

# MODEL HYBRID FOR SALES FORECAST FOR THE HOUSING MARKET OF SÃO PAULO

## Matheus Fernando Moro

Department of Production Engineering and Systems,  
Federal University of Santa Catarina, Brazil  
e-mail: morosmi@hotmail.com, ORCID: 0000-0002-0490-9233

## Andreas Dittmar Weise

Department of Industrial Engineering  
Hochschule 21, Harburger, Germany  
e-mail: mail@adweise.de, ORCID: 0000-0002-0431-4926

## Antonio Cezar Bornia

Department of Production Engineering and Systems  
Federal University of Santa Catarina, Brazil  
e-mail: bornia@ufsc.br, ORCID: 0000-0003-3468-7536

### Abstract

This research proposes a combined model of time series for forecasting housing sales in the city of São Paulo. We used data referring to the time series of sales of residential units provided by SECOVI-SP. The Exponential Smoothing, Box-Jenkins and Artificial Neural Networks models are individually modelled, later these are combined through five forecast combination techniques.

The techniques used are Arithmetic Mean, Geometric Mean, Harmonic Mean, Linear Regression and Principal Component Analysis. The measures of accuracy to measure the results obtained and to select the best model are the RMSE, MAPE and UTheil of forecast. The results showed that Linear Regression with an independent variable, being a combination of the SARIMA model (2,0,0)(2,0,0)12 and MLP/RNA (12,10,1), provided a satisfactory performance, with an RMSE of 368.74, MAPE of 19.2% and UTheil of 0.315.

The combination of time series models allowed a significant increase in forecast performance. Finally, the model was validated, using it to predict housing sales. The results show that the model has a good fit, thus demonstrating that using a housing sales forecasting model helps industry professionals minimize error and make sales and launch decisions.

---

**Key words:** real estate market, sales forecast, real estate management, analysis of real estate, forecast combination.

**JEL Classification:** C53.

**Citation:** Moro, M. F., Weise, A.D., Bornia, A.C. (2020). Model Hybrid for Sales Forecast for the Housing Market of São Paulo. *Real Estate Management and Valuation*, 28(3), 45-64.

**DOI:** <https://doi.org/10.1515/remav-2020-0023>

---

## 1. Introduction

Housing is a necessity of human beings (Temür, et al., 2019); this fact in itself already demonstrates the relevance and evidence that the housing market presents. Furthermore, it's one of the leading sectors of the Brazilian economy, with civil construction being one of its leading employers. In addition to this, it presents itself as an investment industry, in which returns occur in the long term

(Huang, et al., 2011). It should also be noted that the housing market is fundamentally conditioned by the financial market (Anghel & Hristea, 2015), thus, the pace of the negotiations is linked to economic factors (Moro et al. 2016a).

The ability of the housing market to boost the economy can be explored with the aid of computational techniques, such as the construction of models for predicting future scenarios. It should be considered that these models are of use not only to housing incorporations, but also to the entire construction industry involved, and even to government sectors. Understanding the behavior of the housing sector is crucial to the analysis of business cycles of the whole economy (Baghestani, 2017). Ballou (2016) emphasizes that the practice of forecasting is an important factor in the decision-making process of the organization and that predicting sales levels is fundamental to the company.

In housing markets, this need is even more pressing (Gupta, et al., 2019). In any of its segments, there is an expressive interval of time between decision making, investment and the effective entry of the venture in the market dispute. The conditions of the currently perceived attractiveness of the acquisition of terrain and the format of the enterprise can be diverse when it is launched and exposed to competition and its target audience (Gibler & Sherwood Clements, 2011; Hyung & Jung, 2013; Baltrusis & Mourad, 2014).

In the Brazilian scenario, the housing market of the city of São Paulo stands out as the most representative, since it is the largest city and the most important commercial center in Brazil. São Paulo is the second largest city on the American continent and the fifth largest city in the world (United Nations, 2014). Thus, to Moro et al. (2016a), the housing market of São Paulo has favorable characteristics for carrying out housing transactions, as well as follow-up on these by means of studies and research on the behaviors and phenomena they present. The city serves as a base for all research and studies involving housing, as well as exercising transmission power of demand on other metropolises inside the country.

In view of the above, this study aims to contribute to the expansion of literature in this area, and hopes its findings have practical implications for construction companies, housing incorporations, and investors. In addition, the research intends to show that, by using proper data manipulation techniques, data can provide information close to reality, thus demonstrating the importance of data analysis for the housing market. Therefore, this research aims to propose a sales forecast model combining time series forecasting techniques to identify the best accuracy for the housing market in São Paulo. Wang et al. (2018) indicates that forecast combination is one of the most important and effective approaches for time series forecasting, generating better predictions than when choosing just one model (Valle dos Santos & Vellasco, 2015). This statement is supported by studies such as de Temür et al. (2019), which used model combination techniques to generate better estimates of housing sales forecasts.

This research is organized into four sections; in section two, we present a theoretical reference on the housing market of São Paulo and forecast sales in the housing market. In section three, the methodology used is presented, presenting the steps used to reach the proposed objective. Section four presents the results and relevant discussions. Finally, the final considerations present the aspects that guided this work.

## **2. Theoretical Reference**

### **2.1 Housing Market of São Paulo**

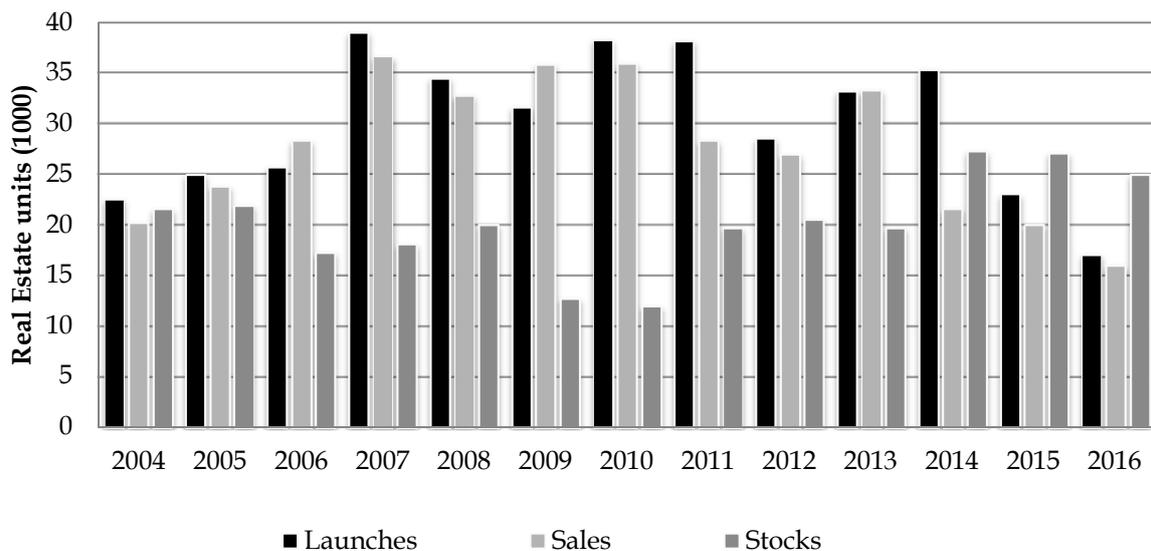
The housing market and all its components are interconnected with the service sector, which plays an important role in the Brazilian economy, accounting for more than half of the country's Gross Domestic Product (GDP) (IBGE 2014), with the housing market being responsible for 9.7% of the national GDP in 2014 (ABECIP 2015).

In this sense, driven by the growth and stability of the national economy, the residential housing market was faced with a vast amount of opportunities thanks to favorable economic conditions. The warming of the labor market brought the unemployment rate down to minimum levels and raised the income of the worker. This factor, coupled with the fall in the interest rate, intensified the long-term financing capacity for the acquisition and construction of housing developments, an energized effect with a government-directed credit policy through public works programs and subsidies for the sector (Mendonça, et al., 2011; Mendonça & Sachsida, 2012; Mendonça, 2013).

These events occurred in the years prior to 2013, when there was a great increase in housing sales in Brazil and especially in the city of São Paulo (Shimbo, 2013), but the stagnation of the national economy in the last three years was also reflected in the sales of housing in fifth largest city in the world. Between 2006 and 2010 there was a sharp increase in the number of residential launches, surpassing all previous records since the 1980s; according to data from the Empresa Brasileira de Estudos do Patrimônio (EMBRAESP 2013) and the Sindicato das Empresas de Compra e Venda de Imóveis em São Paulo (SECOVI-SP), this moment of intense rise in housing production occurred in 2007 and 2010 when, respectively, approximately 62 and 67 thousand residential units were released.

For Mendonça (2013), as of 2004, with governmental stimuli and a favorable international scenario, a change in the direction of the São Paulo housing market can be perceived. As a reflection, the businessmen of the sector returned to invest and provided a strong increase in the number of launches of housing ventures. In 2004 and 2006, the average number of launches was 24,385 units, and in the period between 2007 and 2011, an average of 36,280 units can be observed, signifying a 49% increase in the number of launches (Fig.1).

It is important to note that, in 2012, there was a decrease in the number of housing launched. It is important to note that 80% of the number of launches in that year were in the fourth quarter, tending towards a growth in the coming years. The reduction of economic activity in general contributed to the timid results of the Brazilian economy and, consequently, the national housing market for the year of 2012 (SECOVI-SP 2015). In the following two years, trends show a considerable increase. In 2015, approximately 800 thousand workers with a formal contract lost their jobs; according to data from the Registro Geral de Pessoas Empregadas e Desempregadas (CAGED), one of the sectors most affected was construction, which laid off more than 255 thousand workers in 2015. Together with these negative factors, inflation closed the year of 2015 at 10.67% (IBGE 2016). This combination of negative elements overturned not only the launches, but also the offers of units in the city of São Paulo.



**Fig. 1.** Units launched, sold and in stock in São Paulo city between Jan/2004 and Jun/2016. *Source:* Prepared by authors prepared with data from (SECOVI-SP 2017).

The recovery of the sector, which led to the reoccurrence of levels such as those observed in the 1990s, can be attributed to improvements in financing conditions, falling interest rates, reduction of unemployment, increased income, and, in general, greater confidence of the society in the Brazilian economy. The increase in the number of launches reflects the confidence of entrepreneurs and investors in the sector, but is also the result of the increase in demand for this type of good.

Regarding the demand, the increase in demand for housing and the best conditions for the realization of the dream of home ownership have become a reality in recent years. According to SECOVI-SP (2016), 2007 was the year characterized by the highest sales of housing, i.e. 36,600 units (see Figure 1); this value was driven by the implementation of the PMCMV. In 2008, there was a slight drop, but in the following two years, 2009 and 2010, there was an increase in the supply of credit. After that sales fell, explained by market adjustment. In the year 2015, the fall is evident, and the explanation is the same for the fall of launches: due to political, economic and institutional crises, as

well as layoffs in construction and a high inflation rate.

The success of sales enabled many entrepreneurs to be able to sell stocks, sell units still in the project stage, and market an entire business over just one weekend, a success for the housing businesses and industry investors. But what was a success, has become a problem, seeing as how the number of units stocked in the city of São Paulo in the years of 2014 and 2015 set positive records (SECOVI-SP 2016). From 2010 to 2014, the stock in São Paulo grew 126% and sales accompanied the pessimism, a drop of 40% (see Figure 1). The scenario of 2016, with a scarcity of savings and difficulties in launching loans by banks, led people to think long and hard before making the decision to acquire a loan that would impact from 20% to 30% of the budget for over 20 years or more.

A large quantity of stock increases the operating costs of developers, who cannot lower the prices of new housing, precisely because of their costs. Thus, if the developers used adequate forecasting methods, they could predict this behavior and hence gain the ability to stabilize immobile launches. Within this context, the discussion on forecasting sales in the housing market will be commenced in the next section.

## 2.2 Housing Sales Forecast Studies

In studies of the housing sector, one should consider aspects related to the product offered as well as sale and price formation factors. There is therefore a need to make a sales forecast for the product during the planning of the enterprise. A sensible forecast of sales helps the managers of the developers in making decisions regarding the number of projects, launches and stock. In addition, knowledge of the functioning of the housing market allows the government, for example, to carry out macroeconomic simulations aimed at generating jobs, or to assist in the planning of popular housing policies (Moro, et al., 2016b).

The quantitative studies on housing in Brazil, although frequent, are focused on price estimates. It is therefore possible, through forecasting techniques, to extract information that will allow the modelling of their behavior from the available data. The assumption of continuity in this behavior allows for the realization of forecasts, whose quality and precision are far superior to those of intuitive predictions, based only on the analysts' experience, a characteristic of the sales forecasting models used by the Brazilian incorporators (Moro, et al., 2016b).

Regarding the scientific means, there is scarce research on the of forecasting sales for the housing market. Dua and Miller (1996) developed a Bayesian Vector Autoregressive Model (BVAR) to predict home sales in Connecticut, including household price variables, mortgage interest rates and unemployment rates. The same authors supplemented the paper and used BVAR in the study to examine the utility of leading indicators in predicting US home sales. The Bayesian model of reference included sales of residences, house prices, mortgage rate, real personal disposable income and unemployment rate (Dua, et al., 1999).

Huang et al. (2011) analyzed the main elements that influence China's housing sales. Using regression models, they concluded that the per capita income of the population and the price of housing are the two main factors for modifying sales. Liu and Li (2009) proposed a multiple regression equation to analyze the factors affecting housing sales in Beijing after the 2008 Olympic Games. Per capita income, population size and the average price of commercial housing were the variables found.

The expectation of population aging increases the need for accurate models and reliable forecasts of housing demand (Gibler, et al., 2009; Gibler and Sherwood Clements, 2011). Gibler and Sherwood Clements (2011) aimed at analyzing the only published forecast model, which predicts the real housing movement of older Americans in a retirement situation. This model is based on logistic regression and is used to test the causes that influence the choice of housing in retirement using data from 2002 and 2004. The applied tests showed that more research is needed to develop the model to be able to estimate, with more precision, the housing market demand in countries with aging societies.

In addition, Cseko and Reed (2009) cite that the incorporators will not be able to supply the demand for housing for retirees, due to the lack of studies on this demand. Similarly, Ball and Nanda (2013) conducted a study on the housing demand for retired people, aiming to report the need to estimate future demand for specialized housing for retirees using data for England. They concluded that there is a need to create specialized methods to forecast demand for this sector, since the demand for housing for retired people is high. Chen (2010) proposed an approach that uses financial indicators to predict the sales of large Taiwanese construction corporations; he used 11 macroeconomic variables,

such as GDP, income, price index, etc. The author used hypothesis testing and then a two-stage model combining regression and time series. This model had the objective of predicting the company's turnover rather than the sales of units, but the result shows the applicability of time series models in the housing market.

Temür et al. (2019) uses a data set from 2008 to 2018 to estimate Turkey's housing sales time series. For this, the authors create a hybrid model using ARIMA (Automatic Regressive Integrated Moving Average as a linear model) and LSTM (Long Term Memory as a nonlinear model). Through MAPE and MSE, the authors prove that combined models improve prediction. The authors say that, through this model, it will be possible to make predictions about the future of the country's economy and create useful information for other sectors affected by the housing sector. In addition, they encourage other authors to develop models that combine time series to predict housing sales.

Hassani et al. (2017) compared various models to forecast housing sales in the US and its four separate regions (Northeast, Midwest, South and West). The methods used were two different versions of the single spectrum analysis (SSA) methods, namely recurrent SSA (RSSA) and vector SSA (VSSA), in univariate (UV) and multivariate (MV) structures. The authors concluded that UVVSSA is the best performing model for US aggregate domestic sales, while MV versions of RSSA are the favorite in forecasting home sales for all four regions. The authors state that accurate housing sales forecasts can provide valuable information not only to policymakers, but also to financial institutions and housing professionals.

In relation to Brazilian studies, research related to the sales forecast in the housing market is scarce, and it can be concluded that they are practically nonexistent. Varandas Junior and Alencar (2009) aimed to test the applicability of DW demand prediction models in developing economies. They concluded that once the application of these models requires a historical series of data not available in Brazil, their study becomes relevant for the interpretation of movements of the sector and for the planning the agents involved in this market, thus, creating a demand in research for methods that the current research can provide.

### 3. Data and Methods

This research is classified as quantitative by making use of statistical modeling for forecasting sales in the São Paulo housing market. In view of its object of study, it is an exploratory research, since sales data was collected directly from the sales spreadsheets. The data is spaced monthly, covering the period from January 2004 to June 2016, and provided by CBIC through surveys of SINDUSCON-SP and SECOVI-SP. The current data is no longer offered seeing as how the data disclosure policy has been changed. Figure 2 shows the steps of the research.

At first, a series comprising 150 observations is divided in two, with 75% for modeling/adjustment of the models and another 25% to measure the accuracy of the predictions generated by the models. The first Step of the research corresponds to the analysis of the time series graph to find the characteristics of the series. To confirm the evidence found in the visual analysis of the graphs generated by the decomposition, the Wald-Wolfowitz, Cox-Stuart and Mann-Kendall tests are utilized to test for the presence of trends under the hypotheses  $H_0$ : There is no trend and  $H_1$ : The series presents a trend. To test the presence of seasonality, the Kruskal-Wallis and Friedman tests are used, under the hypotheses  $H_0$ : There is no seasonality and  $H_1$ : There is seasonality. For all the tests, the confidence level was at 95%, realized in the Action Stat software.

The second Step corresponds to the modeling of the Exponential Smoothing methods with state space correction proposed by Ord et al. (1997). This is a representation of SES models through state space models with an only error font. The advantage of this representation is that statistic properties can be studied as an estimate of the average and variance of forecasts, thus generating better forecasts. Altogether, there are 20 methods which depend on the characteristic of the series to be used. For the modelling step of the SES methods, NNQ-Stat software is used. In Hyndman et al. (2002) the representation of smoothing models is proposed under two additive and multiplicative error conditions. Frame 1 shows the notation of these methods.

These are cases already seen in Ord, Koehler and Snyder (1997), who considered the situations with homoscedastic error (constant variance) and heteroscedastic error (non-constant variance). When the type of error is selected, the classification of the model is realized with three letters, being the first letter, the error, additive (A) or multiplicative (M). The second letter is the trend, additive (A),

multiplicative (M), additive damped (Ad) or multiplicative damped (Md). The third letter is the seasonality, additive (A) or multiplicative (M).

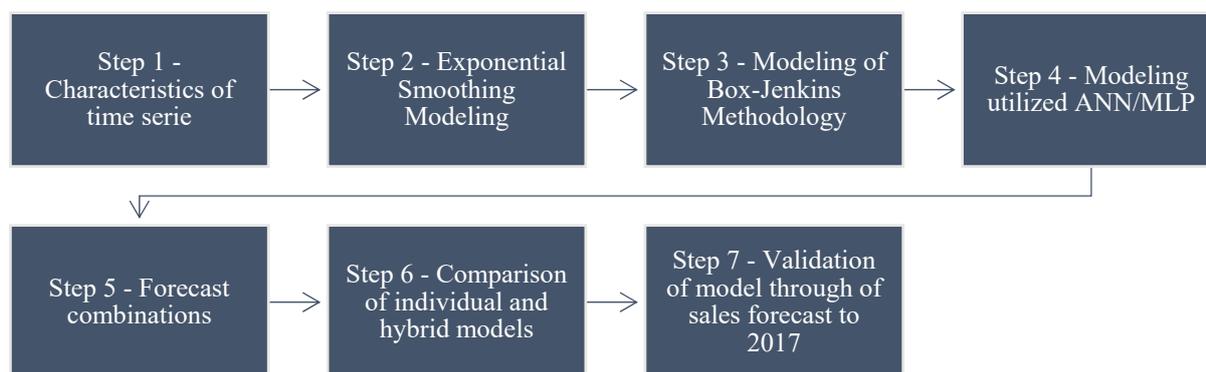


Fig. 2. Methodology steps of research. Source: Prepared by the authors.

Frame 1

Notation of exponential smoothing methods with state-space correction

Trend	Seasonality					
	None (N)		Additive (A)		Multiplicative (M)	
	Error					
	A	M	A	M	A	M
None (N)	ANN	MNN	ANA	MNA	ANM	MNM
Additive (A)	AAN	MAN	AAA	MAA	AAM	MAM
Additive Damped (Ad)	AAdN	MadN	AAdA	MAdA	AAdM	MAdM
Multiplicative (M)	AMN	MMN	AMA	MMA	AMM	MMM
Multiplicative Damped (Md)	AMdN	MMdN	AMdA	MMdA	AMdM	MMdM

Source: Prepared by the authors, adapted from HYNDMAN et al. (2002).

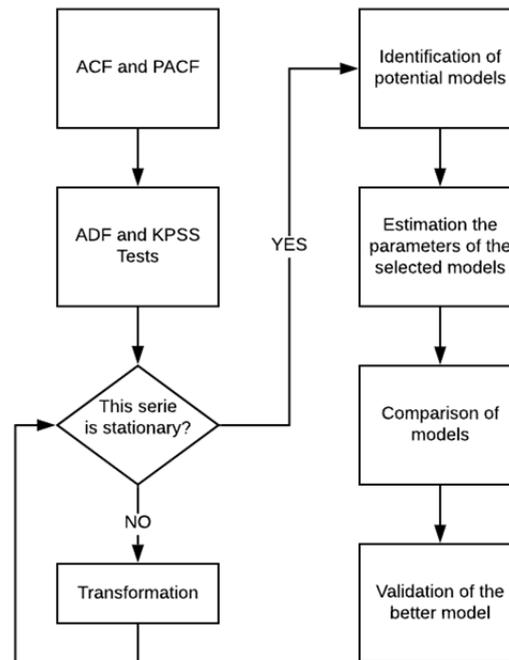
The third Step consists of applying the Box-Jenkins methodology to obtain the ARIMA model that best fits and generates good predictions for the time series being studied. The modelling of the ARIMA process will be based on the sequencing execution of Box-Jenkins methodology (BOX et al. 2008). This sequence is described in Fig. 3, which shows the steps taken to obtain the ARIMA model for time series forecast.

The first step is to perform an earlier analysis of the behavior of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to find evidence of stationary series. Then, the increased Dickley-Fuller (ADF) (DICKEY & FULLER, 1979) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) (Kwiatkowski, et al., 1992) unit root statistical test was applied with the following hypotheses: H0: the series that has unit root; and H1: the series has no unit root (stationarity) for the ADF test and inverted hypotheses for the KPSS test. This was done to confirm the initially established hypothesis regarding the stationarity of the series. The test was applied with a significance level of 1%, 5% and 10%.

After that, the potential ARIMA models were identified through ACF and PACF. Thus, through the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), the best models were selected, which present lower values of the criteria, for estimation of the parameters. Finally, the selected models are compared using the MAE, MAPE, RMSE and UTHEEL criteria (which will be seen in Step 6), thus choosing the model that generates the best prediction. The validation of the model is therefore realized. The validation is realized through the graphic analysis of the standardized residuals and analysis of residuals of the ACF and PACF obtained for the model. If the obtained model has residuals with zero average, constant variance and non-have autocorrelation, this model is validated for use to forecast future values of the time series. For Step 3, the Gretl software was used.

Step 4 aims to perform the modelling and prediction of the series values through the Artificial Neural Network Multilayer Perceptron (ANN/MLP) model (KASHEI, BIJARI 2010). For the modelling of Artificial Neural Networks (ANN), only the Multilayer Perceptron (MLP) topology with a hidden layer was used, also known as the middle layer. Statistica 10.0 software was used for training and modelling ANNs. The first step was to assemble the network architecture, that is how many

neurons each layer will have. The number of neurons in the input layer was defined based on the seasonal characteristic observed in the studied time series. For the intermediate layer, tested with 5, 10, 15, 20, 25 and 30 neurons, the output layer will have 1 neuron as this is a one-dimensional series.



**Fig. 3.** Flowchart of the methodology for obtaining the ARIMA model for forecasting the time series.

*Source:* Prepared for the authors.

After deciding on the architecture of the network, the training of the network starts; in this step, two algorithms were used in two phases. The first phase of the training was performed by the Backpropagation algorithm and the second phase was performed by the Conjugate Gradient Descent algorithm. For all the ANN models used in this study, the number of 20,000 training cycles was stipulated, with the objective of avoiding overadjustment to the training data. After the training of the temporal series, the prediction of 38 steps forwards for each model is carried out.

After the forecast is made, the comparison between all the models using the metrics of accuracy will be seen in Step 6. The best model chosen is the one with the lowest values of the accuracy statistics, and it is also necessary to verify if there was no overadjustment. For verification of the ANNs overadjustment to the data of the series, the analysis consists of comparing the variation of the error measures of the training process and of the ANN validation process. If there are sharp divergences between such measures, there is evidence that a given ANN would have memorized the data provided for its training, losing the capacity of generalization, a situation that does not generate practical results for the study. The MLP nomenclature  $(X, Y, 1)$  was used for simplicity, where  $X$  is the number of neurons in the input layer containing the lag values for  $X$  periods,  $Y$  is the number of neurons in the hidden layer and 1 is the number of neurons in the output layer. For Step 4, Statistica 10.0 software was used.

After the modeling of the three model classes, in Step 5, the predictions generated by these models are combined by Arithmetic Mean - MT (Stock & Watson, 2004), Geometric Mean - MG and Harmonic Mean - MH (Patton & Sheppard, 2009). In addition to these combination methods, the predictions generated by the individual models are combined by Linear Regression - RL (Granger and Ramanathan 1984). Thus, the dependent variable will be the sales series used for validation of the model and the independent variables will be the predictions generated by the three model classes. Finally, the Principal Component Analysis (PCA) method (Skittides & Fruh, 2014) is used to combine the predictions using the first factor to obtain the weights in the predictions. Combinations with MT, MG and MH are performed in Excel, while combinations using PCA and RL are performed using Statistica 10.0 software.

In Step 6, the individual and combined models are compared by means of the following measures of accuracy: U-Theil, Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) shown in (Bratu, 2012). The proposed model, the one that will present the best accuracy, that is, the smallest forecast error, will be validated foreseeing the sales of 2017, this being executed and discussed in Step 7.

#### 4. Empirical results

The first step of the research begins with the intention of understanding the behavior of the series of housing sales in São Paulo. Thus, Fig. 4 presents the graph of the series under study, along with showing the trend line. It is emphasized that only 75% of the series will be used to adjust the forecast models.

It can be seen from the trend line that there is evidence that the series has a positive trend, as shown in the same figure. The time series presents different patterns of behavior in different periods of the series, especially at the beginning (lower variance) in relation to its end (greater variance). In addition to the trend, the series presents evidence of seasonality, since in the month of December there are peaks of sales volumes over the years.

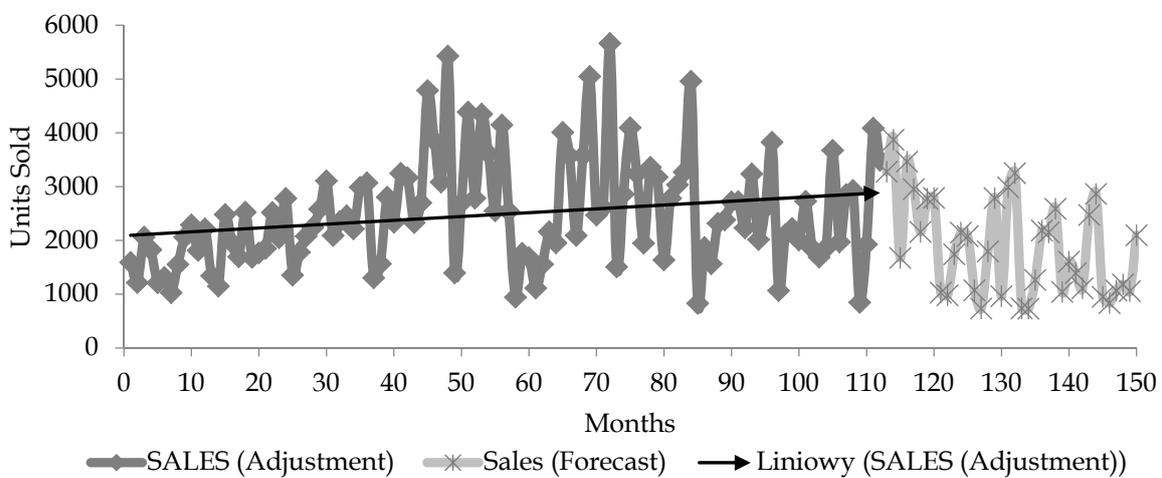


Fig. 4. Housing sales in São Paulo between Jan/04 and Jun/16. Source: Prepared by the authors.

Table 1

		Trend and seasonality detection tests					
		Wald-Wolfwitz		Mann-Kendall		Cox-Stuart	
Trend	Statistic	42		0,18		29	
	p-value	<0.001		<0.001		0.045	
		Seasonality = 3		Seasonality = 6		Seasonality = 12	
Seasonality	Tests	Kruskal-Wallis	Friedman	Kruskal-Wallis	Friedman	Kruskal-Wallis	Friedman
	Statistic	54,81	37,14	46,59	45,59	48,66	50,11
	p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Source: Prepared by authors.

To confirm these hypotheses, Table 1 presents the results of the trend and seasonality tests. By observing the results of the trend tests, the null hypothesis is rejected for all three; since the p-value is lower than the level of significance (5%), it is affirmed that the series under study presents a characteristic trend.

Observing the results of the seasonality tests, it is verified that, for the seasonality sizes 3, 6 and 12, the null hypothesis is rejected, thus it is affirmed with 95% confidence that the series presents a seasonal component of sizes 3, 6 and 12.

In this way we proceeded to model the Exponential Smoothing methods with state space correction. As pointed out, the series of residential housing sales in the city of São Paulo has

characteristics of trend and seasonality. In view of this, only the eight appropriate methods for series with these two characteristics were modeled. Table 2 presents the parameters and accuracy measures calculated for each of the eight methods of SE with trend and seasonality.

Table 2

Parameters and error measures for SES methods with trend and seasonality

	AAA	MAA	AAdA	MAdA	MAM	MAdM	MMM	MMdM
Parameters	$\alpha = 0.32$	$\alpha = 0.08$	$\alpha = 0.32$	$\alpha = 0.08$	$\alpha = 0.30$	$\alpha = 0.30$	$\alpha = 0.31$	$\alpha = 0.29$
	$\beta = 0.01$	$\beta = 0.01$	$\beta = 0.01$	$\beta = 0.01$	$\beta = 0.01$	$\beta = 0.01$	$\beta = 0.01$	$\beta = 0.01$
	$\gamma = 0.01$	$\gamma = 0.25$	$\gamma = 0.01$	$\gamma = 0.23$	$\gamma = 0.01$	$\gamma = 0.01$	$\gamma = 0.01$	$\gamma = 0.01$
			$\varphi = 0.80$	$\varphi = 0.98$		$\varphi = 0.80$		$\varphi = 0.80$
RMSE	1625.733	1145.344	1450.281	<b>1029.571*</b>	1555.361	1403.714	1899.950	1386.785
MAPE	1.134	0.758	1.009	<b>0.670*</b>	1.039	0.942	1.255	0.930
UTHEIL	1.386	0.976	1.236	<b>0.877*</b>	1.326	1.196	1.619	1.182

\*Best model for each measure of accuracy.

Source: NNQ-Stat output adapted by the author.

The model of Additive Tendency Cushioned with Additive Seasonality and Correction of Multiplicative Error (MAdA) obtained the smallest statistics, so the MAdA model is the model that generates the most accurate forecast among the SE methods.

Having the UTheil being smaller than one, that is, the prediction of the model is better than a naive prediction, it turns out that the error values are high for a good prediction, such as a MAPE of 67% error. One caveat to keep in mind is in relation to the value of the smoothing constants  $\alpha$ . The lower the value of  $\alpha$ , the more stable the final predictions, since larger weights are given to past observations, thus, any random fluctuation, at present, will exert a much smaller weight in the calculation of forecasts.

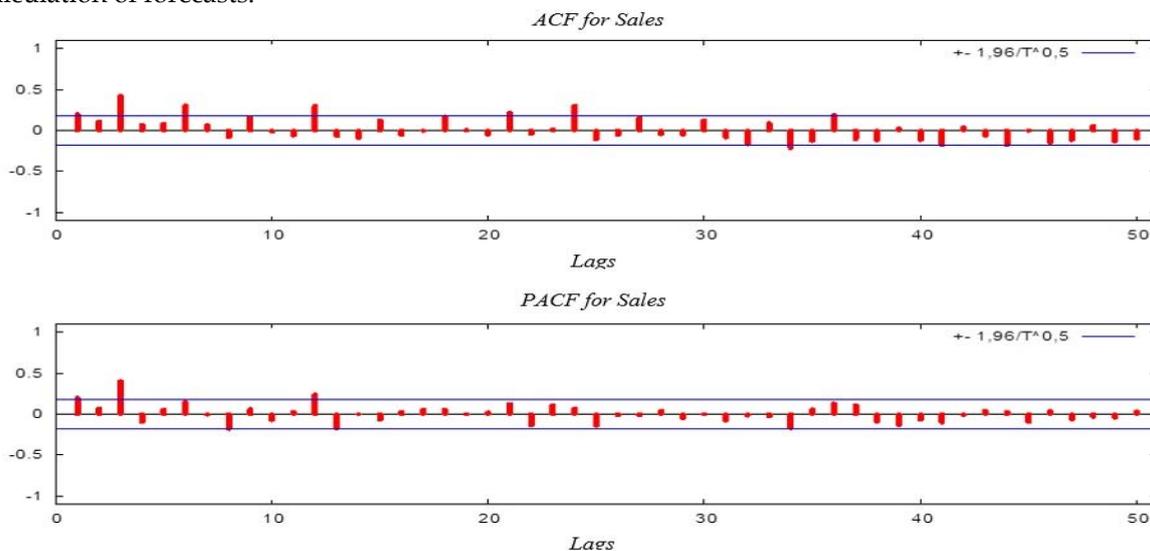


Fig. 5. Correlogram of ACF and PACF of the time series of sales of residential units in São Paulo.

Source: Gretl output.

After modeling the SES methods, one starts to model the Box-Jenkins methodology. The first step in using the Box-Jenkins methodology is the analysis of ACF and PACF. These functions, added to the visualization of the time series graph, help to verify the stationarity condition and contribute to identifying the possible models for the time series. These functions can be observed in Figure 5, where the ACF presents a rapid decay, contrary to the hypothesis that the series is non-stationary, thus providing evidence of the hypothesis that the series is stationary. It is also noted that, in both ACF and PACF, the lag 12 presents a significant behavior, fueling the evidence of autoregressive seasonality. In the PACF, two significant lags can be verified, demonstrating that there may be two self-correcting parameters in the model.

Still in relation to the stationarity of the series, it was chosen to perform the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) unit root test at level, I (0), presented in

Table 3 to reach an adequate conclusion regarding this condition. It is worth noting that, in the ADF test, the series presented a steady behavior at a significance level of 5%, since the critical value of the test is less than -2.88 This fact was confirmed for the KSPP Test and the hypothesis evidenced in the ACF analysis, where rapid decay was observed. Thus, it is affirmed that the series is stationary.

**Table 3**

ADF test applied to housing sales series

		ADF		KPSS	
		t-Statistic	p-value	t-Statistic	p-value
I (0)	Critical	-3.331414	0.0158	0.434665	0.062
	1% level	-3.491345		0.735	
	5% level	-2.888157		0.462	
	10% level	-2.581041		0.349	

Source: Gretl output adapted by authors.

After the unit root verification test, we proceeded to the second phase of the Box-Jenkins methodology, which is the identification of potential models, together with the analysis of ACF and PACF. The analysis of the ACF and PACF allowed for the identification of six potential models for the modeling process of the series, which are presented in Table 4, together with the values of the Akaike Information Criteria (AIC) and Bayesian Information Criterion (BIC) criteria.

**Table 4**

Competing models for parameter estimation and process modeling

	AIC	BIC	Observations
ARIMA (3,0,0)	13.7701	13.7929	
SARIMA (0,0,1)(1,0,0) <sub>12</sub>	13.8516	13.9001	
SARIMA (1,0,0)(1,0,0) <sub>12</sub>	13.7925	13.8410	
SARIMA (2,0,0)(1,0,0) <sub>12</sub>	13.7241	13.7846	Selected model for the estimation
SARIMA (2,0,0)(2,0,0) <sub>12</sub>	13.6255	13.7225	Selected model for the estimation
SARIMA (2,0,0)(3,0,0) <sub>12</sub>	13.5668	13.6882	Selected model for the estimation

Source: Prepared by authors.

**Table 5**

Estimation of parameters and accuracy of adjustment and prediction of concurrent Box-Jenkins models

	SARIMA (2,0,0)(1,0,0) <sub>12</sub>	SARIMA (2,0,0)(2,0,0) <sub>12</sub>	SARIMA (2,0,0)(3,0,0) <sub>12</sub>
Parameters	$\varphi_1 = 0.52288$	$\varphi_1 = 0.47478$	$\varphi_1 = 0.46748$
	$\varphi_2 = 0.32194$	$\varphi_2 = 0.27504$	$\varphi_2 = 0.30427$
	$\Phi_1 = 0.5479$	$\Phi_1 = 0.4091$	$\Phi_1 = 0.2891$
		$\Phi_2 = 0.4051$	$\Phi_2 = 0.3219$
			$\Phi_3 = 0.2961$
RMSE	804.941	<b>726.837*</b>	887.452
MAPE	0.367*	<b>0.442</b>	0.559
UTHEIL	0.686	<b>0.619*</b>	0.756

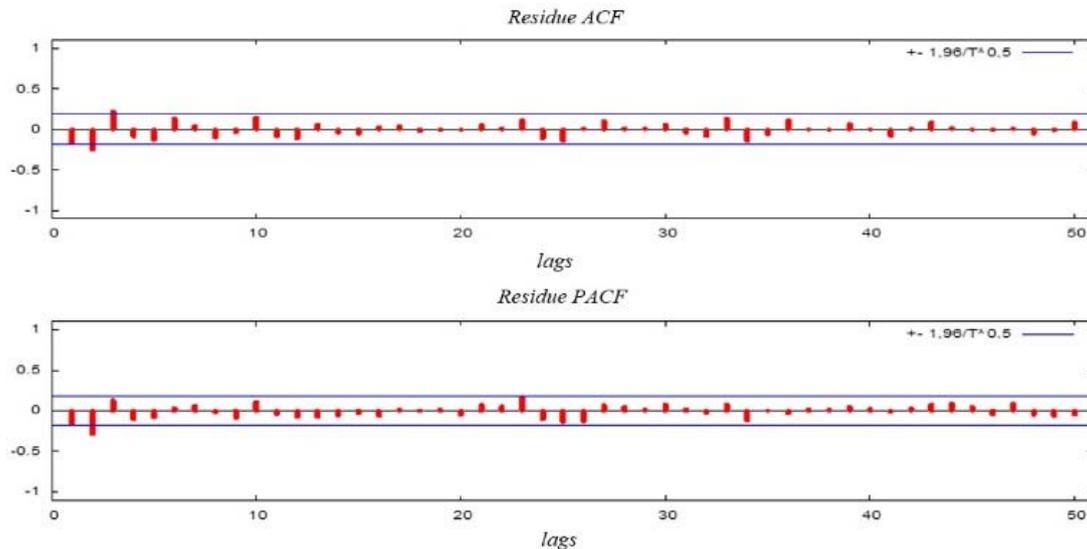
\*Best model for each measure of accuracy.

Source: Prepared by authors.

Several models were tested, but they are not competing models, because their parameters were greater than 1 or were not statistically significant. It was decided to choose the three models with smaller AIC and BIC to estimate the parameters. The next step was to estimate the parameters of the selected models and to compare the values of the accuracy measures for the adjustment and forecast of the series, aiming to define which SARIMA model generates the best prediction. Table 5 shows the results of parameter estimation and the values of the accuracy measurements.

The competing models and potential good predictors found are all models with white noise, i.e., mean equal to zero and constant variance over time, coupled with the fact that the parameters are all

smaller than one. Table 5 shows that the best fit and prediction model of the series is the SARIMA model  $(2,0,0)(2,0,0)_{12}$ , highlighted in bold, presenting the lowest values for the accuracy measurements. This yielded a mean error of 44.2% and a UTheil prediction value of 0.619. After this analysis, to confirm the parsimony of the model, the correlogram of the residues in Fig. 6 is presented.



**Fig. 6.** Residue correlogram SARIMA Model  $(2,0,0)(2,0,0)_{12}$ . Source: Gretl output.

Having performed the ARIMA modeling, a start for RNA / MLP modelling, the first step in this stage was to define the network configuration, where, among others, the number of neurons in the input layer was established. The number of neurons in the hidden layer and the output layer had already been previously defined, based on the research methodology. Considering the existence of a seasonal component, having been verified in the graphical analysis, the seasonal tests and the Box-Jenkins models, the tested configurations for the input layer of the ANN/MLP models were 3, 6, 12 and 18 neurons in the input layer. The network configuration is shown in Frame 2, containing the number of neurons in each layer.

#### Frame 2

Configuration of ANNs modeled in the present study

MLP Network	Number of neurons
Input layer	3, 6, 12 e 18
Hidden layer	5, 10, 15, 20, 25, 30
Output layer	1
Number of models	24

Source: Prepared by authors.

Next, network training was realized, during which the weights of the connections between each layer for each model were established. After the training, the modeling and forecasting of the series were carried out using the trained MLP models, and then the results were recorded. For the sake of organization, the MLP model notation was used in the description of the models, as in the example - MLP (3,5,1) - this being the MLP model with 3 neurons in the input layer, 5 neurons in the hidden layer and 1 neuron in the output layer. The results of modelling and forecasting are presented in Table 6.

**Table 6**

Results of MLP/RNA models utilized for the modelling and time series forecast of residential housing sales of São Paulo

MLP Models	RMSE	MAPE	UTHEIL
<b>MLP (3,5,1)</b>	817.815	0.510	0.697
<b>MLP (3,10,1)</b>	837.853	0.521	0.714
<b>MLP (3,15,1)</b>	743.519	0.463	0.633

MLP (3,20,1)	776.869	0.464	0.662
MLP (3,25,1)	784.551	0.485	0.668
MLP (3,30,1)	720.636	0.385	0.614
MLP (6,5,1)	883.940	0.600	0.753
MLP (6,10,1)	782.992	0.545	0.667
MLP (6,15,1)	836.208	0.521	0.712
MLP (6,20,1)	915.504	0.530	0.780
MLP (6,25,1)	965.077	0.547	0.822
MLP (6,30,1)	992.032	0.558	0.845
MLP (9,5,1)	781.964	0.449	0.666
MLP (9,10,1)	879.196	0.545	0.749
MLP (9,15,1)	1148.434	0.612	0.978
MLP (9,20,1)	839.221	0.455	0.715
MLP (9,25,1)	1104.608	0.621	0.941
MLP (9,30,1)	906.806	0.433	0.772
MLP (12,5,1)	960.429	0.645	0.818
MLP (12,10,1)	<b>644.985*</b>	<b>0.389*</b>	<b>0.549*</b>
MLP (12,15,1)	896.113	0.574	0.763
MLP (12,20,1)	1136.083	0.451	0.968
MLP (12,25,1)	826.470	0.395	0.704
MLP (12,30,1)	1057.881	0.526	0.901
MLP (15,5,1)	865.15	0.497	0.737
MLP (15,10,1)	884.854	0.399	0.754
MLP (15,15,1)	766.42	0.41	0.653
MLP (15,20,1)	1003.613	0.548	0.855
MLP (15,25,1)	1071.706	0.594	0.913
MLP (15,30,1)	815.188	0.42	0.694
MLP (18,5,1)	969.126	0.648	0.825
MLP (18,10,1)	780.521	0.448	0.665
MLP (18,15,1)	834.003	0.485	0.710
MLP (18,20,1)	826.664	0.485	0.704
MLP (18,25,1)	937.652	0.448	0.799
MLP (18,30,1)	979.636	0.440	0.834

\*The best model for each measure of accuracy.

Source: Prepared for authors.

The model highlighted in bold is that which generated the best fit and best prediction. The model RNA/MLP (12,10,1) obtained 38.9% of MAPE and UTHEEL of 0.549, showing that the model generates a good prediction. To be sure that the tested model did not show super adjustment, a simple linear regression was performed, so the degree of adjustment of the simple linear regression model measured how much the MLP model (12,10,1) was able to describe the values presented, being analyzed by the coefficient of determination  $R^2$ . In the training, the target values included in the training set of the models were presented in the test process. The target values not presented in the training were presented, making it possible to verify the potential of the model to predict out-of-sample values and the occurrence of over adjustment.

The  $R^2$  value obtained in the training of the model was equal to 0.831 and, for the validation process, an  $R^2$  value equal to 0.808 was obtained. Thus, it is possible to conclude that, in the training of the MLP model (12,10,1), there was no over-adjustment since there is no large variance between the training  $R^2$  and validation  $R^2$ , thus stating that the model does not show high variance when dealing with unknown data. Another test to confirm that there was no over-adjustment is the training graph analysis of the MLP model (12,10,1), which is presented in Fig. 7. It was observed that the mean error of the training process (0.0095) converged to the stipulated target error of 0.01. The error of the validation process was 0.563, verifying that the model reached acceptable values of error for the 20000 training cycles.

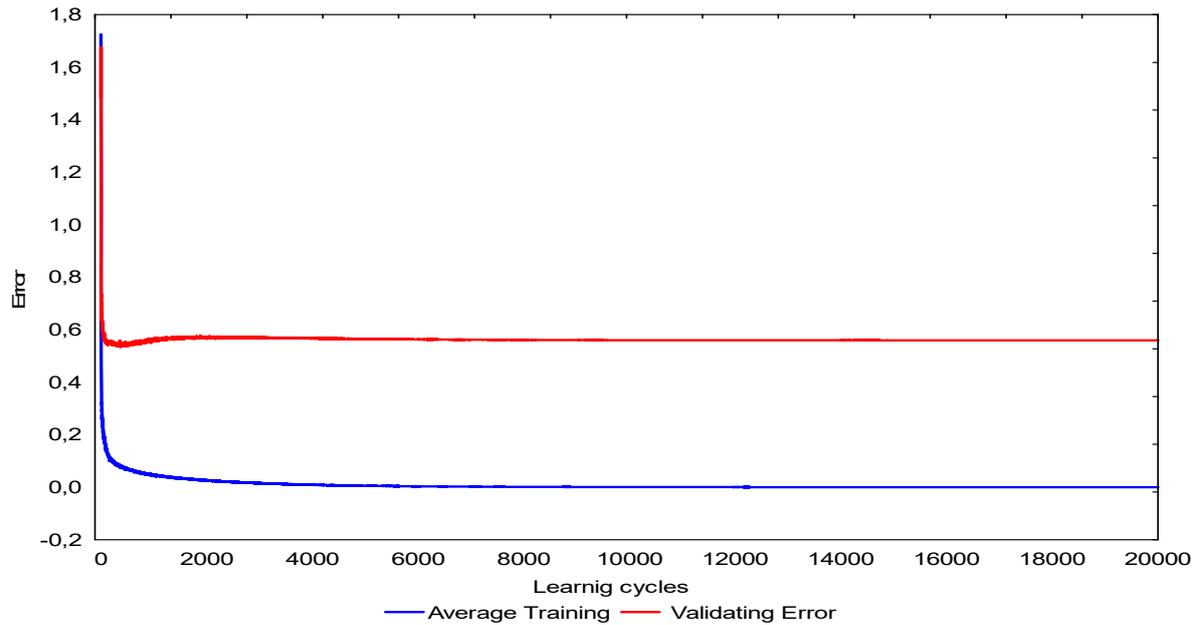


Fig. 7. MLP training graph (12,10,1). Source: Statistica output.

What stands out in the ARIMA model, is that the seasonality was found to be 12. It was expected that the MLP network with 12 neurons in the input layer would produce the best results, a fact which was confirmed in the RNA modeling. This can be explained seeing as how, in December, a portion of the thirteenth salary is paid, so the consumers take advantage of this payment to enter the purchase of new housing.

By obtaining the individual models, we can combine the forecasts for the series using the combination methods described in the methodology section. The individual MAdA, SARIMA (2,0,0) (2,0,0)12 and MLP (12,10,1) models will be combined. For organizational purposes, the underline notation ( ) means a combination of methods, and the bar notation (/) means that a combination method was used and after (/), which combination method was used.

The first combination method used was the Arithmetic Average (MD), generating the respective MAdA\_SARIMA/MD, MAdA\_MLP/MD, SARIMA\_MLP/MD and MAdA\_SARIMA\_MLP/MD models. The second combination method used was the Geometric Average (MG), obtaining the respective models: MAdA\_SARIMA/MG, MAdA\_MLP/MG, SARIMA\_MLP/MG and MAdA\_SARIMA\_MLP/MG. The method of Harmonic Average (MH) was utilized for was utilized for combination, generating the respective models: MAdA\_SARIMA/MH, MAdA\_MLP/MH, SARIMA\_MLP/MH e MAdA\_SARIMA\_MLP/MH. The penultimate method used was a combination of Linear Regression (RL), consequently the presented models are: MAdA\_SARIMA/RL, MAdA\_MLP/RL, SARIMA\_MLP/RL and MAdA\_SARIMA\_MLP/RL. Thus, four linear regressions were performed, one for each method; in Table 7, we can observe the values of the regression coefficients and the Adjusted R<sup>2</sup> for each individual model in front of the combined model.

The last combination method used was Principal Component Analysis (PCA), testing the models: MAdA\_SARIMA/PCA, MAdA\_MLP/PCA, SARIMA\_MLP/PCA e MAdA\_SARIMA\_MLP/PCA. For this method, we tried to obtain only one factor (main component), which has the largest explained variance, since we are working with a maximum of three variables (three individual forecast models). Each factor generates a linear combination of the original variables with the purpose of retaining the maximum information in terms of total variation, that is, the contribution of each variable. Table 8 shows the contribution values of each variable and the respective variance explained by each model.

Table 7

Coefficient  $w$  for each model to define the prediction combination using the RL method

	MAdA_SARIMA	MAdA_MLP	SARIMA_MLP	MAdA_SARIMA_MLP
Intercepto ( $w_0$ )	-257.115	-452.052	-455.375	-530.698

MAdA ( $w$ )	0.536	0.307	-	0.123
SARIMA ( $w$ )	0.607	-	0.364	0.301
ANN ( $w$ )	-	0.831	0.793	0.762
R <sup>2</sup> Adjusted	56.24%	75.30%	76.92%	76.90%

Source: Prepared by authors.

Table 8

Contribution of each individual model in Factor 1 for linear combination using the ACP method

	MAdA_SARIMA	MAdA_MLP	SARIMA_MLP	MAdA_SARIMA_MLP
MAdA	0.428	0.339	-	0.261
SARIMA	0.572	-	0.377	0.297
RNA	-	0.661	0.623	0.442
Variance Explained	85.37%	79%	75.52%	71.26%

Source: Prepared by authors.

Once the model prediction combinations are obtained, we can calculate the results of each combination for the adjustment set and for the forecast set, and then select the best one for the forecast series. Table 9 shows the results of the accuracy measurements for the adjustment and prediction of the combined models.

Table 9

Results of calculations of the accuracy measures for the prediction of the combined models

Combined models	RMSE	MAPE	UTHEIL
MAdA_SARIMA/MD	725.569	0.462	0.619
MAdA_MLP/MD	570.831	0.372	0.487
SARIMA_MLP/MD	447.71	0.274	0.382
MAdA_SARIMA_MLP/MD	560.372	0.358	0.478
MAdA_SARIMA/MG	715.429	0.454	0.61
MAdA_MLP/MG	537.598	0.339	0.459
SARIMA_MLP/MG	436.809	0.258	0.373
MAdA_SARIMA_MLP/MG	530.906	0.332	0.453
MAdA_SARIMA/MH	706.292	0.447	0.602
MAdA_MLP/MH	512.568	0.315	0.437
SARIMA_MLP/MH	430.691	0.252	0.368
MAdA_SARIMA_MLP/MH	506.164	0.307	0.432
MAdA_SARIMA/RL	757.844	0.476	0.646
MAdA_MLP/RL	390.969	0.212	0.334
SARIMA_MLP/RL	<b>368.74*</b>	<b>0.192*</b>	<b>0.315*</b>
MAdA_SARIMA_MLP/RL	371.286	0.198	0.317
MAdA_SARIMA/PCA	706.823	0.447	0.603
MAdA_MLP/PCA	488.075	0.308	0.416
SARIMA_MLP/PCA	418.083	0.253	0.357
MAdA_SARIMA_MLP/PCA	509.813	0.324	0.435

\*The best model for each measure of accuracy.

Source: Prepared by authors.

Table 9 shows that the combination of forecasts that provided the best result for the forecast was SARIMA\_MLP/RL, highlighted in bold, with the lowest values of accuracy measurements. It should be noted that the regression model was validated by analyzing the residues. This model presented a MAP of 19.2% and UTHEIL of 0.315, so it is verified that the prediction generated by the model is considered satisfactory and better than a naive prediction, because of its proximity to zero.

In the sixth step of research, a comparison of the models found in the research is performed; the comparative models are the best SES model: MAdA, best Box-Jenkins model: SARIMA (2,0,0)(2,0,0)12,

the best model of RNA: RNA/MLP (12,10,1) and the best SARIMA\_MLP/RL combined model. Table 10 shows the comparison between the models.

**Table 10**

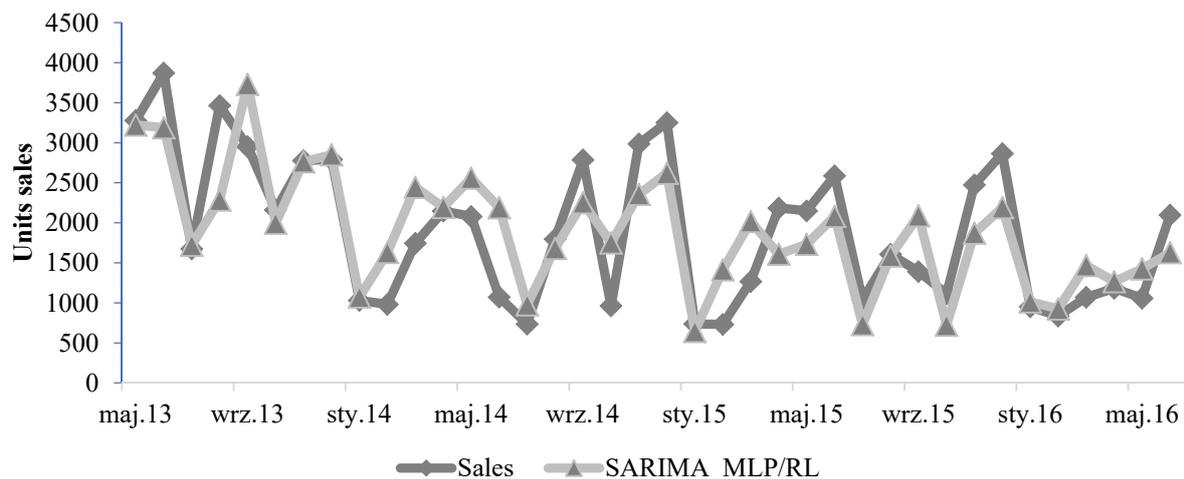
Accuracy measures for comparison between individual and combined models

Models	RMSE	MAPE	UTHEIL
MAdA	1029.571	0.670	0.877
SARIMA (2,0,0)(2,0,0) <sub>12</sub>	726.837	0.442	0.619
MLP (12, 10, 1)	644.985	0.389	0.549
<b>SARIMA_MLP/RL</b>	<b>368.74*</b>	<b>0.192*</b>	<b>0.315*</b>

\*The best model for each measure accuracy.

Source: Prepared by authors.

When comparing the models it was verified that the SARIMA\_MLP/RL combined model highlighted in bold obtained values much smaller than for the other models for all the measurements of accuracy. The model had a 19.2% relative error and a UTHEIL of 31.5. The regression was able to minimize the errors of the residues, for example, a decrease in the prediction MAPE of up to 20%, so the proposed solution is satisfactory. Thus, it is stated that the best model to predict the demand of the housing market in São Paulo is Linear Regression, using the SARIMA model (2,0,0)(2,0,0)<sub>12</sub> and RNA/MLP (12,10,1) as independent variables. Fig. 8 shows the forecast graph generated by the SARIMA\_MLP/RL model compared to the sales series.



**Fig. 8.** Graph of the values predicted by the SARIMA\_MLP/RL model compared to the sales series.

Source: Prepared by authors.

For this research, the data covering the period from January 2004 to June 2016 was used, however, during the survey, sales data was updated by EMBRAESP and SECOVI, with sales data up to October 2017.

In this way, the proposed model is used to forecast the demand for residential housing between July 2016 and December 2017, comparing the months that were not included in the survey, but had already been closed and disclosed the sales, aiming to validate the model. Fig. 9 shows the forecast generated by the updated model compared to the sales series that was updated by EMPRAESP and SECOVI. It should be noted that the forecast generated by the proposed model closely matches the original sales series in the first four months of the second half of 2016.

The model presents a positive trend until the month of December; however, after this, i.e. until February, which make for almost the entire first quarter of 2017, the model shows a decrease in sales. From the second quarter of 2017 onwards, the market begins to recover. In the fourth quarter, we noticed an increase in sales. In addition to the good adjustment to the current sales figures, the result of the model is in line with the updated sales figures. In this way, the model is validated and can be used as an auxiliary tool to predict sales of residential housing in the city of São Paulo.

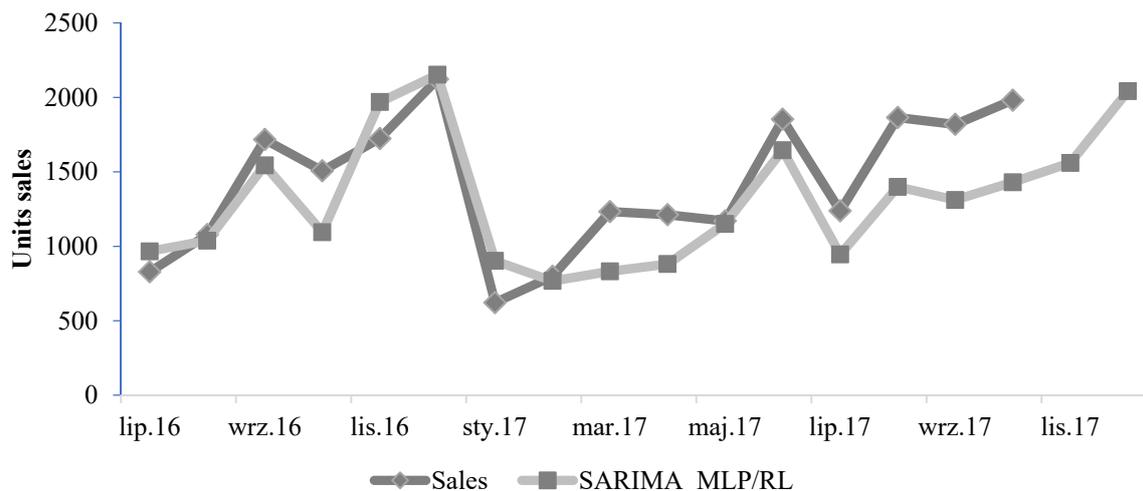


Fig. 9. Forecast generated by the proposed model compared with units sold. *Source:* Prepared by the authors.

The health of the Brazilian housing market has been affected by the financial crisis in recent years, negatively impacting buyers' confidence, who have been more cautious about investing or assuming long-term debt. On the other hand, this scenario was reversed in 2017, generating a cycle of opportunities in the segment for those seeking to invest in real estate, especially in the final stretch of this downward cycle, making the end of 2017 the best time in recent history to buy properties under what were still attractive and advantageous conditions. Driven by the growth trend of late 2017, 2018 started off with high expectations from the sector, mainly due to the promise of falling interest rates. This fact materialized and the market grew 19.2% in 2018 (ISTOÉ, 2019).

Brazil had a period of sharp growth between 2008 and 2014, when property prices soared. In Sao Paulo, the increase was 200% and almost 250% in Rio de Janeiro. This long period of growth was shaken by the late reflection of a world recession, which resulted in real estate stagnation. Banks began a tight credit crunch starting in 2013, cutting funding in half. The Brazilian economic scenario suggests that the 2019-2020 season represents an opportunity for buyers to reach favorable negotiations. These trends are corroborated by the fall in the Selic rate (6.5% in 2018), the decrease in which has a direct impact on real estate credit supply, and controlled inflation should be within the target of 4.25%.

For a long time, the housing market was driven only by speculation and guesswork, that is the expert's feeling. As a result, many investors continued to make mistakes. Although the risk element cannot be eliminated, the use of data analysis, as is the case in this research, enables more informed decisions to be made while minimizing risk (JAFARI, AKHAVIAN 2019). The housing market has a large volume of data (SEHGAL, AGARWAL 2018), which, when well manipulated, can generate information for both housing agencies and large builders.

This research manipulated a time series to prove that, by employing detailed data analysis methods, data can present information that is close to reality. Beyond the investor or housing manager's feeling, companies in the sector can minimize their risks. It should be noted that the manipulated series belongs to the fifth largest city on the planet, and the proposed model can only be used to assist professionals in the São Paulo market; nevertheless, the authors hope that the present study spurs additional research on this topic, leading to further investigation where studies on the real estate market are not as frequent as in Europe and the USA.

### 5. Conclusion

The objective of this study was to propose a forecast model of housing market demand in São Paulo, which would aid housing developers in the decision-making process. To do so, we decided to use 38 individual time series forecasting methods, in addition to the 5 combination prediction methods, generating 20 models combined.

Initially, when using graphic analyses and tests, the presence of tendency and seasonality was affirmed. Based on these results, the eight methods of series with trend and seasonality were used to model the series. In this class, the additive tendency model with additive seasonality and multiplicative error correction - the MA<sub>d</sub>A, obtained the lowest values for the accuracy measurements, presenting 67% of MAPE and UTheil of 0.87.

In the Box-Jenkins methodology, the SARIMA model (2,0,0)(2,0,0) with a size 12 of seasonality, obtained the lowest values of accuracy, i.e. 44.2% of MAPE and UTheil of 0.619. Twenty-four models of Artificial Neural Networks were modeled, and the RNA/MLP model (12,10,1) obtained the least accurate values for prediction error statistics, presenting 38.9% of MAPE and a UTheil of 0.549.

After individual modeling, the models were combined using 5 combination methods. Among the 20 combined models; the combination of the SARIMA model (2,0,0)(2,0,0)12 and RNA/MLP (12,10,1) was the one with the most accurate values. Linear Regression, obtained 19.2% of MAPE and a UTheil of 0.315.

When comparing these methods, it was verified that the combined model generated forecasts superior to the other models, having reduced errors by up to 20%. It was therefore chosen as the best model for predicting housing sales in São Paulo. Finally, this model was used to forecast the sales of the second half of 2016 and the year 2017; the result presented is in line with the updated sales data. Unfortunately, the latest data on housing sales are no longer open, this is a limitation of our research.

Thus, the methodology proved to be efficient for the proposed objective of finding a suitable model for forecasting housing demand. In this way, the combination of the Box-Jenkins models with the MLP/RNA models present advantages that make it interesting to achieve the sales forecast. However, some considerations regarding its use merit comment. The Box-Jenkins and RNA models employed in this research have the advantage of being applied over a wide range of time series. The Box-Jenkins models apply to the forecast of series with seasonality, series with trends and autocorrelated series, in addition to the prediction of series with more than one or all the situations described. On the other hand, the MLP/RNA models, besides these, present the capacity of nonlinear modeling. The combined use of such models, through the combination, makes it possible to capture the various possible patterns in the series, making them interesting for forecasting sales compared to other models in terms of their accuracy.

The study allowed to verify the gain obtained with the analysis of sector data. The difficulty in the practical use of these models and their combinations in organizations for sales forecasting is due to the availability of software and knowledge of forecasting models. From this perspective, the presence of contracted specialists is required, which can generate a greater cost to the planning process. However, the result presented in this research demonstrates that the use of a housing sales forecast model helps developers and builders make decisions on sales and leases, since the model can be used to verify sales trends in the medium term.

We would like to emphasize that the proposed model is valid when there isn't any change or event in the national economy, such as the creation of another housing program. The model can only be used for the city of São Paulo, however processes like this can be used by other researchers to develop models for other cities in Brazil, as well as in other countries.

## 6. References

- ABECIP. Crédito imobiliário com recursos da poupança cresceu 3,4% em 2014. (2015). <https://www.abecip.org.br/imprensa/abecip-na-midia/credito-imobiliario-com-recursos-da-poupanca-cresceu-3-4-em-2014>. Accessed 15 jan. 2019.
- Anghel, I., & Hristea, M. (2015). Some considerations regarding the international real estate market – present and future predications. *Procedia Economics and Finance*, 32, 1442–1452. Advance online publication. [https://doi.org/10.1016/S2212-5671\(15\)01520-8](https://doi.org/10.1016/S2212-5671(15)01520-8)
- Baghestani, H. (2017). Do consumers' home buying attitudes explain the behaviour of US home sales? *Applied Economics Letters*, 24, 779–783. Advance online publication. <https://doi.org/10.1080/13504851.2016.1229401>
- Ball, M., & Nanda, A. (2013). Household attributes and the future demand for retirement housing. *International Journal of Housing Markets and Analysis*, 6(1). <https://doi.org/10.1108/17538271311306002>
- Ballou, R. (2006). *Gerenciamento da cadeia de suprimentos/Logística empresarial* (5th ed.). Bookman.

- Baltrusis, N., & Mourad, L. N. (2014). Política habitacional e locação social em Salvador. *Caderno CRH*, 27(71), 267–284. <https://doi.org/10.1590/S0103-49792014000200004>
- Box, G., Jenkins, G., & Reinsel, G. (2008). *Time series analysis - forecasting and control* (4th ed.). Prentice Hall.
- Bratu, M. (2012). The reduction of uncertainty in making decisions by evaluating the macroeconomic forecasts performance in Romania. *Economic Research Journal*, 25, 239–262. Advance online publication. <https://doi.org/10.1080/1331677X.2012.11517506>
- Chen, L. H. (2010). Using Financial and Macroeconomic Indicators to Forecast Sales of Large Development and Construction Firms. *The Journal of Real Estate Finance and Economics*, 40, 310–331. Advance online publication. <https://doi.org/10.1007/s11146-008-9158-7>
- Csesko, M., & Reed, R. (2009). Will residential aged care facilities meet long-term demand? *Property Management*, 27(1). <https://doi.org/10.1108/02637470910932674>
- Dua, P., & Miller, S. M. (1996). Forecasting Connecticut home sales in a BVAR framework using coincident and leading indexes. *The Journal of Real Estate Finance and Economics*, 13. Advance online publication. <https://doi.org/10.1007/BF00217392>
- Dua, P., Miller, S. M., & Smyth, D. J. (1999). Using Leading Indicators to Forecast U.S. Home Sales in a Bayesian Vector Autoregressive Framework. *The Journal of Real Estate Finance and Economics*, 18, 191–205. Advance online publication. <https://doi.org/10.1023/A:1007718725609>
- Dickey, D., & Fuller, W. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*. Advance online publication. <https://doi.org/10.2307/2286348>
- EMBRAESP. (2013). *Crise estrutural no setor imobiliário ou aonde que se ergueu no mar*. <https://incorporacaoimobiliaria.com/tag/embraesp/>. Accessed 26 dec 2016.
- Gibler, K. M., & Sherwood Clements, J., III. (2011). Testing a forecast model to predict movement of older Americans into retirement housing. *International Journal of Housing Markets and Analysis*, 4(1). <https://doi.org/10.1108/17538271111111811>
- Gibler, K. M., Talvavull, P., Casado-Díaz, J. M., & Rodriguez, V. (2009). Examining retirement housing preferences among international retiree migrants. *International Real Estate Review*, 12(1), 1–22.
- Granger, C. W., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3, 197–204. Advance online publication. <https://doi.org/10.1002/for.3980030207>
- Gupta, R., Lau, C. K.M., Plakandaras, V., & Wong, W. (2019). The role of housing sentiment in forecasting U.S. home sales growth: evidence from a Bayesian compressed vector autoregressive model. *Economic Research-Ekonomska Istraživanja*, 32(1). <https://doi.org/10.1080/1331677X.2019.1650657>
- Hassani, H., Ghodsi, Z., Gupta, R., & Segnon, M. (2017). Forecasting Home Sales in the Four Census Regions and the Aggregate US Economy Using Singular Spectrum Analysis. *Computational Economics*, 49, 83–97. Advance online publication. <https://doi.org/10.1007/s10614-015-9548-x>
- Huang, Y., Wang, Y., & Gai, S. (2011). The application and research of a new combinatorial analysis and forecasting method in real estate area based on Grey System Theory and Multivariate Linear Regression. *Procedia Engineering*, 15, 4532–4537. Advance online publication. <https://doi.org/10.1016/j.proeng.2011.08.851>
- Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18, 439–454. Advance online publication. [https://doi.org/10.1016/S0169-2070\(01\)00110-8](https://doi.org/10.1016/S0169-2070(01)00110-8)
- Hyung, C. J., & Jung, J. U. (2013). Simulation Model on Long-Term Housing Demand Forecasting for the Utilization of Real Estate Knowledge. *Journal of CEO and Management Studies*, 16, 119–139.
- IBGE. (2014). *Pesquisa anual dos serviços*. <http://ibge.gov.br/home/estatistica/economia/comercioeservico/pas/analisepas99.shtm>. Accessed 02 dec. 2018.
- IBGE. (2016). *Série histórica IPCA*. [http://www.ibge.gov.br/home/estatistica/indicadores/precos/inpc\\_ipca/ipca-inpc\\_201601\\_1.shtm](http://www.ibge.gov.br/home/estatistica/indicadores/precos/inpc_ipca/ipca-inpc_201601_1.shtm). Accessed 27 jan. 2019.
- ISTOÉ. *Mercado imobiliário fecha 2018 com alta de 19,2% em vendas, diz CBIC*. (2019). <https://istoe.com.br/mercado-imobiliario-fecha-2018-com-alta-de-192-em-vendas-diz-cbic/>. Accessed 02 jan. 2020. JAFARI, A., AKHAVIAN, R. Driving forces for the US residential housing price: a predictive analysis. Built Environment Project and Asset Management. Article In Press.

- Khashei, M., & Bijari, M. (2010). An artificial neural network (p,d,q) model for time series forecasting. *Expert Systems with Applications*, 37, 479–489. Advance online publication. <https://doi.org/10.1016/j.eswa.2009.05.044>
- Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root? *Journal of Econometrics*. Advance online publication. [https://doi.org/10.1016/0304-4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y)
- Liu, L., & Li, W. (2009). Research on the forecast of Real Estate market demand in Beijing. Paper presentation at International Conference on Construction & Real Estate Management, Beijing, China.
- Mendonça, M. J., Medrano, L. A., & Sachsida, A. (2011). Avaliando o efeito de um choque de política monetária sobre o mercado imobiliário. *Texto para Discussão IPEA*, 1631, 1–31.
- Mendonça, M. J., & Sachsida, A. (2012). Existe bolha no mercado imobiliário brasileiro? *Textos para Discussão IPEA*, 1762, 1–66.
- Mendonça, M. J. (2013). O crédito imobiliário no Brasil e sua relação com a Política Monetária. *Revista brasileira de Economia* 67(4). <https://doi.org/10.1590/S0034-71402013000400005>
- Moro, M. F., Reis, C. C. C., Flores, S., Weise, A., & Vincenzi, S. (2016a). São Paulo Real Estate Market: Interference of economic indicators in sales of residential Real Estates. *International Journal of Real Estate Studies*, 10(2).
- Moro, M., de Almeida Flores, S., dos Reis, C. C. C., Weise, A. D., & Vincenzi, S. L. (2016b). Previsão de vendas: uma adequação da metodologia ARIMA para o mercado imobiliário de São Paulo. Paper presentation at the Conferência Internacional da Latin American Real Estate Society, São Paulo, Brazil. [https://doi.org/10.15396/lares-16-moro\\_previsao\\_artigo](https://doi.org/10.15396/lares-16-moro_previsao_artigo)
- Ord, K., Koehler, A., & Snyder, R. (1997). Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models. *Journal of the American Statistical Association*, 92, 1621–1629. Advance online publication. <https://doi.org/10.1080/01621459.1997.10473684>
- Patton, A., & Sheppard, K. (2009). Optimal combinations of realised volatility estimators. *International Journal of Forecasting*, 25, 218–238. Advance online publication. <https://doi.org/10.1016/j.ijforecast.2009.01.011>
- SECOVI-SP. (2015). Na contramão do mercado imobiliário, VCG cresce 155% em 2014. <https://incorporacaoimobiliaria.com/tag/mercado-imobiliario/>. Accessed 25 jan. 2019.
- SECOVI-SP. (2016). Série histórica com a evolução do IVV e do número de unidades ofertadas, vendidas e lançadas em São Paulo. <http://www.secovi.com.br/pesquisas-e-indices/pesquisa-mensal-do-mercado-imobiliario>. Accessed 15 jan 2017.
- Sehgal, S., & Agarwal, M. (2018). Analogous Examination of Various Machine Learning Algorithm Applied to Big Data. Paper presentation at the International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India. <https://doi.org/10.1109/ICACCCN.2018.8748855>
- Shimbo, L. (2013). O aquecimento imobiliário e o mercado de habitação na área central de São Paulo (2001-2010). *Revista Latinoamericana de Estudios Urbanos e Regionales*, <https://doi.org/10.4067/S0250-71612013000200010>
- Skittides, C., & Fruh, W. (2014). Wind forecasting using Principal Component Analysis. *Renewable Energy*, 69, 365–374. Advance online publication. <https://doi.org/10.1016/j.renene.2014.03.068>
- Stock, J., & Watson, M. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23, 405–430. Advance online publication. <https://doi.org/10.1002/for.928>
- Temür, S. A., Akgün, M., & Temür, G. (2019). Predicting housing sales in turkey using arima, lstm and hybrid models. *Journal of Business Economics and Management*, 20, 920–938. Advance online publication. <https://doi.org/10.3846/jbem.2019.10190>
- UNITED NATIONS. (2014). *World's population increasingly urban with more than half living in urban areas*. [http://www.un.org/en/development/desa/news/population/world-urbanizationprospects2014.html?utm\\_term=United+Nations,+DESA,+Social,+Economic&utm\\_content=DESA+assists+countries+in+agendasetting+and+decisionmaking+with+the+goal+of+meeting+their+economic,+social+and+environmental+challenges.&utm\\_source=twitterfeed&utm\\_medium=twitter](http://www.un.org/en/development/desa/news/population/world-urbanizationprospects2014.html?utm_term=United+Nations,+DESA,+Social,+Economic&utm_content=DESA+assists+countries+in+agendasetting+and+decisionmaking+with+the+goal+of+meeting+their+economic,+social+and+environmental+challenges.&utm_source=twitterfeed&utm_medium=twitter). Accessed 18 mar 2019.

- 
- Valle dos Santos, R. O., & Vellasco, M. M. (2015). Neural Expert Weighting: A NEW framework for dynamic forecast combination. *Expert Systems with Applications*, 42, 8625–8636. Advance online publication. <https://doi.org/10.1016/j.eswa.2015.07.017>
- Varandas Júnior, J., Alencar, C. (2009). Modelos de projeção de demanda aplicado ao mercado imobiliário. Paper presentation at Conferência Internacional da Latin American Real Estate Society, São Paulo, Brazil.
- Wang, L., Wang, Z., Qu, H., & Liu, S. (2018). Optimal Forecast Combination Based on Neural Networks for Time Series Forecasting. *Applied Soft Computing*, 66, 1–17. Advance online publication. <https://doi.org/10.1016/j.asoc.2018.02.004>