Revista Gestão Industrial
ISSN: 1808-0448
http://periodicos.utfpr.edu.br/revistagi

Predicting the GFCF of the Brazilian Construction Industry: A Comparison Between Holt Winters' and SARIMA Models

Paulo Siga Thomaz
paulosigathomaz@gmail.com
FURG, Rio Grande, RS, Brasil

Viviane Leite Dias de Mattos
viviane.leite.mattos@gmail.com
FURG, Rio Grande, RS, Brasil

ABSTRACT

The present study focuses on creating a forecasting model in order to predict the behavior of the economic indicator known as Gross Fixed Capital Formation (GFCF) of the Brazilian construction industry, which reflects the amount of investment in the construction industry sector. The data set consists of monthly observations from January 1996 to December 2016, the year of 2016 is used as validation for the forecast model. As strong seasonality was identified in the time series, Seasonal Autoregressive Integrated Moving Average (SARIMA) and Holt Winters' models are applied and the results are compared. After the evaluation of the selected models, the ARIMA $(2,1,2) \times (0,1,1)_{12}$ is identified as the best forecast model with reasonable deviations. However, the damped multiplicative Holt Winters' model also produces good results, despite its inability in eliminating the autocorrelation in the residuals. Therefore, both models can predict the GFCF with good accuracy, which can be useful for decision-making by investors and business managers.

KEYWORDS: SARIMA. Box-Jenkins. Forecasting. Holt Winters. GFCF.

INTRODUCTION

The construction industry sector has strong influence in country's economic policy, mainly due to its complex productive chain, ranging from mining activity to the generation of infrastructure, as highroads, railroads, bridges, harbors, transportation systems, buildings, among others. Bon (1992) points out that due to the process of industrialization and urbanization in developing countries, the economic growth leads to an increase in the participation of the construction industry in the total income produced. It reflects the extensive investment in infrastructure during early stages of country's development. Since Brazil is a developing country, infrastructure is the center of the policy agenda, which can be evidenced, according to Amman et. al. (2016), by the creation of the Growth Acceleration Program (PAC, in Portuguese) in 2007, which raised the investments in the construction industry sector exponentially.

The Gross Fixed Capital Formation (GFCF) of the construction industry is an economic indicator that reflects the amount of investment in the sector. Predicting this indicator can be useful in order to provide us a picture of the Brazilian government economic policy for the next years, helping with the budget allocation planning of companies and people in general, besides it can supply a general view on the economic stability and growth for the coming years.

There are several prediction techniques presented in literature, some of them are suitable for specific situations, while others can be applied in an extensive set of time series. Holt (1957), Brown (1959) and Winters (1960) introduced the exponential smoothing methods, which consider the weighted average of past observations to compute fitted and predicted values of the model. Usually, these models are presented in three different variations: simple exponential smoothing, trend-corrected exponential smoothing (or linear Holt method) and Holt-Winters' method (BILLAH et al., 2006). Hyndman & Athanasopoulos (2013) proposed a range of modifications in the components (level, trend and seasonality) of the previous models, improving their performance in certain situations, these models are called Innovation State Space models. Holt Winters' method is a special case of these models, it takes into account the seasonality present in the time series, which can be considered as additive or multiplicative, and considers the trend as linear.

An alternative for the exponential smoothing techniques is the Box-Jenkins methodology, represented by the Autoregressive Integrated Moving Average (ARIMA) models (BOX & JENKINS, 1976). These models are extensively used, since they are able to produce reasonable results for time series with different configurations, they are also simple to use and are included in most of the econometric and statistics software. If seasonality is included into the model, it can be called Seasonal ARIMA, or simply SARIMA.

The goal of this paper is to create a forecasting model in order to predict the GFCF of the Brazilian construction industry. For this purpose, several SARIMA models and variations of the Holt Winters' method are applied and prediction results are compared.

THE CONSTRUCTION INDUSTRY IN THE BRAZILIAN ECONOMIC SCENARIO

According to Bon (1992), in the first stages of a country's development, the civil construction is the sector that contributes more to the Gross Domestic Product (GDP) growth, mainly due to the processes of urbanization and industrialization, which put the whole productive chain of the construction sector in operation. Whereas, in more advanced stages of development, the increase in per capita income leads to an increase in service activity, making this sector more relevant to GDP growth than the construction industry.

As Brazil is still a developing country, the civil construction sector is constantly growing, and can be considered as a thermometer of the national economic activity. This is confirmed by some data presented by Passos et al. (2012), who affirm that since 1975 the contribution of construction industry to GDP has been increasing, while the manufacturing industry contribution has declined. Between 1975 and 1985, for example, the industrial sector's contribution to GDP was around 40%, while construction contributed with 6.5%. In 1990, the construction industry contributed with 6.9%, whilst the industrial sector's contribution to GDP fell to 34.3%. The contribution of civil construction to the economy increased even more in 2001, reaching 15.6% of the national GDP.

When it comes to investment, civil construction also shows great relevance. This is evidenced by the creation of programs such as PAC (Growth Acceleration Program), which have stimulated even more the growth of the construction sector through investment in residences, highways and other infrastructure works. The amount of money invested in the construction of habitations, between the years of 2011 and 2013, for example, reached R$ 773,4 billion (FIALHO ET AL., 2014). Farias (2015) also shows the importance of these programs, stating that they were responsible for an increase of 25.8% in the total produced by the national economy in the period between 2007 and 2013.

Another important factor to consider is the strong bond of civil construction with other sectors, since, through the generation of infrastructure, the construction industry contributes to the development of all branches of the economy. Teixeira & Carvalho (2005), for example, say that in 2009, despite the international crisis, formal employment in the construction sector increased by 9.17%. The authors also say that the PAC was responsible for an average growth of 3.17% in the total number of employment in the country, from 2011 to 2013.

The data presented show that the construction industry is a key sector in Brazil’s development, being a priority in the investment strategies of the government. Therefore, the analysis of the economic indicators of this sector can provide important subsidies for decision-making by investors and administrators.

THEORETICAL BASIS

LITERATURE REVIEW

Holt Winters' models are widely used in literature. Tratar & Strmcnik (2016) applied this technique in order to predict heat load with daily, weekly and monthly data sets. The authors found out that Holt Winters' multiplicative method performed better than multiple regression for the monthly data. In another study, Grubb & Mason (2001) used a Holt Winters' model with damped trend to forecast the number of air passengers in United Kingdom. The results of the study showed that the forecast was improved due to the modification in the trend. In the financial area, Brazenas & Valakevicius (2015) applied different Holt Winters approaches in order to predict USD/EUR exchange rate pair, in which multiplicative Holt Winter presented the best results. These studies show that Holt Winters' models can be a good option for situations where seasonality is present.

There are also numerous examples of SARIMA applications in several different areas of research. Taneja et al. (2016) used SARIMA models in order to predict the aerosol optical depth (AOD) over New Delhi, India. The models were unable to simulate extreme values, however they can be used do predict AOD with good accuracy. In the health area, Martinez et al. (2011) created a forecast model in order to predict the monthly incidence of dengue fever in the city of São Paulo, Brazil. As the monthly incidence of dengue fever in a population tends to be subject to seasonal variations, SARIMA models were adequate and could produce good results. Tsui et. al. (2013) extended SARIMA models applications to the multivariate case. The authors applied a SARIMA model with exogenous variables (SARIMAX) in order to predict the monthly passenger's traffic demand in Hong Kong. As a result of the study, the authors stated that the forecasting results were satisfactorily accurate.

Taking into account all the studies previously presented, it is possible to notice that both Holt Winters and Box-Jenkins methodologies are largely used in different situations. However, it is difficult to define in advance which model is more appropriate for the time series of interest. For example, in a study conducted by Milenković et al. (2015), SARIMA models presented better results than exponential smoothing techniques in the prediction of railway passenger flow. However, Puthran et al. (2014), in a study about the Indian motorcycle industry, obtained better results for the Holt Winters' method. Lepojević & Anđelković-Pešić (2011) and Veiga et al. (2014) also carried out similar studies and came to different conclusions. Therefore, there is no consensus regarding which model is better, it depends entirely on the time series under study.

HOLT WINTERS' METHOD

Holt Winters' method is composed of three smoothing equations, one for level, one for trend and one for seasonality. The latter can be considered additive or multiplicative. A Holt Winters' model with additive seasonality is described by Equations (1-4).

$l_t = \alpha(y_t - S_{t-s}) + (1-\alpha)(l_{t-1} + b_{t-1})$
(1) $b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$
(2) $S_t = \gamma(y_t - l_{t-1}) + (1-\gamma)S_{t-s}$

(3) $F_{t+m} = l_t + b_{t-m} + S_{t-s-m}$

(4) where l_t, b_t, S_t and y_t are respectively the level, trend and seasonal components; y_t is the observed value at time *t*; *s* is the seasonality period, which can be considered as 12 for monthly data; α, β and γ are the weighting parameters with values ranging from 0 to 1; F_{t+m} is the forecast value at *m* steps ahead.

The multiplicative Holt Winters' model can be written similarly, with some modifications to the components. The trend equation remains unchanged, since it is linear in both models and does not include seasonality. Level, seasonality and forecast equations for the multiplicative case are shown in Equations (5-7).

$$l_t = \alpha(y_t/S_{t-s}) + (1-\alpha)(l_{t-1} + b_{t-1})$$ (5)

$$S_t = \gamma(y_t/l_{t-1}) + (1-\gamma)S_{t-s}$$ (6)

$$F_{t+m} = (l_t + b_{t-m})S_{t-s-m}$$ (7)

The additive method is more suitable for situations where seasonal variations are constant through the time series, whereas the multiplicative case is preferred when seasonality varies proportionally to the time series level (HYNDMAN & ATHANASOPOULOS, 2013). However, both models consider the linear trend increasing or decreasing indefinitely towards the future, which can be a problem, since not all trended time series have this kind of behavior. Gardner & McKenzie (1989) added an extra parameter to the Holt Winters' method in order to damp the trend and fix this problem, originating the Holt Winters' models with damped trend. The components for the additive case are shown in Equations (8-10).

$$l_t = \alpha(y_t/S_{t-s}) + (1-\alpha)(l_{t-1} + \phi b_{t-1})$$ (8)

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)\phi b_{t-1}$$ (9)

$$F_{t+m} = l_t + \phi b_{t-m} + S_{t-s-m}$$ (10)

where 0 < ϕ < 1 is the damping parameter. Note that when ϕ = 1, the model is identical to the regular additive Holt Winters' model. The seasonality equation does not change, since it does not include a trend component. Analogously, the multiplicative Holt Winters' model with damped trend is given by Equations (11) and (12).

$$l_t = \alpha(y_t/S_{t-s}) + (1-\alpha)(l_{t-1} + \phi b_{t-1})$$ (11)

$$F_{t+m} = (l_t + \phi b_{t-m})S_{t-s-m}$$ (12)

It is desirable that the Holt Winters' models produce residuals without autocorrelation, since its presence means that there are information among the observations the model is not capturing (HYNDMAN & ATHANASOPOULOS, 2013).

SARIMA MODELS

The general formula for an Autoregressive Moving Average (ARMA) model for a stationary time series is represented by Equation (13).

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=0}^{q} \theta_j \varepsilon_{t-j} \tag{13}$$

where y_t is the observed value at time t; *c* is a constant; *i* and *j* are the autoregressive and moving average process lags respectively; ϕ_i is the autoregressive coefficient at lag *i*; θ_j is the moving average coefficient at lag *j*; *p* and *q* are the autoregressive and moving average orders respectively.

If the time series is non-stationary, it is possible to make it stationary by the process of differencing, which consists in compute the difference between consecutive observations until the time series reaches stationarity or, in other words, mean and variance constant. The combination of the autoregressive and moving average terms and differencing results in an ARIMA (*p, d, q*) model, where *d* denotes the order of the differencing process.

Seasonality can be defined as a pattern repeating itself over a fixed time interval (TANEJA ET AL., 2016). In the case of seasonality present in the time series, it is necessary to include seasonal terms into the ARIMA model. Taking this into consideration, a multiplicative ARIMA denoted as ARIMA $(p,d,q)\times(P,D,Q)_{12}$ can be used, where *s* is the lag of seasonality and *P*, *D*, and *Q* are the orders of the seasonal part. The ARIMA with seasonality can be called SARIMA and its general equation can be written as shown in Equation (14) (MARTINEZ ET AL., 2011).

$$\Phi(L^s)\phi(L)(1-L)^d(1-L^s)^D y_t = \Theta(L^s)\theta(L)\varepsilon_t \tag{14}$$

where *L* is the lag operator; $\phi(L)$ is the autoregressive polynomial function of order *p*; $\theta(L)$ is the moving average polynomial function of order *q*; $\Phi(L^s)$ is the seasonal autoregressive polynomial function of order *P*; $\Theta(L^s)$ is the seasonal moving average polynomial function of order *Q*; *d* is the differencing order; *D* is the seasonal differencing order and ε_t is the error term.

In order to the SARIMA model be valid, its residuals must be normally distributed, not autocorrelated and homoscedastic.

METHODOLOGY

The data used in this study were obtained from the Institute of Applied Economic Research (IPEA) database (IPEA, 2017). The Brazilian construction industry GFCF consists of monthly observations from January 1996 to December 2016, the first 19 years (228 observations) were used as training set, whilst the year of 2016 (12 observations) was used as validation for the forecast models.

In order to apply the Holt Winters' models, it was necessary to define the initial values l_0, b_0, and S_0, which was done by maximizing the likelihood function of the models. The parameters α, β and γ were chosen using the same procedure. The goodness of fit in the training sample was measured by the Root

Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). In addition, the Ljung-Box test (LB) was applied to the residuals.

For the application of the SARIMA models, the first step was to identify if GFCF time series is stationary. In order to do this the Augmented Dickey-Fuller (ADF) test was used. ADF examines the existence of unit root through the estimation of a linear model with the first difference value as the response variable, and the value of the time series at time t - 1 and the previous first difference values as input variables (TRATAR & STRMCNIK, 2016). The hypothesis is then tested where the presence of unit root is considered as null hypothesis and the absence of it as alternative hypothesis. However, as asserted by Ng & Perron (2001), ADF test is very sensitive depending on the lag length chosen. Therefore, the lag length was determined using the following technique proposed by Bueno (2008): first a maximum lag length was fixed by the Schwert criterion (SCHWERT, 1989) $p_{max} = 12int[\left(\frac{T}{100}\right)^{\frac{1}{4}}]$, where T is the number of observations; from this maximum lag, the test was performed and the significance of the coefficient corresponding to the last term of the regression model was verified. If this coefficient was significant the result of the test was taken into account, otherwise the test was executed again for $p_{max} - 1$, this process was repeated until the last coefficient be significant. As seasonality was taken into account, it was necessary to apply the seasonal difference to the original time series, which consists in the difference between an observation and the corresponding observation of the previous year (HYNDMAN & ATHANASOPOULOS, 2013). Then, the ADF test was applied again in order to check if the seasonal component is stationary.

Continuing the implementation of the SARIMA models, the orders *p*, *q*, *P* and *Q* were defined. This process was done iteratively, checking the significance of the coefficients and analyzing the patterns of the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. The former takes into account the implicit relationship between observations, whereas the latter considers this relationship individually (BUENO, 2008). In this study, the coefficients of the models were computed through the maximum likelihood estimation method.

Finally, with the SARIMA models defined it was necessary to perform the residual diagnostics, in other words, to check the normality, autocorrelation and heteroscedasticity in the residuals. For checking the normality, the Jarque-Bera test was performed, which statistically tests if the asymmetry is equal to 0 and if the kurtosis is equal to 3 (null hypothesis) (MORETTIN & TOLOI, 2005). The normality of the residuals is important, since the maximum likelihood estimation requires normality, otherwise the estimation can be inconsistent. In order to check the autocorrelation in the residuals, the Ljung-Box test was applied. Finally, the ARCH-LM (ALM) test was executed for checking if the ARCH effect was present, which is considered as alternative hypothesis. If the residuals are heteroscedastic, it can mean that an ARCH or GARCH model is more adequate for the time series under analysis. The three tests, Jarque-Bera, Ljung-Box and ARCH-LM, use the χ^2 distribution for testing the hypothesis.

As an additional step, the information criterion of Akkaike (AIC) and Hannan-Quinn (HQIC), as well as the error measures RMSE and MAPE, were used to identify the best fitted SARIMA model. The information criterion is a way to find the ideal number of parameters in the model, it associates a penalty corresponding to the addition of a regressor, if the decrease in the sum of the residuals is greater than the penalty, the regressor should be added into the model (YIP ET AL., 2013).

The SARIMA and Holt Winters' models with the best performance in the training sample were then applied to predict the Brazilian construction industry GFCF. The best forecast model was defined through the error measures MAPE and RMSE, as well as using the U's Theil parameter. The latter measures the relative performance of the forecast in relation to the naive prediction method.

All the analysis in this study was executed through R and GRETL (Gnu Regression, Econometric and Time-series Library) software.

RESULTS AND DISCUSSION

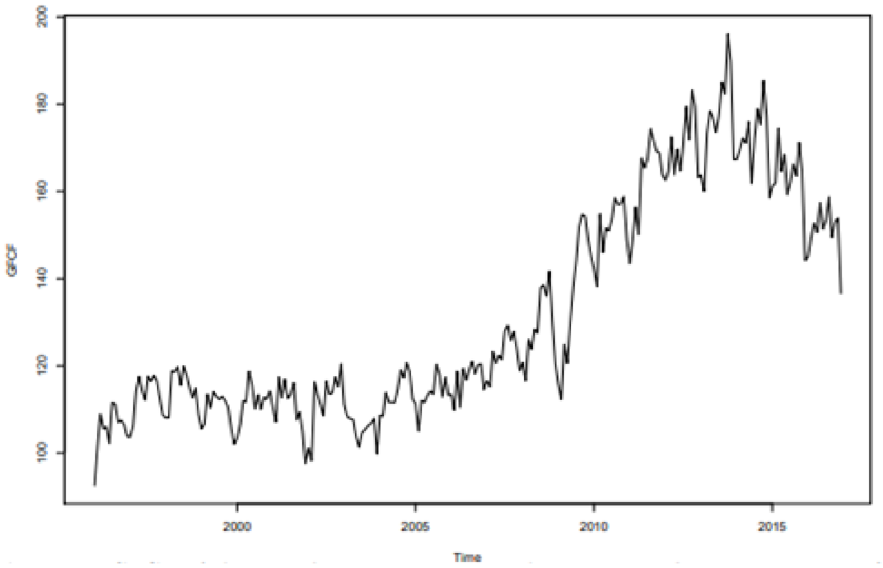
DATA

The behavior of the time series can be seen in Figure 1, in which the y-axis represents the GFCF of the construction industry sector and the x-axis represents the time in years. It is possible to observe that there are two distinct moments in the time series. From January 1996 to December 2006, the series has an apparent stationary trend, since at first sight the mean seems to be constant in this period. From January 2007 onward, the GFCF grows almost exponentially until November 2013. It is a reflection of the economic policy adopted by the Brazilian government as from January 2007, in this month the Government created the Growth Acceleration Program (known as PAC) in order to stimulate the economic activity of the construction industry sector. The downfall of investment from November 2013 until nowadays is a consequence of the economic slowdown that the country has been experiencing.

In addition, it is evident in Figure 1 that the time series has strong seasonality. In almost every year, there is a peak in the months of January and February, and then the time series falls again and goes up one more time in the months of July and August, reaching its yearly peak one last time and then going down again from September to December.

R. Gest. Industr., Ponta Grossa, v. 15, n. 1, p. 180-196, jan./mar. 2019.

Figure 1 – GFCF of Brazilian construction industry sector, between January of 1996 and December of 2016.



Source: IPEADATABASE (2017)

HOLT WINTERS MODELLING

In this study, four variations of Holt Winters' models were considered for fitting the time series: Additive Holt Winters, Multiplicative Holt Winters, Damped Additive Holt Winters and Damped Multiplicative Holt Winters. The goodness of fit in the training sample is shown in Table 1, as well as the result of the Ljung-Box test. It is possible to notice that the Damped Multiplicative Holt Winters has the smallest errors among the models. In addition, it seems that the damped trend enhances the performance of the multiplicative model, whereas it does little effect for the additive case. However, all models have autocorrelated residuals, which means they are not able to capture all available information.

Table 1 - Performance in the training sample and residual diagnostics for the Holt Winters' models

Models	RMSE	MAPE	LB (p-value)
AHW	4,2150	2,4057	<0,0001
MHW	4,1700	2,4760	<0,0001
DAHW	4,2055	2,3972	<0,0001
DMHW	3,9718	2,3294	<0,0001

OBS: AHW – Additive Holt Winters; MHW – Multiplicative Holt Winters; DAHW – Damped Additive Holt Winters; DMHW – Damped Multiplicative Holt Winters; LB – Ljung-Box.
Source: Own Authorship (2018)

SARIMA MODELLING

This study applies the ADF test for checking the stationarity. The test indicates that the original time series is non-stationary, so the first difference is taken. As the time series has a strong seasonal pattern, the seasonal difference is also taken and is evaluated as non-stationary, and then an additional difference is performed. Therefore, it is possible to say that the differencing orders are $d = 1$ and $D = 1$. The results of the ADF test are summarized in Table 2, where t is the test statistics and p-value is the probability of significance.

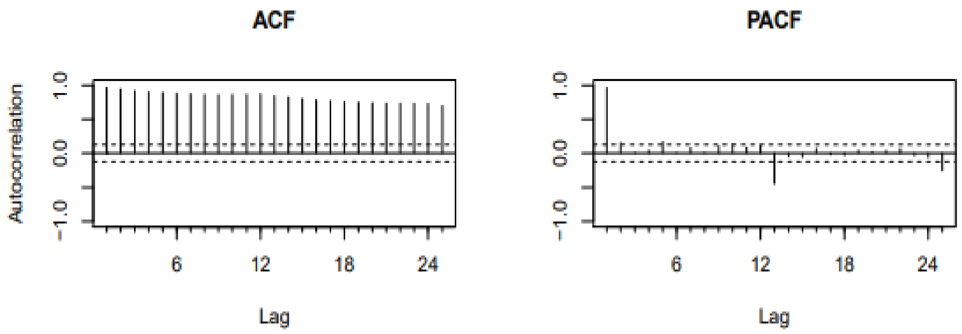
Table 2 – Augmented Dickey-Fuller test results

Models	t	p-value
Original	-1,5459	0,8139
First Difference	-4,5621	0,0002
Seasonal Difference	-1,3900	0,1533
First difference and seasonal difference	-8,1571	< 0,0001

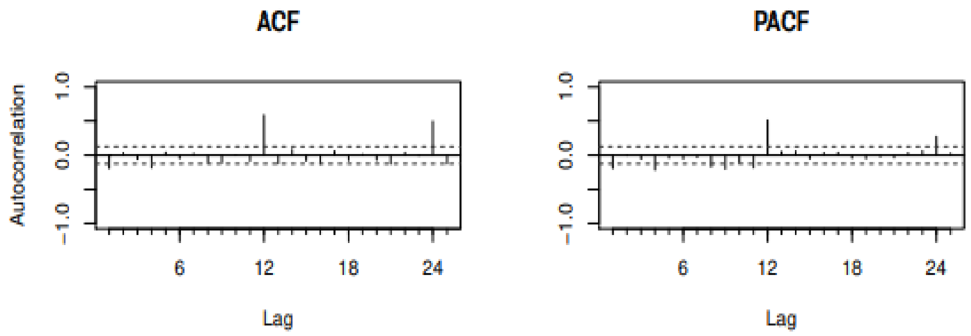
Source: Own Authorship (2018)

Figure 2 (a-d) shows the plots of the autocorrelation functions with different orders of differencing, the y-axis represents the values of autocorrelation, whereas the x-axis is the lag length. The dashed lines indicate the 95% confidence interval calculated by the formula $1{,}96/T^{0,5}$, where T is the number of observations. It is possible to observe in Figure 2 (a, c) that the decay of the ACF function is slow, which indicates the non-stationarity of the original and seasonal differenced time series. In Figure 2 (b, d) it is possible to notice, in both functions, two significant spikes at lag 1 and 12, which indicates the presence of seasonality. In addition, the functions decay quickly toward zero, being an indicative of stationarity.

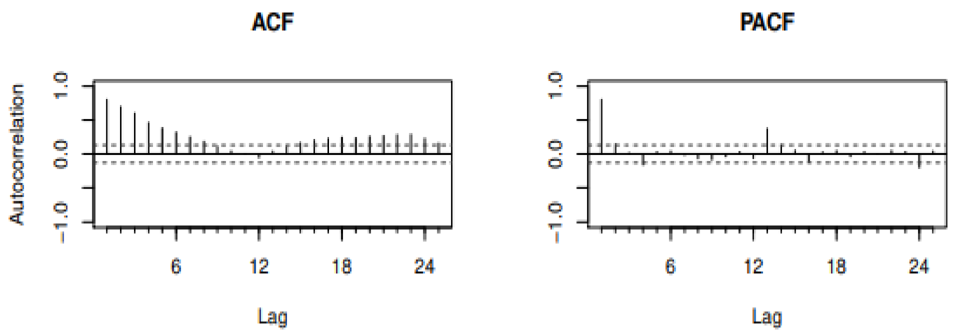
Figure 2 - ACF and PCF for original time series (a), first difference only (b), seasonal difference only (c), first difference and seasonal difference (d).



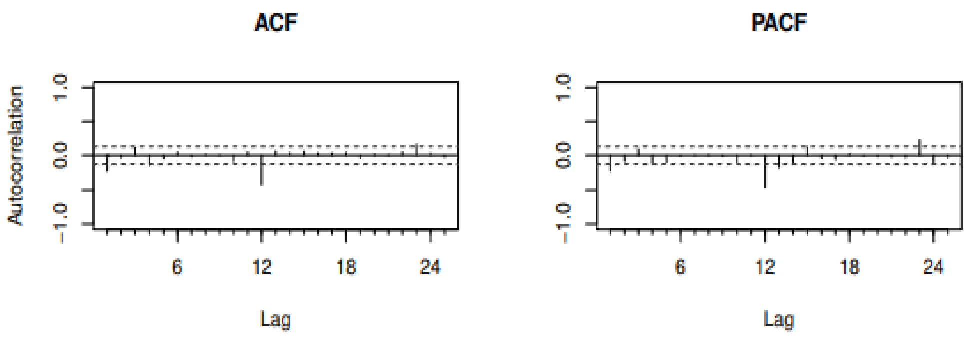
(a)



(b)



(c)



(d)

Source: Own authorship (2018)

The performance of the SARIMA models in the training sample is shown in Table 3 for the six best models obtained in this study, whereas the results of the Jarque-Bera (JB), Ljung-Box (LB) and ARCH-LM (ALM) tests are presented in Table 4. The results show that ARIMA $(2,1,2)\times(0,1,1)_{12}$ is the model with less deviations and lowest information criteria, so it is possible to state that it is the best SARIMA model in this study. In addition, all models present residuals normally distributed, non-autocorrelated and homoscedastic at 1% significance level.

Table 3 - Performance in the training sample for the SARIMA models

Models	RMSE	MAPE	AIC	HQIC
ARIMA(2,1,1)×(1,1,1)$_{12}$	3,9537	2,2474	1305,76	1326,31
ARIMA(2,1,1)×(0,1,1)$_{12}$	4,0070	2,6683	1303,78	1330,64
ARIMA(2,1,2)×(0,1,1)$_{12}$	3,9537	2,2474	1299,91	1320,46
ARIMA(2,1,1)×(2,1,1)$_{12}$	3,9957	2,2549	1306,66	1330,64
ARIMA(2,1,2)×(2,1,1)$_{12}$	3,9456	2,2437	1302,93	1330,33
ARIMA(2,1,2)×(2,1,2)$_{12}$	3,9356	2,2431	1303,90	1334,73

Source: Own authorship (2018)

Table 4 – Residual diagnostics for the SARIMA models

Models	JB (p-value)	LB (p-value)	ALM (p-value)
ARIMA(2,1,1)×(1,1,1)$_{12}$	0,0120	0,0254	0,0390
ARIMA(2,1,1)×(0,1,1)$_{12}$	0,0135	0,0349	0,0219
ARIMA(2,1,2)×(0,1,1)$_{12}$	0,0195	0,2142	0,0264
ARIMA(2,1,1)×(2,1,1)$_{12}$	0,0115	0,0263	0,0122
ARIMA(2,1,2)×(2,1,1)$_{12}$	0,0200	0,1822	0,0517
ARIMA(2,1,2)×(2,1,2)$_{12}$	0,0160	0,1714	0,0510

Source: Own authorship (2018)

COMPARISON AND PREDICTION RESULTS

The best SARIMA and Holt Winters' models are compared, taking into account their performance in the training and validation data sets. The results are summarized in Table 5. Since both models have U's Theil values less than 1, their use as forecasting methods are justified. In addition, it is possible to observe that the results for the SARIMA models are better in both training and validation samples. However, despite the Damped Multiplicative Holt Winters' model having autocorrelated residuals, it also produces good results.

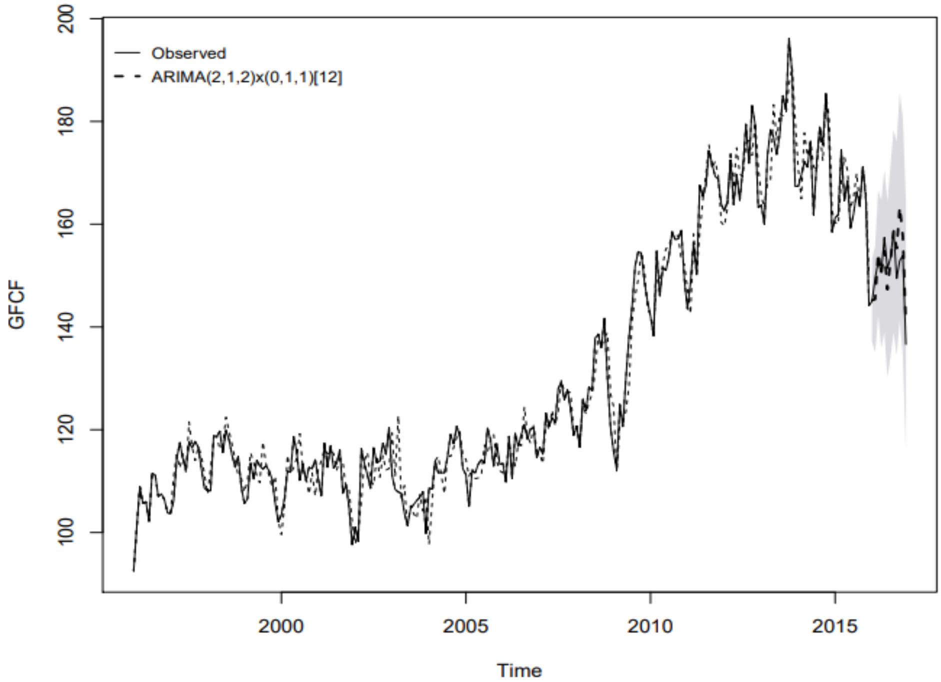
Table 5 - Training sample and validation sample performance for the SARIMA and Holt Winters' models

Models	Training Sample		Validation Sample		
	RMSE	MAPE	RMSE	MAPE	U's THEIL
ARIMA(2,1,2)×(0,1,1)$_{12}$	3,9718	2,3972	4,5898	2,4195	0,6815
DMHW	3,9537	2,2474	4,2644	2,0974	0,6506

OBS: DHMW – Damped Multiplicative Holt Winters
Source: Own Authorship (2018)

Since the ARIMA(2,1,2)×(0,1,1)$_{12}$ model presented the best results and its residuals fulfilled the requirements of normality and non-autocorrelation, this model is chosen to forecast the GFCF of the Brazilian construction industry. Figure 3 shows the comparison between the SARIMA model and the original time series. The forecasting confidence interval of 95% is used and is represented by the gray area. Note that all predicted values are inside the confidence interval, therefore the short-term forecast results have good accuracy.

Figure 3 - Observed values and SARIMA$(2,1,2)\times(0,1,1)_{12}$ model



Source: Own authorship (2018)

FINAL CONSIDERATIONS

This study aimed to create a forecast model in order to predict the GFCF of the Brazilian construction industry. The construction industry is at the center of the Brazilian economic policy, therefore predicting this indicator is important, since it can provide an estimate of the economy's path for the next years. Since the time series presents strong seasonality, SARIMA models and variations of the Holt Winters' method were used.

The results showed that the ARIMA$(2,1,2)\times(0,1,1)_{12}$ model surpass the Damped Multiplicative Holt Winters' model in the fit and prediction of the construction industry GFCF, agreeing to the studies proposed by Milenković et al.. (2015) and Lepojević & Anđelković-Pešić (2011). In addition, the SARIMA model is able to eliminate the autocorrelation in its residuals, whereas the Holt Winters' model is not. Nevertheless, the Damped Multiplicative Holt Winters' model produces good results and in practical terms can be used to forecast the GFCF indicator. In addition, it is important to emphasize that the damped trend enhances the performance of the multiplicative Holt Winters' model in comparison with the regular one, which is in accordance with Gardner & McKenzie (1989) and Grubb & Mason (2001).

Therefore, both SARIMA and Holt Winters' methodologies are suitable to capture the seasonality and predict the GFCF of the Brazilian construction industry, which can be useful for decision-making by investors and business managers. Future studies should be conducted in order to investigate the poor performance of the Holt Winters' models in eliminating autocorrelation.

Prevendo a FBCF da Indústria da Construção Civil: Uma Comparação Entre Modelos de Holt Winters e SARIMA

RESUMO

O foco deste estudo é criar um modelo de previsão com a finalidade de prever o comportamento do indicador econômico conhecido como Formação Bruta de Capital Fixo (FBCF) da indústria da construção, o qual reflete a quantidade de investimento no setor da construção civil. O conjunto de dados consiste em observações mensais datando de janeiro de 1996 até dezembro de 2016, este último ano é usado como validação para o modelo de previsão. Como foi identificada presença de forte sazonalidade na série temporal, Modelos Autorregressivos Integrados de Médias Móveis Sazonais (SARIMA) e modelos de Holt Winters foram aplicados. Após a avaliação dos melhores modelos selecionados, o modelo ARIMA $(2,1,2)\times(0,1,1)_{12}$ é identificado como o melhor modelo para a previsão com resultados razoáveis. No entanto, o modelo de Holt Winters multiplicativo suavizado também apresenta bons resultados, apesar de ser incapaz de eliminar a autocorrelação nos resíduos. Portanto, ambos os modelos podem ser utilizados para prever a GFCF com boa precisão, o que pode ser útil para auxiliar nas tomadas de decisão por parte de investidores e administradores.

PALAVRAS-CHAVE: SARIMA. Box-Jenkins. Previsão. Holt Winters. FBCF.

REFERÊNCIAS

AMANN, E.; BAER, W.; TREBAT, T.; LORA, J. Infrastructure and its role in Brazil's development process. **The Quarterly Review of Economics and Finance**, n. 62, p. 66-73, 2016. crossref

BILLAH, B.; KING, M.L.; SNYDER, R. D.; KOEHLER, A.B. P.R. Exponential smoothing model selection for forecasting. **International Journal of Forecasting**, n. 22, p. 239-247, 2006. crossref

BON, R. The future of international construction: secular patterns of growth and decline. **Habitat International**, n. 16, p. 119-128, 1992. crossref

BOX, G.E.P.; JENKINS, G.M. **Time series analysis forecasting and control**. San Francisco: Holden-Day, 1996.

BUENO, R.L.S. **Econometria de Séries Temporais**. São Paulo: Cengage Learning, 2008.

BROWN, R.G. **Statistical forecasting for inventory control**. New York: McGraw Hill, 1959.

FIALHO, K.E.R; COSTA, H.N.; LIMA, S.H.Q.; NETO, J.P.B. Aspectos econômicos da construção civil no Brasil. **XV Encontro Nacional de Tecnologia do Ambiente Construído**. Nov 12-14; Maceió, Brazil, 2014.

GARDNER, E.S.; MCKENZIE, E. Note—Seasonal Exponential Smoothing with Damped Trends. **Management Science**, n. 35, p. 372-376, 1989. crossref

GUJARATI, D.N.; PORTER, D.C. **Basic Econometrics**. New York: McGraw Hill; 2009.

GRUBB M, MASON A. Long lead-time forecasting of UK air passengers by Holt–Winters methods with damped trend. **International Journal of Forecasting**, n. 17, p. 71-82, 2001. crossref

HOLT, C.C. The future of international construction: secular patterns of growth and decline. ONR Memorandum 52, Carnegie. Institute of Technology, Pittsburgh, 1959. Published in **International Journal of Forecasting**, n. 20, p. 5-13.

HYNDMAN, R.J; ATHANASOPOULOS, G. **Forecasting: principles and practice**. Melbourne: OTexts, 2017. http://otexts.org/fpp/

IPEADATABASE. **IPEA indicator of GFCF**. 2017. Available from: http://www.ipea.gov.br/cartadeconjuntura/index.php/tag/fbcf/

LEPOJEVIĆ, V.; ANĐELKOVIĆ-PEŠIĆ, M. Forecasting electricity consumption by using Holt-Winters and seasonal regression models. **Series: Economics and Organization**, n. 08, p. 421-431, 2011.

KURESKI, R; RODRIGUES, R.L.; MORETTO, A.C.; FILHO, U.A.S; HARDT, L.P.A. O macro setor da construção civil na economia brasileira em 2004. **Ambiente construído**, n. 01, p. 07-19, 2008.

MARTINEZ, E.Z., SILVA, E.A.S.; FABBRO, L.D. A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. **Journal of the Brazilian Society of Tropical Medicine**, n. 44, p. 436-440, 2011.

MORETTIN, P.A.; TOLOI, C.M.C. **Análise de séries temporais**. 2 ed. São Paulo: editora Blucher, 2006.

MILENKOVIĆ, M.; ŠVADLENKA, L.; MELICHAR, V., BOJOVIĆ, N.; AVRAMOVIĆ, Z. SARIMA modelling approach for railway passenger flow forecasting. **Transport**, n. 31, p. 01-08, 2015.

NG, S.; PERRON, P. Lag length selection and the construction of unit root tests with good size and power. **Econometrica: Journal of the Econometric Society**, n. 06, p. 1519-1554, 2001. crossref

PASSOS, A.G; MACIEL, M. A. C.; DORIA, M. R.; OLIVEIRA, R. B.; RUSSO, S. L. Análise estatística da evolução do produto interno bruto da indústria da construção civil brasileira utilizando regressão linear simples. **Revista Geintec**, n. 5, p. 505-514, 2012.

PUTHRAN, D.; SHIVAPRASAD, H.C.; KUMAR, K.S.K.; MANJUNATH, M. Comparing SARIMA and Holt-Winters' forecasting accuracy with respect to Indian motorcycle industry. **Transactions on Engineering and Sciences**, n. 02, p. 25-28, 2014.

SCHWERT, G.W. **Journal of Business and Economic Statistic,** n. 07, p. 147-159, 1989.

TANEJA, K.; AHMAD, S.; AHMAD, K.; ATTRI, S.D. Time series analysis of aerosol optical depth over New Delhi using Box-Jenkins ARIMA modeling approach. **Atmospheric Pollution Research**, n. 07, p. 585-596, 2016. crossref

TRATAR, F. L.; STRMCNIK, E. The Comparison of Holt-Winters method and Multiple regression method: A case study. **Energy**, n. 109,p. 266-276, 2016. crossref

TEIXEIRA, L.P.; CARVALHO F.M.A. A construção civil como instrumento de desenvolvimento da economia brasileira. **Revista Paranaense de Desenvolvimento**, n. 109, p. 09-26, 2005.

TSUI, W.H.K, BALLI, H.O., GILBEY, A., GOW, H. Forecasting of Hong Kong airport's passenger throughput. **Tourist Management**, n. 42, p. 62-76, 2013.

VALAKEVICIUS, E.; BRAZENAS, M. Application of the Seasonal Holt-Winters Model to Study Exchange Rate Volatility. **Inzinerine Ekonomika-Engineering Economics**, n. 26, p. 384-390, 2015.

WINTERS, P.R. Forecasting sales by exponentially weighted moving averages. **Management Science**, n. 06, p. 324-342, 1960. crossref

YIP, H., FAN, H. AND CHAING, Y. Predicting the maintenance cost of construction equipment: Comparison between general regression neural network and Box-Jenkins time series models. **Automation in Construction**, n. 38, p. 30-38, 2013.

Recebido: 21 jul. 2018
Aprovado: 15 fev. 2019
DOI: 10.3895/gi.v15n1.8590
Como citar:
THOMAZ, P. S.; MATTOS, V. L. D. Predicting the GFCF of the Brazilian Construction Industry: A Comparison Between Holt Winters' and SARIMA Models. **R. Gest. Industr.**, Ponta Grossa, v. 15, n. 1, p. 180-196, jan./mar. 2019. Disponível em: <https://periodicos.utfpr.edu.br/rgi>. Acesso em: XXX.
Correspondência:
Paulo Siga Thomaz
Rua Professor Doutor Araújo, número 2149, apt. 303, Bairro Centro, Pelotas, RS, Brasil.