). Exponential smoothing methods mentioned above have been studied in the framework of state space models. The SS innovations models include those underlying the additive and multiplicative methods of Holt Winters.

Extending the method of HW over a seasonal pattern, Taylor (2003) incorporates a second seasonal component, and when the number of seasonal components is large, it can be difficult to estimate the parameters and the seed values. Furthermore, if the seasonal period is large, the model obtained is likely to be overparametrized. This can be simplified when a seasonal period is multiple of the other. The exponential smoothing model assumes that the process of white noise is serially uncorrelated. This assumption is not always true in practice because sometimes behave as an AR (1) process.

De Livera et al. (2011) propose modifications to the ETS models in order to include a wide variety of seasonal patterns and solve the problem of correlated errors.

To avoid falling into nonlinearity problems, these authors restricted the models to those homoskedastic and in the the Box-Cox transformations (Box and Cox, 1964) is used in case of some type of specific non-linearity. The model including the transformation of Box and Cox, ARMA errors and seasonal patterns can be written as follows

$$y_t^{w} = \begin{cases} \dfrac{y_t^{w} - 1}{w}, & w \neq 0 \\ \log y_t, & w = 0 \end{cases}, \tag{1.a}$$

$$y_t^{w} = \ell_{t-1} + \phi b_{t-1} + \sum_{i=1}^{T} s_{t-m_i}^{i} + d_t, \tag{1.b}$$

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha d_t, \tag{1.c}$$

$$b_t = (1-\phi)b + \phi b_{t-1} + \beta d_t, \tag{1.d}$$

$$s_t^{i} = s_{t-m_i}^{i} + \gamma_i d_t, \tag{1.e}$$

$$d_t = \sum_{i=1}^{p} \varphi_i d_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t . \tag{1.f}$$

y_t^{w} is used to represent the Box and Cox transformed observation of with parameter ω, where y_t is the observation at time t. $m_1, m_2, ..., m_T$ are seasonal periods, ℓ_t is the local level in the period t, b is the long-term trend, b_t is the short-term trend in the period t, s_t^{i} represents the i-th seasonal component in the time t, d_t can be an ARMA model (p, q) and the process ε_t is a white noise Gaussian process with zero mean and constant variance σ^2. Smoothing parameters are given by α, β and γ_i for $i = 1, ..., T$. ϕ is the damping constant of the trend. This change ensures that the value of the short-term b_t trend converges to the value b (Long-term trend), rather than towards zero.

These models are called BATS as an anachronism of the key features of the model: Box and Cox transformation, ARMA errors, trend and seasonal components. The arguments $\omega, \phi, p, q, m_1, m_2, ..., m_T$ are the Box-Cox parameter, damping parameters, ARMA model parameters and seasonal periods respectively.

The BATS model is the most obvious generalization of the traditional seasonal innovation models that allow multiple seasonal periods.

To provide a flexible and parsimonious approach De Livera et. Al. (2011) introduced a trigonometric representation of the seasonal components based on West and Harrison Fourier series form (1997) and Harvey (1989) in the following manner:

$$s_t^{i} = \sum_{j=1}^{k_i} s_{j,t}^{i}, \qquad (2.a)$$

$$s_{j,t}^{i} = s_{j,t-1}^{i}\cos\lambda_j^{i} + s_{j,t-1}^{*i} sen\lambda_j^{i} + \gamma_1^{i} d_t, \qquad (2.b)$$

$$s_{j,t}^{*i} = -s_{j,t-1}^{i} sen\lambda_j^{i} + s_{j,t-1}^{*i}\cos\lambda_j^{i} + \gamma_2^{i} d_t, \qquad (2.c)$$

where γ_1^{i} and γ_2^{i} are the smoothing parameters and $\lambda_j^{i} = \frac{2\pi j}{m_i}$ describes the stochastic level of the i-th seasonal component as $s_{j,t}^{i}$ .Then, $s_{j,t}^{*i}$ is stochastic growth of the i-th seasonal component, needed to describe seasonal changes over time. The number of harmonics required for the i-th seasonal component is denoted by k_i.

A new class of MEE innovations is obtained by replacing the seasonal component in Equation (1c) by the trigonometric formulation, and then the measurement equation is given by

$$y_t^{w} = \ell_{t-1} + \phi b_{t-1} + \sum_{i=1}^{T} s_{t-1}^{i} + d_t$$

This class of models is called TBATS, initial T refers to trigonometrical transformation. If the arguments are listed, the TBATS model is written: TBATS $\omega,\phi,p,q,\ m_1,k_1\ ,\ m_2,k_2\ ,...,,\ m_T,k_T$ where each argument has the same meaning as in the models BATS , and k_i is the number of harmonics for the seasonal component s_t^{i} . TBATS model requires the estimation of $2\ k_1 + k_2 + ... + k_T$ initial values, a number that is generally smaller than the number of seed parameters in BATS models. Another advantage is that the trigonometric function can be used for models with non-integer seasonal frequency (eg daily for a year as 365.25 to contemplate leap years).

In a general linear approach ETS unknown, smoothing and damping parameters are estimated using the sum of squares errors of a Gaussian likelihood. In this context we also have to estimate the Box and Cox transformation parameter, and ARMA coefficients.

The forecasts distribution for the future period in the transformed space, given the final state vector and parameters, is Gaussian. The associated random variable $y_{n+h|n}^{w}$ has mean $E\ y_{n+h|n}^{w}$ and variance $V\ y_{n+h|n}^{w}$, considering the Box and Cox transformation.

Akaike criterion $AIC = \mathbf{L} + 2j$ is used to perform the selection, $\mathbf{L}$ is the likelihood and j is the number of parameters.

Model is repeatedly adjusted with a gradual growth to determine the number of harmonics for each seasonal component maintaining constant all other harmonics to achieve the minimum AIC.

A two stages method is used to select the orders p and q of the ARMA model a. First of all, a suitable model is selected regardless of the ARMA model for the residuals, then the automatic ARIMA algorithm of Hyndman and Khandakar (2008) is applied to determine the orders of the residual ARMA model (assuming stationarity). The selected model is fitted again but with an ARMA p,q model for the error component, where ARMA coefficients are estimated jointly with the other parameters. The ARMA component is allowed only if the AIC of the resulting model is lower than the model without it.

Mean Average Percentage Error (MAPE) is used to estimate the efficiency of the models to forecast h steps ahead.

$$MAPE\ h = \frac{1}{h}\sum_{i=1}^{h}\left|\frac{\hat{y}'_{n+i|n} - y_{n+i}}{y_{n+i}}\right| \times 100$$

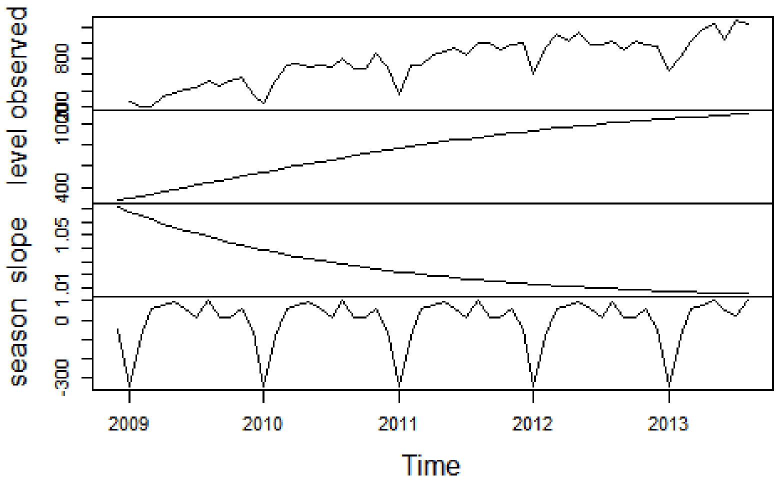
III. Application [1]

Firstly two sets of sales data, with a single seasonality are analyzed using ETS models.

Series 1: Number of axles sold by a metallurgical company Alvear (Santa Fe, Argentina) in the period January 2009 to August 2013. The series shows increasing trend and seasonality. The best model from Akaike criterion is ETS (A, M $_{d,}$ A), that means an additive error damped multiplicative trend and additive seasonality. The decomposition properly represents the features of the time series, Figure 1. Out-of sample 5-steps-ahead forecasts and the prediction intervals are presented in Figure 3. All actual values fall within the prediction intervals. Out of sample for 5 month ahead is 9.4%.

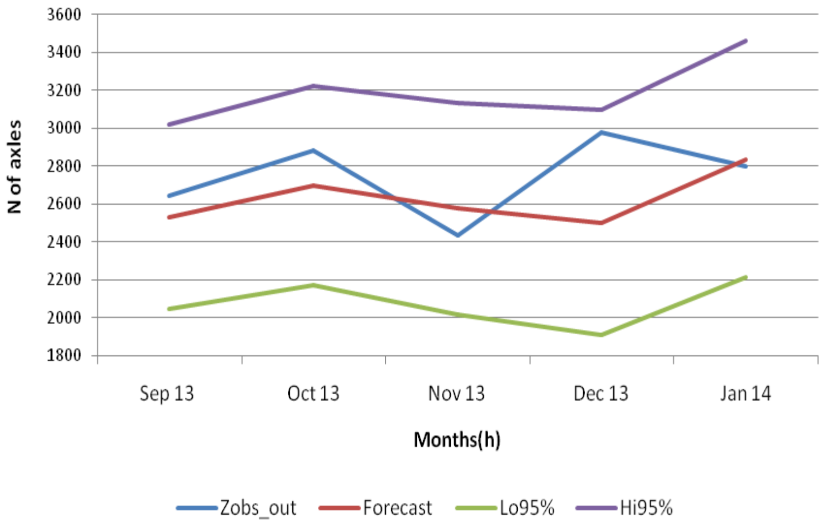
Figure 1. Decomposition of the number of axes sold by a metallurgical company from Alvear (Santa Fe, Argentina). January 2009 to August 2013.

[1] R software is used in all the application (Development Core Team, 2011)



Source: Own calculations based on data from a private company.

Figure 2. Forecasts of the number of axles sold by a metallurgical company from Alvear (Santa Fe, Argentina).September 2013-January 2014.

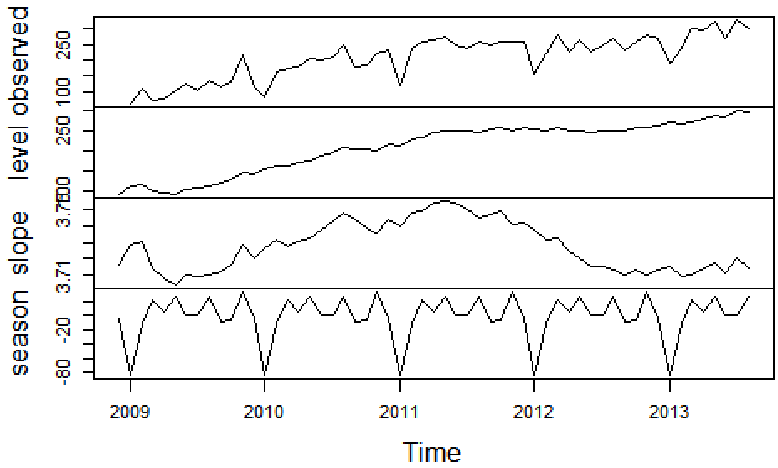


Source: Own calculations based on data from a private company.

Series 2: Number of suspensions sold by a metallurgical company from Alvear (Santa Fe, Argentina).January 2009 to August 2013.

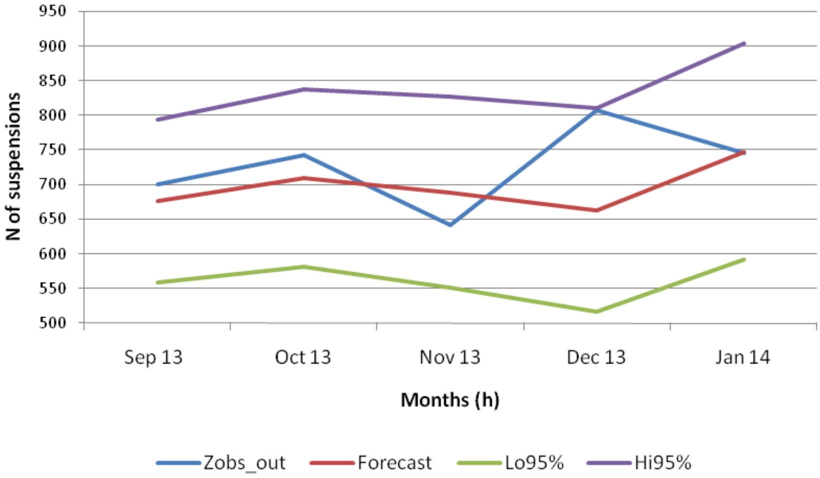
This series is more irregular than the series 1 tends to fluctuate over time and shows more irregular seasonality. The best model according to the Akaike criterion is ETS (A, A, A), which means that error, trend and seasonality are additive. The decomposition of the series is presented in Figure 3. As for the number 1 all the values fall within the prediction intervals except 4 steps forward, that is in the upper limit of the range, Figure 4. The average error for five months ahead out of sample, measured by MAPE is 6.9%.

Figure 3. Decomposition of the number of suspensions sold by a metallurgical company from Alvear (Santa Fe, Argentina). Period January 2009 to August 2013.



Source: Own calculations based on data from a private company.

Figure 4. Forecasts of number of suspensions sold by a metallurgical company from Alvear (Santa Fe, Argentina) (September 2013 - January 2014).



Source: Own calculations based on data from a private company.

Then two series with multiple seasonal (complex) components are analyzed using BATS and TBATS models.

Series **3:** Total number of vehicles per day passing through General Lagos Toll Booth Station of Buenos Aires-Rosario highway in the period from April 22, 2010 to December 31, 2012. The measurements were made using a sensor that counts the number of vehicles.

This series shows annual and weekly seasonality. The annual seasonal period is not an integer (365.25 days). Then the shorter seasonal period: the week (7 days), is not divisor of the larger period. This seasonal behavior is complex and requires unconventional models for the analysis.

There is a missing value in the series due to a flaw in the measuring instrument which is estimated by interpolation. Other flaws in the instrument caused outliers in November 2012.

In order to take into account the behavior of the holidays: a prior correction to the data is done using: 1) a regression model with trend, a unique seasonal component (weekly) and three variables for the holidays (F1: holiday week, F2: last day of long weekend and F3: day before the long weekend) to estimate the effect of these days on the daily number of vehicles passing through the toll (this correction consider the trend and seasonality as deterministic) and 2) the estimated values are corrected using the coefficients obtained for each type of holiday.

The models are fitted using data from April 22, 2010 to October 31, 2012, leaving the November and December to assess the out of sample forecasting performance of the models.

Following the models proposed by De Livera et al (2011) TBATS and BATS models are fitted:

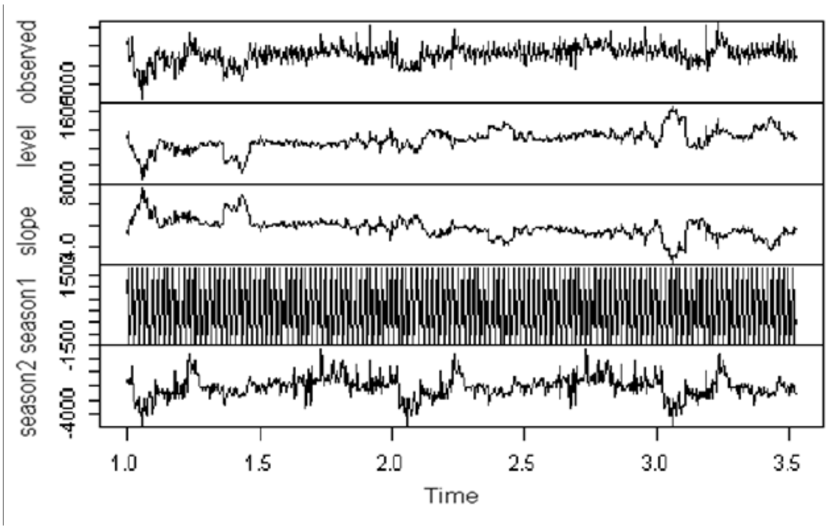
BATS $\omega = 1, \phi = 0.998, p = 3, q = 1, m_1 = 7, m_2 = 365$

and TBATS $\omega = 1, \phi = 0.999, p = 1, q = 2,\ m_1, k_1 \ = \ 7, 3\ ,\ m_2, k_2 \ = \ 365.25, 4$,

$\omega = 1$ means that the series does not need transformation, $\phi \cong 1$ indicates that the damping is very small and the harmonics in the TBATS model reduce significantly the seed values to be estimated.

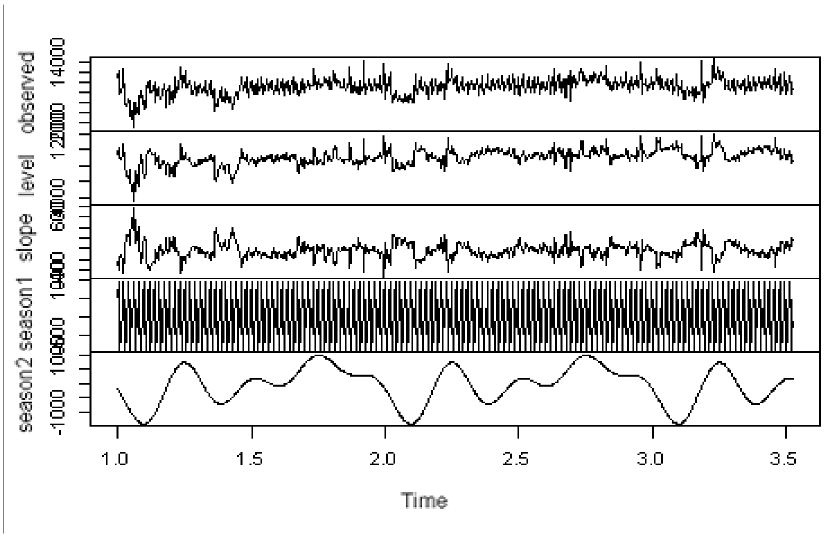
The decomposition of the time series of daily number of vehicles passing through the toll booth General Lagos by both methods shows a slightly increased level a slightly declining growth rate and two seasonal patterns weekly and annual.

Figure 5. Decomposition of number of daily vehicles passing through the toll booth General Lagos by BATS model. (April 22, 2010 to October 31, 2012).



Source: Author's calculations based on data from OCCOVI.

Figure 6. Decomposition of the number of daily vehicles passing through the toll booth General Lagos by TBATS model (April 22, 2010 to October 31, 2012)..

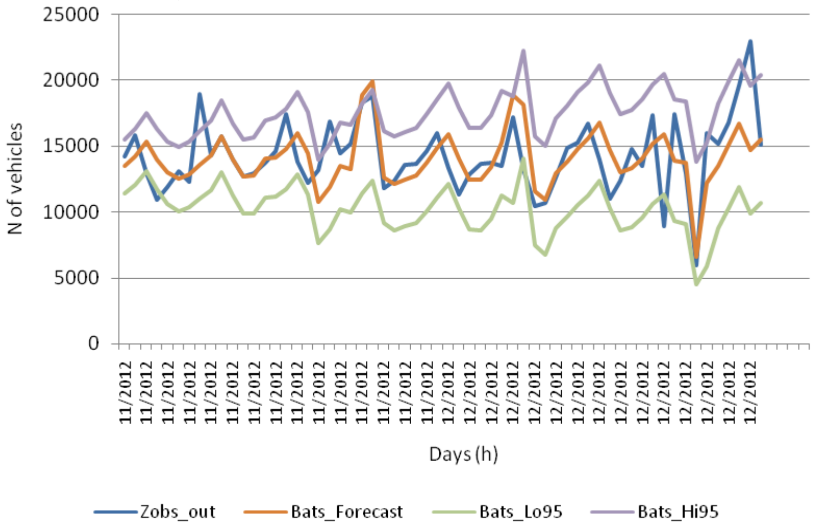


Source: Author's calculations based on data from OCCOVI

The BATS model estimates the components with more irregularity than the TBATS model.

As mentioned above months of November and December of 2012 are left to evaluate the out-of-sample forecasts of the two models. The forecast for the days: 11/09/2012 and 12-11-2012 were not taken into account because they can be considered outliers due to a flaw in the sensor that measures the passage of vehicles. The forecasts were corrected using the coefficients obtained for the holidays.

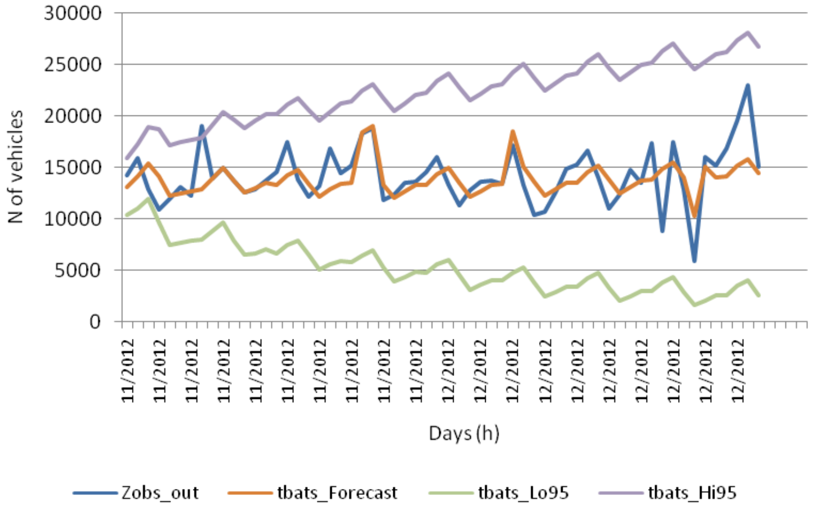
Figure 7. Forecasts of the number of vehicles per day passing through the toll booth General Lagos with their respective 95% confidence intervals for the model BATS (09.11.2012 to 12-11-2012)



Source: Author's calculations based on data from OCCOVI.

Most of the observed values fall within the 95% of the forecast intervals. Over the last days of December, the variability of the observed values increased considerably and some of them fall outside the forecast interval.

Figure 8. Predictions of the number of vehicles per day passing through the toll booth General Lagos with their 95% confidence intervals for the model TBATS (09.11.2012 to 12-11-2012)



Source: Author's calculations based on data from OCCOVI.

Both models have a good performance, presenting the TBATS model some advantages over the BATS model such as: i) lower AIC (AIC = 19742.28 $_{TBATS,}$ AIC = 19885.2 $_{BATS),}$ ii) significantly less computation time given the estimation of less seeds values and iii) lower error (MAPE (30 days) $_{TBATS}$ and MAPE = 9.64% (30 days) = 10.66% $_{BATS\ BATS}$ and MAPE (61 days) = 13.00% and MAPE $_{TBATS}$ (61 days) = 12.79%).

Table 1. MAPE of BATS and TBATS models for different horizons (h).

Days (h)	BATS	TBATS
7	11.58	11.13
15	10.53	10.04
30	10.66	9.64
45	10.91	9.95
61	13.00	12.79

Series **4:** Las Rosas (Santa Fe, Argentina) daily average consumption of gas in cubic meters (m^3) during the period March 1, 2008 to August 31, 2011. To fit the model the two last months are left (July-August) to assess the out-of-sample goodness of forecasts.

This series also has two types of seasonality, weekly (7-day) and annual (365.25), so it can be considered as complex seasonality. Another feature of this series is that it is more volatile in winter than in summer due to the use of gas for heating.

This series was previously studied by Acosta, P. (2013), using a state-space model (Harvey, Koopman, 1993), with *spline*. This model considers the annual seasonality, and includes explanatory variables as temperature and holidays. The series presented outliers that were took into account for the analysis. The forecasts made were very good, with a 31-day MAPE 5.25% and 5.83% for 62 days.

In the present application the BATS and TBATS models do not include explanatory variables. A possible loss of accuracy due to the use of simpler models is evaluated.

In order to take into account the particular behavior of the holidays, as for Series 3, a prior correction of the data is done using: 1) a regression model with trend, a unique seasonal component (weekly) and a variable for public holidays, to estimate the effect of these days on the average consumption of gas (this correction considers the trend and seasonality as deterministic) and 2) the observed values are corrected using the estimated coefficients for the public holidays.

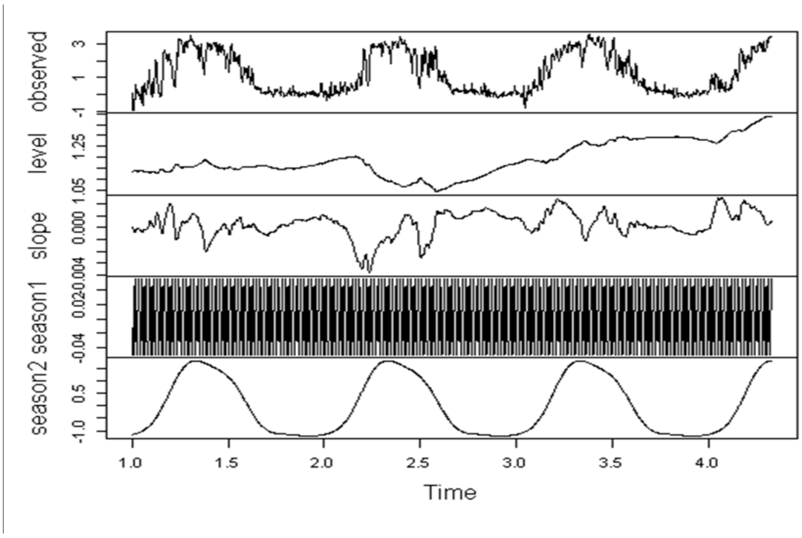
Following the models proposed by De Livera et al (2011) presented in section I model a TBATS[2] is fitted to the data. The estimated model is:

TBATS $\omega = 1, \hat{\phi} = 0.971, p = 4, q = 5, \ m_1, k_1 \ = \ 7,1 \ , \ m_2, k_2 \ = \ 365.25, 4$,

A transformation is not needed ($\omega = 1$), $\phi \cong 1$ indicates that the damping is slow and harmonics in the TBATS model reduce significantly the number of seed values to be estimated.

The shows an increasing level, a growth rate of almost zero and two seasonal patterns weekly and annual.

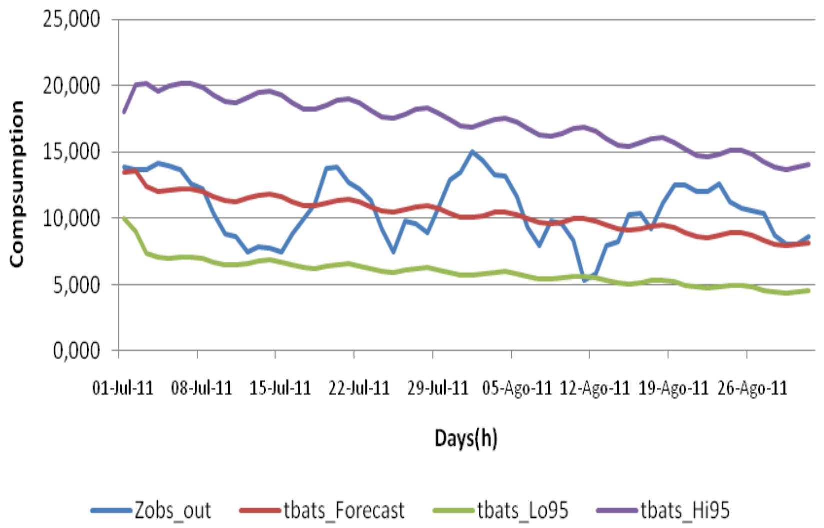
Figure 9. Decomposition of Las Rosas (Santa Fe, Argentina) daily average consumption of gas in cubic meters (m^3) for the model TBATS



Source: Own calculations based on data from a private company.

Figure 10. Predictions of Las Rosas (Santa Fe, Argentina) daily average consumption of gas in cubic meters (m^3) with their respective 95% confidence intervals for the model TBATS (November 9, 2012 to November 12, 2012).

[2] BATS model is not presented because a proper fit is not achieved.



Source: Own calculations based on data from a private company.

Most of the observed values fall within the 95% forecast intervals.

Table 2. MAPE of the TBATS model for different horizons (h).

Days (h)	TBATS
7	7. 6
15	22. 8
30	18. 8
45	21 1.
62	19. 7

If the values in Table 2 are compared with that by Acosta, P. (2013), the value obtained here for 30 days is three times the other and the value obtained for 62 days is almost four times the one obtained by Acosta P., showing the superiority to forecast of the more complete model.

IV. Concluding Remarks

This paper discusses the benefits of forecasting using state space innovations models, for series with a single seasonal period and for a complex seasonality. TBATS and BATS models are used with Argentinean time series.

The two time series of sales (axles and suspensions) of a metalworking firm in Argentina, show a good performance, as it was expected for this type of series The average percentage of error for forecasts with a horizon of five months ahead out of sample, is 9.4% and 6.9% respectively, with a controllable level of uncertainty.

These results are consistent with those reached by the authors, who recommend this type of models for utility demand series.

TBATS and BATS models are used for series with complex seasonal patterns: The total number of vehicles per day passing through the toll booth station of General Lagos Buenos Aires-Rosario highway. Results show that both types of models are suitable for describing and predicting this series. The TBATS model has some advantages over the

BATS model such as: i) better goodness of fit (lower AIC), ii) lower percentage of error in their out of sample forecasts for different forecast horizons (measured by MAPE); reduction in computation time to estimate the model, due to the lower number of seeds values.

In series 4, Las Rosas (Santa Fe, Argentina) daily average consumption of gas in cubic meters (m^3) during the period March 1, 2008 to August 31, 2011, however, the performance of the proposed models is not as good, the BATS model did not provide good fit and in the case of the TBATS model, although it fits the data well, the forecasts have more error than the ones obtained using a SSM with *Spline.* One possible explanation for the lower quality of forecasts by TBATS model is that no explanatory variables were included in these models and in this application climate variables have much influence. Climate variables were included if in SSM approach. However, given the simplicity of the model for use TBATS causes cannot be completely discarded.

Future research will continue investigating the validity of the models, both, theoretically and empirically, to improve the quality of forecasts made in highly fluctuating series.

IV. References

ACOSTA, P. (2013) *Time series models for daily consumption of gas.* Master Thesis in Applied Statistics, Faculty of Economics and Statistics Sciences, National University of Rosario.

BOX, GEP, and JENKINS, GM (1970) *Time Series Analysis: forecasting and control* (1 st ed) San Francisco. Holden-Day.

DE LIVERA, AM, Hyndman, RJ, and Snyder, RD (2011) Forecasting time Series with Complex Smoothing exponential Using seasonal patterns. *Journal of the American Statistical Association.* **106.** From 1513 to 1527.

HARVEY, A. (1989) *Forecasting Structural Time Series Model and the Kalman Filter,* New York: CambridgeUniversity press.

HARVEY, A. AND KOOPMAN, SJ (1993) Forecasting Hourly Electricity Demand Using Tie Varying Splines. *Journal of the American Statistical Association.* **88** 1228-1236.

HYNDMAN, RJ, KOEHLER, AB, ORD, JK and SNYDER, RD (2008) *Forecasting with Exponential Smoothing: The State Space Approach,* Berlin, Springer-Verlag.

HYNDMAN, RJ and KANDHAKAR, Y. (2008) Automatic Time Series Forecasting: the Forecast Package for R. *Journal of Statistical Software,* **26** (3) 1-22.

R DEVELOPMENT CORE TEAM (2011). R: A language and environment for statistical Foundation for Statistical Computing computing.R, Vienna, Austria.ISBN 3-900051-07-0, URL http://www.R-project.org/ .

TAYLOR, JW (2003). Short-Term Electricity Demand Forecasting *Journal of the Operational Research Society,* **54,** 799-805.

TAYLOR, JW (2010) Seasonal Triple Methods for Short-Term Electricity Demand Forecasting. *European Journal of Operational Research,* **204,** 139-152.

TAYLOR, JW and Snyder, RD (2009) Forecasting Intraday Time Series With Seasonal Cycles Multiple Seasonal Exponential Smoothing Using Parsimonious. Technical Report 09/09. Dept. of Econometric and Business Statistics, Monash University.

WINTERS, PR (1960) Forecasting sales by exponentially weighted moving averages, *Management Science,* **6,** 324-342.

WEST, M. and HARRISON, J. (1997) *Bayesian Forecasting and Dynamic Models* (2nd ed.) New York: Springer-Verlag.