

**UNIVERSIDADE FEDERAL DE SANTA MARIA  
CENTRO DE TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM  
ENGENHARIA DE PRODUÇÃO**

**Matheus Fernando Moro**

**MODELO HÍBRIDO DE SÉRIES TEMPORAIS PARA PREVISÃO DE  
DEMANDA DO MERCADO IMOBILIÁRIO DE SÃO PAULO**

Santa Maria, RS, Brasil  
2017



**Matheus Fernando Moro**

**MODELO HÍBRIDO DE SÉRIES TEMPORAIS PARA PREVISÃO DE DEMANDA  
DO MERCADO IMOBILIÁRIO DE SÃO PAULO**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia da Produção, Área de Concentração em Gestão da Produção, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do título de **Mestre em Engenharia de Produção**.

Orientador: Prof. Dr. Andreas Dittmar Weise

Santa Maria, RS, Brasil  
2017

Ficha catalográfica elaborada através do Programa de Geração Automática da Biblioteca Central da UFSM, com os dados fornecidos pelo(a) autor(a).

Moro, Matheus Fernando  
MODELO HÍBRIDO DE SÉRIES TEMPORAIS PARA PREVISÃO DE  
DEMANDA DO MERCADO IMOBILIÁRIO DE SÃO PAULO / Matheus  
Fernando Moro.- 2017.  
146 p.; 30 cm

Orientador: Andreas Dittmar Weise  
Dissertação (mestrado) - Universidade Federal de Santa  
Maria, Centro de Tecnologia, Programa de Pós-Graduação em  
Engenharia de Produção, RS, 2017

1. Previsão de demanda 2. Mercado Imobiliário 3.  
Combinação de previsões I. Weise, Andreas Dittmar II.  
Título.

**Matheus Fernando Moro**

**MODELO HÍBRIDO DE SÉRIES TEMPORAIS PARA PREVISÃO DE DEMANDA  
DO MERCADO IMOBILIÁRIO DE SÃO PAULO**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção, Área de Concentração em Gestão da Produção, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Mestre em Engenharia de Produção**.

**Aprovado em 15 de fevereiro de 2017:**

---

**Andreas Dittmar Weise, Dr. (UFSM)**  
(Presidente/Orientador)

---

**Lynceo Falavigna Braghirolli, Dr. (UFSM)**

---

**Antonio Cezar Bornia, Dr. (UFSC)**

Santa Maria, RS  
2017



## AGRADECIMENTOS

Desafio tão grande quanto escrever essa dissertação é limitar-se a palavras para agradecer as pessoas que fizeram parte dessa minha passagem por Santa Maria e contribuíram, de alguma forma, para a concretização dessa conquista.

Em princípio agradeço a minha família, a qual dedico esse trabalho, minha mãe Ivete e meu pai Nilso, bem como a minha irmã Marcella, obrigado por tudo, principalmente por aguentarem os telefones estressados e peço desculpas pela ausência nesse período.

Agradeço aos guris da confraria dos illuminatis, que desde o primeiro momento aqui nessa cidade me receberam tão bem em sua casa. Os percalços durante esses dois anos aconteceram, mas soubemos contornar bem cada situação e viver com tranquilidade e ao mesmo tempo com muito agito nesse período.

Agradeço ao meu orientador, professor Andreas Dittmar Weise, por esse tempo de aprendizado, discussões acerca de variados assuntos. Obrigado por esse período que trabalhamos juntos e agradeço por sempre acreditar no meu potencial. Agradeço também a professora Silvana, que me ajudou em todos os momentos de dúvida, você foi primordial para que essa dissertação fosse realizada. De forma geral, agradeço a todos os servidores da UFSM que sempre se mostraram dispostos e competentes para sanar qualquer dúvida.

Agradeço as novas amigas que fiz aqui e que me proporcionaram ótimos momentos, agradeço principalmente aos colegas de mestrado, que sempre estiveram prontos para me ajudar de qualquer forma, obrigado por tornar a convivência mais tranquila. Em especial agradeço aos meus brothers Murilo, Wagner, Cyro e Bruno por todos os momentos que me apoiaram, me aguentaram, bem como pelos embalos no Centro de Eventos e outras noites na cidade. De forma mais lisonjeada agradeço ao kaisergelo team pela viagem sensacional, a qual serviu, além de tudo para garantir uma apresentação na defesa tranquila e serena.

E agora, vem a consistência, o apoio, o sólido... sem vocês duas esse mestrado não teria sido o mesmo. Agradeço imensamente a Sandrine e a Camila por todas as situações que passamos nesses dois anos. Foram momentos intensos, de muito compartilhamento de ideias, sentimentos, teorias e conteúdo. Se existe uma parceria que deu certo foi essa de Palmeira das Missões com Medianeira, meus sinceros agradecimentos a vocês duas. Ainda falando de consistência, não posso deixar de expressar em evidência aquela que foi minha base nessa cidade, muitos passaram por mim aqui, mas quem nunca me deixou desistir, que sempre esteve do meu lado - por vezes brigando - marrenta do jeito que é, mas sempre ali quando

precisei, agradeço imensamente a Camila por todas as situações vividas aqui em Santa Maria e pelo apoio incondicional. Do mesmo modo, agradeço meu irmão que Santa Maria me deu, Bruno, obrigado por tudo man, valeu por me aguentar com todas as particularidades, obrigado por me escutar e por me ajudar quando mais precisei, sempre vou lembrar positivamente da nossa amizade.

Por fim agradeço a todos que de alguma forma contribuíram para essa conquista.



## RESUMO

### MODELO HÍBRIDO DE SÉRIES TEMPORAIS PARA PREVISÃO DE DEMANDA DO MERCADO IMOBILIÁRIO DE SÃO PAULO

AUTOR: Matheus Fernando Moro

ORIENTADOR: Prof. Dr. Andreas Dittmar Weise

Um dos maiores problemas associados com o uso de previsões de demanda no apoio à tomada de decisão é a escolha do método de previsão a ser implementado. Nesse contexto, por terem um comportamento diferente dos demais setores, o setor imobiliário possui dificuldade de encontrar métodos corretos para prever sua demanda, de fato, devido ao expressivo intervalo de tempo entre a tomada de decisão do projeto, de investimento e a entrada efetiva do empreendimento na disputa de mercado. Essa complexidade acarreta na escolha de métodos errôneos, ocasionando em grandes estoques de unidades residenciais, gerando altos custos para as construtoras e incorporadoras, como acontece desde 2014 na cidade de São Paulo, o mercado imobiliário mais representativo do país. Diante disso, essa pesquisa tem como objetivo propor um modelo híbrido de séries temporais para previsão de demanda de imóveis na cidade de São Paulo. Para isso, são utilizados dados referentes à série temporal de vendas de unidades residenciais, fornecidos pelo SECOVI-SP. Os modelos de Suavização Exponencial, de Box-Jenkins, de Heterocedasticidade Condicional e de Redes Neurais Artificiais são modelados individualmente, posteriormente estes são combinados por meio de seis técnicas de combinação de previsão. As técnicas utilizadas são Média Aritmética, Média Geométrica, Média Harmônica, Variância Mínima, Regressão Linear e Análise de Componentes Principais. As medidas de acurácia para mensurar os resultados obtidos e selecionar o melhor modelo, são o RMSE, MAPE, MAE e UTheil de previsão. Os resultados mostraram que a Regressão Linear com variável independente sendo a combinação do modelo SARIMA (2,0,0)(2,0,0)12 e MLP/RNA (18,25,1) via Análise de Componentes Principais forneceu um desempenho de previsão satisfatório, com RMSE de 349, 21, MAPE de 17,1%, MAE de 287, 62 e UTheil de 0,298. Assim, demonstrando que a combinação e hibridização de modelos de séries temporais possibilitou um aumento significativo no desempenho de previsão. Por fim, utilizou-se o modelo proposto para previsão da demanda de imóveis entre julho de 2016 a dezembro de 2017. Os resultados foram ao encontro de estimativas de especialistas da área, constatando que em 2017 o mercado imobiliário vai se recuperar, entretanto enquanto estes estimam que o mercado cresça 10% em 2017, o modelo revela um crescimento de 19%.

**Palavras-chave:** Previsão de demanda. Mercado imobiliário. Combinação de previsões.



## ABSTRACT

### HYBRID MODEL OF TIME SERIES FOR FORECASTING DEMAND OF THE REAL ESTATE MARKET OF SÃO PAULO

AUTHOR: Matheus Fernando Moro  
ADVISOR: Prof. Dr. Andreas Dittmar Weise

One of the biggest problems associated with the use of demand forecasts at the support the decision-making is the choice of forecasting method to be implemented. In the context, present a behavior different from other sectors, the real estate market has difficulty in finding correct methods to predict its demand, indeed, due to the significant time interval between the project decision making, investment, and the actual entry of the enterprise in the market dispute. This complexity leads to the choice of wrong methods, resulting in large inventories of residential units, generating high costs for builders and incorporators, as it has since 2014 in São Paulo, the most representative real estate market of Brazil. Therefore, this research aims to propose a hybrid model of time series for forecasting demand of real estate in the city of São Paulo. For this, will be used data referring to the time series of residential units sales, provided by SECOVI-SP. At first, the Exponential Smoothing, Box-Jenkins, Conditional Heteroskedasticity and Artificial Neural Networks models are modeled individually, posteriorly these are combined by means of six forecast combining techniques. The techniques used are Arithmetic Mean, Geometric Mean, Harmonic Mean, Minimum Variance, Linear Regression and Principal Component Analysis. The measures of accuracy to measure the obtained results and to select the best model are the RMSE, MAPE, MAE and UTheil of forecast. The results showed that the Linear Regression with independent variable being the combination of the SARIMA(2,0,0) (2,0,0)<sub>12</sub> and MLP/RNA(18,25,1) models through Principal Component Analysis provided a performance satisfactory prediction, with RMSE of 349.21, MAPE of 17.1%, MAE of 287.62 and UTheil of 0.298. Thus, demonstrating that the combination and hybridization of time series models allowed a significant increase in prediction performance. Finally, we used the proposed model to forecast the demand of real estate between July 2016 and December 2017. The results were in agreement with estimates of specialists in the area, stating that in 2017 the real estate market will recover, however while these estimate that the market grows 10% in 2017, the model shows a growth of 19%.

**Keywords:** Demand forecast. Real Estate Market. Combination of forecasts.



## LISTA DE FIGURAS

Figura 1 - Classificação do mercado imobiliário .....	22
Figura 2 - Unidades lançadas na cidade de São Paulo entre Jan/2004 e Dez/2015.....	27
Figura 3 - Unidades residenciais disponíveis para venda na cidade de São Paulo.....	29
Figura 4 - Unidades residenciais vendidas na cidade de São Paulo .....	30
Figura 5 - Evolução das unidades em estoque entre jan/2004 e dez/2015 .....	31
Figura 6 - Características de uma série temporal .....	33
Figura 7 - Série estacionária com ruído aleatório (a); Série estacionária com ruídos autocorrelacionados; Série não estacionária (c).....	43
Figura 8 - Sequencia de execução do método de Box-Jenkins.....	47
Figura 9 - Série temporal não estacionária na média (a); Série temporal não estacionária na média e na declividade (b) .....	54
Figura 10 - Redução da não estacionariedade de uma série temporal após sucessivas diferenciações.....	55
Figura 11 - Neurônio biológico .....	65
Figura 12 - Modelo de neurônio artificial .....	66
Figura 13 - Modelo estrutural de uma RNA para a previsão de séries temporais.....	69
Figura 14 - A retro propagação do erro em uma rede MLP .....	72
Figura 15 - Etapas para o desenvolvimento da pesquisa.....	84
Figura 16 - Metodologia utilizada na Etapa 2 de modelagem dos métodos de Suavização Exponencial.....	86
Figura 17 - Fluxograma da metodologia para a obtenção do modelo ARIMA para a previsão da série temporal .....	87
Figura 18 - Fluxograma da metodologia utilizada para a obtenção do modelo MLP .....	89
Figura 19 - Metodologia para a combinação das previsões.....	90
Figura 20 - Vendas de imóveis em São Paulo entre Jan/04 e Jun/2016 .....	95
Figura 21 - Decomposição da componente tendência da série temporal de vendas de unidades de imóveis residências na cidade de São Paulo .....	96
Figura 22 - Série temporal de vendas de unidades de imóveis residências na cidade de São Paulo decomposta sem a tendência.....	97
Figura 23 - Decomposição da componente sazonalidade da série temporal de vendas de unidades de imóveis residências na cidade de São Paulo .....	98
Figura 24 - Índices sazonais da série temporal de vendas de unidades de imóveis residências na cidade de São Paulo.....	98
Figura 25 - Comparação das previsões geradas pelos modelos de MAdA comparada à série de Vendas de imóveis residências na cidade de São Paulo .....	101
Figura 26 - Correlograma da FAC e da FACP da série temporal de vendas de unidades residências em São Paulo .....	102
Figura 27 - Correlograma dos resíduos Modelo SARIMA(2,0,0)(2,0,0) <sub>12</sub> .....	105
Figura 28 - Gráfico dos resíduos do modelo SARIMA(2,0,0)(2,0,0) <sub>12</sub> comparado com o tempo .....	105
Figura 29 - Série de vendas e série ajustada pelo modelo SARIMA (2,0,0)(2,0,0) <sub>12</sub> .....	106
Figura 30 - Previsão gerada pelo modelo SARIMA (2,0,0)(2,0,0) <sub>12</sub> comparada com a série de vendas entre Mai/13 e Jun/16.....	107
Figura 31 - Série de vendas de imóveis de São Paulo e série ajustada pelo modelo RNA/MLP (18, 25, 1).....	110
Figura 32 - Previsão gerada pelo modelo RNA/MLP (18,25,1) comparada com a série de vendas entre Mai/13 e Jun/16.....	111

Figura 33 - Gráfico de treinamento da rede MLP(18,25,1) .....	112
Figura 34 - Arquitetura da rede do modelo MLP/RNA(18,25,1) .....	112
Figura 35 - Gráfico dos valores previstos pelo modelo SARIMA_RNA/RL comparado com a série de vendas.....	116
Figura 36 - Gráfico dos valores previstos pelo modelo de RL(SARIMA_RNA/ACP) comparado com a série de vendas .....	117
Figura 37 - Etapas pertinentes a modelagem do modelo proposto RL(SARIMA_RNA/ACP) .....	119
Figura 38 - Previsão gerada pelo modelo proposto.....	120

## LISTA DE QUADROS

Quadro 1 - Resumo dos métodos de suavização exponencial proposta por Taylor .....	40
Quadro 2 - Resumo dos métodos de suavização exponencial com correção de espaço de tempo proposta por Ord et al. (1997).....	41
Quadro 3 - Comparação entre modelos ARIMA e MLP .....	68
Quadro 4 - Configuração das RNAs modeladas no presente estudo.....	108
Quadro 5 - Resultados dos pressupostos do modelo SARIMA_RNA/RL .....	143
Quadro 6 - Resultados dos pressupostos do modelo SARIMA_RNA/RL .....	145





## LISTA DE ABREVIATURAS E SIGLAS

ABECIP	Associação Brasileira das Entidades de Crédito Imobiliário e Poupança
ACF	<i>Autocorrelation Function</i>
ACP	Método de combinação de previsão Análise de Componente Principal
ADF	<i>Augmented Dickey-Fuller test</i>
AIC	<i>Akaike Information Criterion</i>
AR	Autorregressivo
ARCH	Autorregressivo para heterocedasticidade condicional
ARCH-M	Autorregressivo para heterocedasticidade condicional em média
ARIMA	Autorregressivo Integrado de Média Móvel
ARMA	Autorregressivo de Média Móveis
BIC	<i>Bayesian Information Criterion</i>
CAGED	Cadastro Geral de Empregados e Desempregados
CBIC	Câmara brasileira da Indústria da Construção
CUB	Custo Unitário da Construção
EMBRAESP	Empresa brasileira de Estudos de Patrimônio
ENESEP	Encontro Nacional de Engenharia de Produção
FGTS	Fundo de Garantia do Trabalhador Social
FIPE	Fundação Instituto de Pesquisas Econômicas
FIPE/ZAP	Índice FIPE ZAP de Preços de Imóveis Anunciados
GARCH	Autorregressivo para Heterocedasticidade Condicional Generalizado
I	Integrado
IBGE	Instituto Brasileiro de Geografia e Estatística
IGARCH	Autorregressivo para heterocedasticidade condicional Integrado
IMA	Integrado de Média Móvel
JCR	<i>Journal Citation Reports</i>
LARES	Conferência da Sociedade Latina Americana de Mercado Imobiliário
MA	Média Móvel
MAE	<i>Mean Absolute Error</i>
MAPE	<i>Mean Absolute Percentage Error</i>
MG	Método de combinação de previsão Média Geométrica
MH	Método de combinação de previsão Média Harmônica
MLP	<i>Perceptron</i> Multi camadas
MT	Método de combinação de previsão Média Aritmética
PACF	<i>Partial Autocorrelation Function</i>
PIB	Produto Interno Bruto
PMCMV	Programa Minha Casa Minha Vida
RL	Método de combinação de previsão pela Regressão Linear
RMSE	<i>Root Mean Square Error</i>
RNA	Rede Neural Artificial
RNN	Rede Neural Natural
SARIMA	Autorregressivo Integrado de Média Móvel Sazonal
SES	Suavização Exponencial Simples
SEH	Suavização Exponencial de Holt
SEHW	Suavização Exponencial de Holt-Winters
SBPE	Sistema brasileiro de Poupança e Empréstimo
SECOVI-SP	Sindicato das Empresas de Compra e Venda de Imóveis em São Paulo

SFH	Sistema Financeiro de Habitação
SINDUSCON-SP	Sindicato da Indústria da Construção Civil do Estado de São Paulo
VAM	Método de combinação de previsão pela Variância Mínima
VAR	Vetor Autorregressivo
VG	Valor Geral de Vendas

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>13</b>
1.1	CONTEXTUALIZAÇÃO E PROBLEMÁTICA .....	13
1.2	OBJETIVOS .....	15
<b>1.2.1</b>	<b>Objetivo Geral</b> .....	<b>15</b>
<b>1.2.2</b>	<b>Objetivos Específicos</b> .....	<b>15</b>
1.3	JUSTIFICATIVA .....	16
1.4	ESTRUTURA DA PESQUISA.....	19
<b>2</b>	<b>REFERENCIAL TEÓRICO</b> .....	<b>21</b>
2.1	MERCADO IMOBILIÁRIO .....	21
<b>2.1.1</b>	<b>Mercado Imobiliário brasileiro</b> .....	<b>21</b>
<b>2.1.2</b>	<b>Mercado Imobiliário de São Paulo</b> .....	<b>25</b>
2.2	PREVISÃO DE DEMANDA.....	31
<b>2.2.1</b>	<b>Modelos de Previsão de Demanda</b> .....	<b>34</b>
2.2.1.1	<i>Modelos de Suavização Exponencial</i> .....	34
2.2.1.1.2	Modelo de Suavização Exponencial Simples .....	35
2.2.1.1.3	Modelo de Suavização Exponencial de Holt .....	37
2.2.1.1.4	Modelo de Suavização Exponencial Sazonal de Holt-Winters.....	38
2.2.1.2	<i>Modelos de Box-Jenkins</i> .....	42
2.2.1.2.1	Conceitos básicos para percepção dos modelos Box-Jenkins.....	42
2.2.1.2.2	Modelos Autorregressivos .....	47
2.2.1.2.3	Modelos de Médias Móveis .....	50
2.2.1.2.4	Modelos mistos Autorregressivos e de Médias móveis .....	52
2.2.1.2.5	Modelos não estacionários .....	53
2.2.1.2.6	Modelos Sazonais .....	57
2.2.1.2.7	Identificação do Modelo .....	58
2.2.1.2.8	Estimativa dos parâmetros e verificação dos modelos.....	59
2.2.1.2.9	Previsão .....	60
2.2.1.3	<i>Modelos de Heterocedasticidade Condicional</i> .....	60
2.2.1.4	<i>Modelos de Redes Neurais Artificiais</i> .....	64
2.2.1.4.1	Aspectos conceituais sobre Redes Neurais Artificiais .....	64
2.2.1.4.2	Previsão de séries temporais utilizando Redes Neurais Artificiais.....	67
2.2.1.5	<i>Combinação de previsão</i> .....	72
<b>2.2.2</b>	<b>Medidas de Acurácia</b> .....	<b>75</b>
2.2.2.1	<i>Medidas estatísticas padrão</i> .....	75
2.2.2.2	<i>Estatística U de Theil</i> .....	77
2.3	PREVISÃO DE DEMANDA NO MERCADO IMOBILIÁRIO .....	78
<b>3</b>	<b>METODOLOGIA</b> .....	<b>83</b>
3.1	CLASSIFICAÇÃO DA PESQUISA .....	83
3.2	OBTENÇÃO DE DADOS .....	84
3.3	ETAPAS DA PESQUISA .....	84
3.4	DELIMITAÇÕES E LIMITAÇÕES DA PESQUISA .....	92
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b> .....	<b>95</b>
4.1	CARACTERÍSTICAS DA SÉRIE TEMPORAL .....	95
4.2	MODELAGENS DA SÉRIE TEMPORAL .....	99
4.3	COMBINAÇÕES E MODELAGEM HÍBRIDA .....	113
4.4	COMPARAÇÃO DOS MÉTODOS .....	118
4.5	PREVISÃO .....	119
<b>5</b>	<b>CONCLUSÃO</b> .....	<b>123</b>

5.1	CONCLUSÕES ACERCA DOS RESULTADOS OBTIDOS .....	123
5.2	SUGESTÕES DE PESQUISA FUTURAS .....	125
	<b>REFERÊNCIAS .....</b>	<b>127</b>
	<b>ANEXO A - TABELA DE FORMULAÇÃO DOS MODELOS DE SUAVIZAÇÃO EXPONENCIAL POR MEIO DE MODELOS DE ESPAÇO DE ESTADO</b>	<b>142</b>
	<b>APÊNDICE A – ANÁLISE DOS RESÍDUOS DO MODELO SARIMA_RNA/RL.....</b>	<b>143</b>
	<b>APÊNDICE B – ANÁLISE DOS RESÍDUOS DO MODELO RL(SARIMA_RNA/ACP) .....</b>	<b>145</b>

# 1 INTRODUÇÃO

Nesta seção inicial são expostas as premissas conceituais sobre as quais o trabalho será desenvolvido, apresentando a contextualização e problemática do tema, justificativa da relevância do assunto e os objetivos propostos. Por fim, é exposta a estruturação do desenvolvimento do trabalho, bem como as etapas percorridas para que os objetivos propostos fossem alcançados.

## 1.1 CONTEXTUALIZAÇÃO E PROBLEMÁTICA

Na dinâmica do atual mercado, as empresas confrontam-se continuamente com questões críticas para assegurar a competitividade futura e seu crescimento organizacional (JOZEFczyk; GASIOR, 2015). Dentro dessa dinâmica que envolve as organizações, o processo de tomada de decisão gerencial, desempenha um papel formidável, influenciando a competitividade e os resultados. Deste modo, orientações quanto a diretrizes futuras de fatores e variáveis que influenciam os planejamentos nos diferentes horizontes de tempo são primordiais (VANCLAY, 2015; MARTINS; 2015; HANSSON et al., 2016).

Uma instrução correta aumentará as chances de sucesso das empresas, evitará perigos futuros, ajudará na manutenção de vantagens competitivas, influenciará as organizações em escolhas e investimentos corretos, e favorecerá o aproveitamento de oportunidades vindouras (ADYA et al., 2001; ARMNSTRONG; COLLOPY; YOKUM, 2005; BREM; MAIER; WIMSCHNEIDER, 2016). Para obter-se um planejamento estratégico eficiente nas empresas, são necessárias a identificação e a previsão correta das mudanças em potencial no ambiente de negócios, o que torna a previsão de demanda um elemento fulcro na tomada de decisão (DUUS, 2016). Para melhorar seus resultados, as empresas podem antecipar problemas e, assim, desenvolver planos para responder a estes problemas, bem como serem capazes de antecipar grande volume de vendas (STAUDT; GONÇALVES; RODRIGUEZ, 2015; CECATTO; BELFIORE, 2015; VAN DONSELAAR et al., 2016).

Ballou (2006) destaca que a prática da previsão é um importante fator no processo decisório da organização e que prever níveis de demanda é fundamental à empresa como um todo. Esse processo fornece as entradas básicas para planejamento e controle de todas as áreas funcionais (HOFER; EISL; MAYR, 2015; HAJEJ; REZG; ALI, 2016). Diante do exposto, ganhos de produtividade e eficiência em qualquer mercado exigem transparência e fartura de informações para os diversos agentes que nele atuam. Indicadores que espelhem o

comportamento de preços e volumes das transações ao longo do tempo são de extrema relevância para que empresas e investidores tomem suas decisões, e se posicionem com mais segurança na arena de mercado (OSADCHIY; GAUR; SESHADRI, 2013; WU, ZHANG; CHEN, 2015).

Nos mercados de *real estate* esta necessidade é ainda mais premente. Em qualquer um dos seus segmentos há um expressivo intervalo de tempo entre a tomada de decisão, de investimento e a entrada efetiva do empreendimento na disputa de mercado. As condições de atratividade percebida no momento da aquisição do terreno e formatação de certo empreendimento pode ser diversa quando do seu lançamento e exposição à concorrência e ao seu público alvo (GIBLER; CLEMENTS, 2011; HYUNG; JUNG, 2013; BALTRUSIS; MOURAD, 2014).

A compreensão dessa dinâmica neste mercado proferido econômico requer a utilização de métodos de investigação adequados ao estudo do mercado imobiliário. As incorporadoras usualmente utilizam estudos de demanda de mercado elaborados por consultores ou imobiliárias, entretanto, raramente são executados estudos de oferta e demanda no Brasil para o desenvolvimento de pesquisas, principalmente no meio científico. Ao se referir a estudos de imóveis no Brasil, é imprescindível citar a cidade de São Paulo, por ser a maior cidade e o mais importante centro comercial do Brasil, possui o mercado imobiliário mais representativo do país e serve de base para todas as pesquisas e estudos envolvendo imóveis, bem como exerce poder de transmissão de demanda imobiliária sob outras metrópoles dentro do país (MORO et al., 2016).

Diante do exposto, o presente estudo visa contribuir para a ampliação da literatura sobre essa área, como também seus achados possuem implicações práticas para as construtoras, incorporadoras, tal como os investidores. Assim sendo, este trabalho se propõe realizar uma análise apurada sobre os métodos de previsão de demanda e métodos de combinação de previsão, para modelá-los e assim sugerir um modelo que possa auxiliar aos profissionais da área imobiliária na tomada de decisões em suas operações e empreendimentos.

Sob esta ótica, esta dissertação tem como problema de pesquisa responder o seguinte questionamento: Qual método de previsão de demanda, utilizando série temporal tem a melhor acurácia, ou seja, métodos individuais ou métodos de combinações de previsões são mais adequados para prever a demanda do mercado imobiliário da cidade de São Paulo?

## 1.2 OBJETIVOS

Nesta seção são explanados os objetivos desta pesquisa, ou seja, a finalidade pela qual se realizará este estudo. Este tópico divide-se em: objetivo geral e objetivos específicos.

### 1.2.1 Objetivo Geral

O objetivo geral dessa dissertação é propor um modelo de séries temporais, utilizando combinações de previsões para a previsão de demanda do mercado imobiliário da cidade de São Paulo.

### 1.2.2 Objetivos Específicos

Para atingir o objetivo geral, os seguintes objetivos específicos foram delineados:

- a) Analisar a série temporal de unidades vendidas no mercado imobiliário de São Paulo a fim de encontrar características da série: tendência, sazonalidade, padrão cíclico ou aleatório;
- b) Estimar modelos de Suavização Exponencial, de Box-Jenkins, de Heterocedasticidade Condicional e os modelos de Redes Neurais Artificiais de previsão de séries temporais;
- c) Combinar os modelos de previsões de séries temporais com melhores acurácias por meio da Média Aritmética, Média Geométrica, Média Harmônica, Variância Mínima, Regressão Linear e Análise de Componentes Principais;
- d) Averiguar qual o método de melhor acurácia entre modelos individuais e modelos combinados; e
- e) Utilizar o modelo de melhor acurácia para prever a demanda do mercado imobiliário de São Paulo para o ano de 2017, e propô-lo como modelo de previsão de demanda para o mercado imobiliário paulistano.

### 1.3 JUSTIFICATIVA

O tema desta dissertação é a previsão de demanda no mercado imobiliário de São Paulo por meio de métodos quantitativos, utilizando série temporal de vendas de unidades de imóveis.

Lindberg e Zackrisson (1991) e Hofer, Eisl e Mayr (2015) afirmam que um dos maiores problemas associados com o uso de previsões no apoio à tomada de decisões é a escolha do método aplicado para gerar a previsão. Segundo Armstrong (2001), pesquisas que contribuem para o desenvolvimento e aprimoramento de guias para seleção de métodos de previsão são sempre úteis, independente da área.

Utilizar as ferramentas de previsão sabiamente requer conhecimento sobre onde cada tipo de ferramenta trabalha de forma eficiente (MOON et al., 1998; TERRA; PASSADOR, 2012; JALES; SILVA, 2016). Um dos problemas principais que os métodos de previsão deparam-se é que, enquanto há muitos métodos para escolher, há pouca indicação sobre quais são mais eficientes em cada situação (LYNN; SCHNAARS; SKOV, 1999; MORETTIN; TOLOI, 2006; HOFER; EISL; MAYR, 2015). Kahn (2002) diz que o uso impróprio de métodos de previsão sugere a falta de conhecimento acerca da previsão e, de modo correspondente, indica a necessidade de direcionadores sobre quando usar determinado método de previsão.

É nesse sentido que o objetivo dessa pesquisa se insere: propor um modelo híbrido de séries temporais para a previsão de demanda para o mercado imobiliário de São Paulo, tema o qual nota-se uma escassez de pesquisa com foco em métodos de previsão para o mercado de imóveis, isso pode ser verificado por meio de uma bibliometria realizada pelo autor.

Para garantir a consistência dos dados, sua relevância e integridade, os termos considerados para a pesquisa são: (a) Mercado Imobiliário (no inglês, *Real Estate*); (b) Previsão de demanda (no inglês, *Demand Forecast*); e (c) Combinação (no inglês, *Combining*), pesquisando no título, resumo e palavras-chave no período que compreende os anos de 2007 a 2016. A escolha destas palavras-chaves se destina a concentrar a pesquisa no foco de interesse dessa pesquisa, que é a proposição de um modelo de combinação de previsão de demanda para o mercado imobiliário da cidade de São Paulo.

A busca foi realizada nas principais bases de dados e em dois anais de eventos, o evento mais importante do mercado imobiliário na América Latina, a Conferência da Sociedade Latina Americana de Mercado Imobiliário (LARES) e o maior encontro relacionado ao curso que essa pesquisa de pós-graduação está inserida, o Encontro Nacional



de Engenharia de Produção (ENEGEP). A escolha das bases é justificada pelo fato de que somadas, as bases *Web of Science*, *Scopus*, *Science Direct* e *Emerald* disponibilizam acesso a cerca de 32 mil periódicos, sendo que somente a *Science Direct* contém mais de 25% de toda a informação científica publicada mundialmente, bem como essas bases possuem grande número de periódicos que publicam na área de Engenharias III, segundo critérios do portal *WebQualis*, além de que estes serem periódicos de alto fator de impacto JCR (*Journal Citation Report*). Concomitantemente a escolha da base Scielo deve-se ao fato que as principais revistas brasileiras estão indexadas nessa base. Na Tabela 1 é apresentado o número de publicações por cada base e nos dois anais de eventos.

Tabela 1 - Publicações nos últimos 10 anos nas bases de dados e anais em estudo

Combinações	Bases de dados				Eventos		
	<i>SCOPUS</i>	<i>Science Direct</i>	<i>Web of Science</i>	<i>Emerald</i>	Scielo	ENEGEP	LARES
<i>Real Estate</i>	6127	1089	6985	3047	42	19	343
<i>Demand Forecast</i>	621	2488	7892	238	44	112	9
<i>Combining</i>	172097	193349	868009	8543	3092	85	38
<i>Real Estate + Demand Forecast</i>	1	5	40	18	0	0	2
<i>Demand Forecast + Combining</i>	21	275	796	31	0	0	0
<i>Real Estate + Combining</i>	69	56	347	176	0	0	0
<i>Real Estate + Demand Forecast + Combining</i>	0	0	0	0	0	0	0

Fonte: Elaborado pelo autor.

Um número considerável de trabalhos é encontrado pesquisando cada palavra-chave separadamente, ao cruzá-las percebe-se uma escassez de trabalhos, principalmente para *Real Estate + Demand Forecast* e *Real Estate + Demand Forecast + Combining*, nesta última combinação não se encontrou trabalhos, sendo essas três palavras-chaves desse projeto de pesquisa. Salienta que para a base Scielo e os eventos ENEGEP e LARES foram pesquisados os termos no idioma português e para as outras bases em inglês. Nas bases pesquisadas em português não encontrou pesquisas publicadas com nenhuma das combinações de palavras, com exceção da LARES, a qual encontraram-se dois trabalhos com a combinação *Real Estate + Demand Forecast*.

O surpreendente número de apenas dois artigos no maior evento de mercado imobiliário da América Latina chama atenção e transcreve que o campo tem espaço pra pesquisa pelo quase ineditismo, ao menos em âmbito continental. Um dos trabalhos

encontrados na LARES é o de Varandas Júnior e Alencar (2009), os autores realizam uma síntese bibliográfica das literaturas referentes aos modelos de regressão de variáveis por meio de séries históricas, focados em rotinas aplicadas a projeção de demanda por unidades habitacionais. Já o número de zero artigo encontrado no maior evento de Engenharia de Produção no Brasil não traduz o crescente espaço que o mercado imobiliário vem sendo discutido na área de Engenharia de Produção.

Corroborando com isto, uma pesquisa realizada pelo autor nos anais do ENEGEP mostra que entre os anos de 2005 e 2010, seis artigos referentes ao mercado imobiliário foram publicados, já entre os anos de 2011 e 2016 foram quinze artigos, mostrando o crescente interesse pela área.

É importante salientar que mesmo sendo encontrada essa quantidade de artigos nas combinações, analisando cada trabalho em separado é possível afirmar que praticamente sua totalidade é formada por previsão de preços no mercado imobiliário e somente em três pesquisas, Varandas Júnior e Alencar (2009), Liu e Li (2009) e Gibler e Clements (2011), citam especificamente métodos de previsão de demanda no mercado imobiliário. Os autores supracitados são unânimes em dizer que não existem métodos de previsão de demanda consistentes aplicáveis em mercados imobiliários de países emergentes e em processo de mudança de pirâmide etária. Sobre esta ótica, somado ao fato do resultado da combinação das três palavras-chave não ter encontrado pesquisas sobre o tema, corresponde que há uma lacuna a ser estudada.

Corroborando com isso, o cenário macroeconômico brasileiro (país emergente e em mudança na pirâmide etária), dominante nos últimos anos, de quedas abruptas dos juros, crescimento econômico e elevação da classe C para B, incentivou o desenvolvimento do mercado imobiliário, a partir da valorização dos aluguéis e dos preços dos imóveis tanto residenciais como comerciais. Porém, nos dois últimos anos, de 2014 e 2015, este cenário começou a sofrer algumas mudanças, influenciando diretamente na área imobiliária.

Em um estudo de mercado habitacional, devem ser levados em consideração aspectos relativos ao produto ofertado e aos fatores geradores de demanda e de formação dos preços. Assim, há a necessidade de se fazer uma previsão da demanda pelo produto durante o planejamento do empreendimento.

Uma previsão de demanda sensata auxilia os gestores das incorporadoras na tomada de decisão, sobre o número de projetos, lançamentos e estoque. Além disso, o conhecimento do funcionamento do mercado habitacional permite ao governo, por exemplo, realizar

simulações macroeconômicas objetivando a geração de empregos, ou ainda auxiliar o planejamento de políticas voltadas à habitação popular.

Com todas estas peculiaridades, a coleta de informações a respeito do comportamento do mercado imobiliário possui grande importância para o planejamento e gerenciamento de projetos imobiliários. Porém, como visto, no Brasil existem poucos trabalhos acadêmicos a respeito do comportamento do mercado imobiliário.

Os estudos quantitativos sobre habitação no Brasil, embora frequentes, estão focados em estimativas e disposição de estoques e preços, sendo assim, por intermédio das técnicas de previsão, que são o objeto de estudo dessa dissertação, é possível extrair dos dados passados disponíveis sobre um processo de demanda, informações que permitirão a modelagem de seu comportamento. A hipótese de uma continuidade nesse comportamento permite a realização de previsões, cuja qualidade e precisão são muito superiores aquelas das previsões realizadas com base intuitiva, baseadas unicamente na experiência dos analistas, característica dos modelos de previsão de demanda qualitativo. De forma adicional, os modelos são atualizáveis e, uma vez atualizados, eles passam de imediato, a refletir as alterações do processo, fornecendo prontamente informações às novas tomadas de decisão.

Vê-se, então, que o uso da previsão viabiliza, nas incorporadoras e até para o setor público, a realização de tomadas de decisão mais ágeis e com maior acurácia, as quais se refletirão em maior velocidade de resposta, em menores perdas e, portanto, em uma maior competitividade no mercado para as empresas.

#### 1.4 ESTRUTURA DA PESQUISA

O presente estudo está dividido em cinco capítulos. O primeiro apresenta a introdução, contextualizando e apresentando a problemática, bem como seus objetivos e a justificativa da escolha do tema. Ademais, abordaram-se sucintamente os principais tópicos da pesquisa, ressaltando a importância de um estudo apurado sobre o tema.

Na sequência, mais precisamente no Capítulo 2 é exposto o referencial teórico do estudo, a fim de proporcionar um embasamento literário consistente. Para isso, este se encontra dividido em três subcapítulos mestre, sendo o primeiro tratando sobre o mercado imobiliário, o segundo sobre os métodos de previsão de demanda e o terceiro realizando uma união entre os dois primeiros. Em primeiro momento explana-se sobre o mercado imobiliário nacional, sua classificação e tendências. Posteriormente remete-se ao mercado imobiliário de São Paulo, apresentando a importância deste para o Brasil, bem como comentando o

comportamento deste mercado em relação aos lançamentos, estoques e vendas. No segundo subcapítulo apresenta-se os métodos de previsão de demanda utilizando séries temporais, mostrando as vantagens e desvantagens de cada modelo, bem como suas formulações. Findando o Capítulo 2, trabalhos encontrados no meio científico abordando previsão no mercado imobiliário são citados.

No Capítulo 3 apresenta-se um relato detalhado sobre metodologia utilizada na pesquisa, ou seja, classificação da pesquisa, a forma de coleta dos dados, as etapas da pesquisa, bem como as delimitações e limitações.

O Capítulo 4 dedica-se aos resultados e discussão do presente estudo, sendo este dividido em 5 subcapítulos referentes a cada objetivo específico. O primeiro versa sobre as características da série, o segundo apresenta as modelagens e previsões dos modelos de Suavização Exponencial, Box-Jenkins, Heterocedásticos Condicional e Redes Neurais Artificiais. O terceiro subcapítulo destina-se a combinação das previsões e modelagem híbrida, em seguida os modelos individuais e combinados são comparados e por fim o modelo de menor erro é utilizado para previsão.

Finalmente, o Capítulo 5 alude às conclusões e reflexões obtidas com a presente pesquisa. Este se desdobra em dois subcapítulos, o primeiro trata de conclusões acerca dos dados obtidos com as análises realizadas, enquanto o segundo propõe ideias para futuras pesquisas.

## **2 REFERENCIAL TEÓRICO**

Neste capítulo apresentar-se-á uma breve exposição sobre o mercado imobiliário brasileiro, bem como o mercado imobiliário específico da cidade de São Paulo. Em uma segunda parte são explanados os métodos de previsão por séries temporais e, por fim, são apresentados alguns trabalhos disponíveis sobre a previsão de demanda no mercado imobiliário.

### **2.1 MERCADO IMOBILIÁRIO**

Esta seção está dividida em duas partes, a primeira abordando o mercado imobiliário brasileiro em geral e a segunda especificamente sobre o mercado imobiliário paulistano.

#### **2.1.1 Mercado Imobiliário brasileiro**

Baptistella (2005) articula que o mercado é o cenário onde são realizadas as trocas de direitos ou de bens tangíveis e intangíveis. Sendo assim, Braulio (2005) caracteriza-o por uma competição acirrada, com participação espontânea, onde as transações são vastas e independentes, com numerosas ofertas, visto que, a retirada de amostras não afetam os negócios.

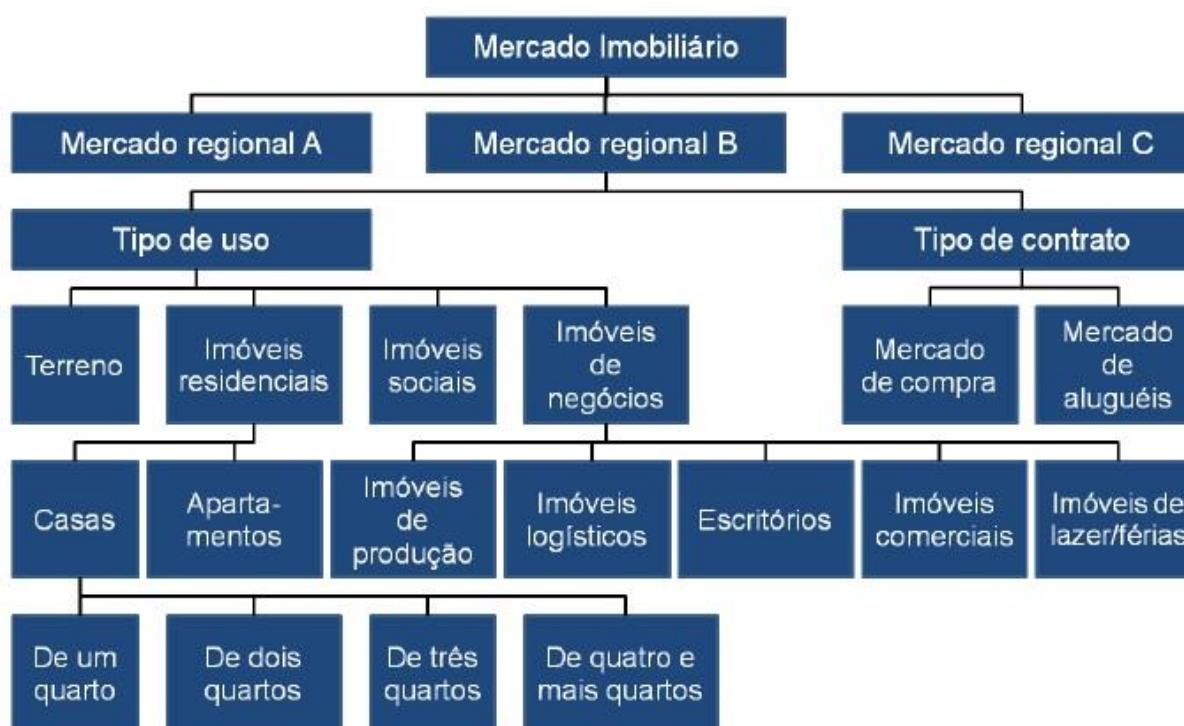
Para González e Formoso (2003) e Siqueira (2013), o mercado imobiliário tem um comportamento muito diferente dos mercados de outros bens. As distinções decorrem em relação às características especiais dos imóveis, dentre elas podem-se citar: a longevidade, a fixação espacial, a heterogeneidade, o elevado prazo de maturação, o alto custo das unidades e as interferências das leis municipais, estaduais ou federais (GONZÁLEZ, FORMOSO, 2003; WEISE, 2009).

Dentro desse contexto, o mercado imobiliário brasileiro engloba produtos que são representados pelos imóveis, que por sua vez, são compostos por materiais e sistemas (MENDONÇA, 2013). Além desses produtos, têm-se planejamentos, construções e comercializações envolvendo diferentes profissionais gerando emprego e renda. Afirma-se assim, que o mercado imobiliário movimentava diferentes campos da sociedade sendo um dos setores com maior capacidade de impulsionar a economia (MATOS; BARTKIW, 2013).

O mercado imobiliário e todos seus componentes estão interligados ao setor de serviços, o qual possui papel importante na economia brasileira, pois representa mais da

metade do Produto Interno Bruto (PIB) do país (IBGE, 2014), sendo que somente o mercado imobiliário foi responsável por 9,7% do PIB nacional de 2014 (ABECIP, 2015). Porém, o mercado de imóveis é constituído de uma série de mercados menores, referente a regiões específicas, como também, de mercados específicos para tipos de imóveis, sendo que cada qual, apresenta características temporais e transações complexas de serem mensuradas e previstas. Dessa forma, faz-se necessário uma melhor definição do setor, onde os imóveis urbanos são classificados de acordo com a sua natureza, essa classificação pode ser observada na Figura 1.

Figura 1 - Classificação do mercado imobiliário



Fonte: Brauer (2011 apud MEDEIROS, 2015, p. 27).

Dentro dessa classificação apresentada, o presente estudo está inserido no mercado de imóveis residenciais, abrangendo casas e apartamentos.

Nesse sentido, puxado pelo crescimento e estabilidade da economia nacional, o mercado imobiliário residencial viu-se diante de um colosso de oportunidades por meio das favoráveis condições conjunturais (CANDIDO, 2012). O aquecimento do mercado de trabalho levou a taxa de desemprego a patamares mínimos e elevou a renda do trabalhador. Esse fator, somado à queda da taxa de juros, intensificou a capacidade de financiamento de

longo prazo para a aquisição e construção de empreendimentos imobiliários, efeito energizado com uma política de crédito direcionada por parte do governo por meio de programas de obras públicas e subsídios para o setor (MENDONÇA; MEDRANO; SCHSIDA, 2011; IPEA, 2012; MENDONÇA, 2013).

Todo esse efeito na política monetária facilitou aquisição de um imóvel, e permitiu a uma grande fatia da população conseguir a tão almejada casa própria (ARRAES; SOUSA FILHO, 2008; OLIVEIRA, 2011; MENDONÇA, 2013). Diante do exposto, o processo nacional de aquecimento do mercado imobiliário, em especial, os lançamentos residenciais pertencentes ao “segmento econômico” que compreende os imóveis residenciais com valores de até R\$ 200.000,00 (em dezembro 2010), destinados às famílias com renda mensal entre três e dez salários mínimos, ou seja, aos setores de renda média e média baixa, ganhou expressividade, sobretudo, a partir de 2006, quando uma injeção de recursos públicos e privados alavancou a produção de moradias pelo mercado, um exemplo disso é o Programa Minha Casa Minha Vida (SHIMBO, 2013; SILVA; ALVES, 2014).

O processo de estabilização da economia brasileira, principiado no Plano Real, consolidou-se nos anos 2000 e repercutiu em crescimento econômico acarretando em uma considerável crescente no poder de consumo da população, em especial. Ao mesmo tempo, o país passou a integrar, mais fortemente, às operações do mercado financeiro mundial (PAULANI, 2009; MENDONÇA, 2013). Por um lado, esse crescimento do poder de consumo possibilitou o aumento do número de sujeitos passíveis de obterem créditos, dentre eles, o habitacional. Por outro, essa faixa de consumo foi “descoberta” pelas grandes empresas construtoras e incorporadoras. E esse consumo foi incentivado pelo poder público via crédito habitacional, potencializado pela entrada de capital financeiro nessas grandes empresas (ROCHA LIMA JR; MONETTI; ALENCAR, 2011; SHIMBO, 2013; SOLA, 2014).

No âmbito do Estado, a “nova” política nacional de habitação elaborada em 2004 pelo primeiro governo de Luiz Inácio Lula da Silva (2003-2006), visava ampliar maciçamente os recursos de fundos públicos e semi públicos para os financiamentos habitacionais e arregimentar outros fundos para esse fim, na perspectiva de aumentar a produção subsidiada de habitação (para as faixas de menor renda) e de fomentar o mercado privado de habitação (BRASIL, 2004).

Como resultados dessa política, Shimbo (2013) cita que por um lado, houve um aumento exponencial nos valores de financiamento habitacional contratados pelo Sistema Financeiro Habitacional (SFH), contando com recursos do Fundo de Garantia do Tempo de

Serviço (FGTS) e do Sistema Brasileiro de Poupança e Empréstimo (SBPE). Sola (2014) diz que em 2003, quando se iniciou o primeiro governo Lula, cita que o valor total contratado pelo SFH era de aproximadamente cinco bilhões de reais e em 2008, na metade do segundo governo, esse valor excedeu quarenta bilhões de reais. Em 2009, o governo federal formalizou o incentivo à produção privada de moradias ao lançar o Programa “Minha Casa, Minha Vida” (PMCMV), com a meta de construir um milhão de unidades habitacionais, injetando um montante de recursos ainda maior do que nas linhas de financiamento anteriores.

Shimbo (2013) e Sola (2014) citam que em 2011, na segunda etapa do programa, a promessa foi de mais dois milhões de unidades. Outras medidas regulatórias também favoreceram as atividades imobiliárias e o setor da construção civil, como, por exemplo, relacionadas à segurança jurídica na transação de imóveis e à desoneração de impostos (SOLA; PORTO, 2014).

O conjunto total das estratégias econômicas e políticas adotadas pelo governo federal favoreceu, em grande medida, o forte aquecimento do mercado habitacional no Brasil. Porém, no ano de 2012, o mercado nacional fechou em queda em relação aos anos anteriores, os principais mercados imobiliários registraram queda nas vendas de unidades, bem como no lançamento de unidades, por intermédio do Valor Geral de Vendas (VGV Lançado), sendo apurado um valor 7% menor que no ano de 2011. Lopes (2013) explica que o motivo dessa queda, deve-se a uma aceleração muito forte nos lançamentos nos últimos anos e em 2012 o mercado passou por um processo de ajuste. Destaca-se também a redução no crédito imobiliário concedido tanto em quantidade quanto em valores, primeira vez desde 2002.

Em 2013 o cenário foi favorável, os lançamentos aumentaram 24% e o VGV Lançado aumentou 41,4% em relação a 2012. Um fator que influenciou positivamente para a alta dos índices do mercado imobiliário em 2013 foi o aumento do limite de utilização do FGTS para a aquisição de imóveis. Em outubro de 2013 o governo federal decidiu elevar o valor máximo de compra de imóveis com recursos do FGTS de R\$ 500 mil para R\$ 750 mil nos estados do Rio de Janeiro, São Paulo, Minas Gerais e no Distrito Federal. Nos demais estados, passou para R\$ 650 mil. Isso fez com que muitas pessoas tivessem mais facilidade para comprar imóveis que antes não se encaixavam no valor limite para utilização do benefício.

Em 2014, a procura por imóveis teve grande queda, os fatores que explicam essa queda são diversos. Com a estagnação da economia, PIB 0,1%, a condição econômica da população ficou instável, os índices de juros e desemprego aumentaram e a oferta de crédito diminuiu. Isso levou a população a ficar receosa e não investir com muita tranquilidade em imóveis próprios, como vinha fazendo. Ressalta-se que outros mercados também sofrem as



consequências da crise econômica, mas que isso fica mais claro no setor imobiliário, em função de sua ascensão contínua nos últimos quatro anos. Finalizando o ano, os estoques aumentaram e as vendas caíram 25% (EL PAÍS, 2015).

Em 2015, de janeiro a novembro houve queda de 5,7% das vendas nos imóveis novos. O acumulado de janeiro a novembro de 2015 foi o pior resultado registrado desde 2004, o que explica esse fato é a conjunção de fatores econômicos e institucionais, os quais vêm influenciando negativamente o mercado imobiliário desde 2014. Para o SECOVI-SP (2015a) essa situação somente será amenizada se houver clara intenção do governo de recolocar a economia nos eixos, possibilitando a volta da confiança e dos investimentos.

Por ser a maior cidade do Brasil, o município de São Paulo não ficou imune a esse processo, sendo que o mercado imobiliário da cidade é o mais representativo do país e serve de base para todas as pesquisas e estudos envolvendo imóveis (SHIMBO, 2013; MENDONÇA, 2013). Na próxima seção é apresentado um breve referencial sobre o mercado imobiliário paulistano.

### **2.1.2 Mercado Imobiliário de São Paulo**

Nos anos que antecederam 2013 houve grande ascensão na demanda por imóveis no Brasil e principalmente na cidade de São Paulo (SHIMBO, 2013), mas a estagnação da economia nacional nos últimos dois anos acabou refletindo também na demanda de imóveis na maior cidade brasileira. Mais especificamente nessa cidade, o mercado imobiliário tem um notável destaque, com uma expectativa de aproximadamente 12 milhões de habitantes (IBGE, 2015), é uma das regiões do país mais urbanizadas e de alta densidade demográfica, com uma arrecadação que beira os 500 bilhões de reais, o que representou sozinho 11,4% do PIB brasileiro em 2012.

Por um processo de conurbação, o município de São Paulo e arredores formam uma grande região metropolitana. A região metropolitana de São Paulo, segundo dados do IBGE, compreende uma área de 7.947 km<sup>2</sup>, dividida em 39 municípios, com população estimada em quase 20,2 milhões de habitantes. Está-se falando de um local que por si só é mais populosa que qualquer outro estado brasileiro, e se fosse um país figuraria entre os sessenta mais populosos do mundo.

Com todos esses fatores, trata-se sem dúvida de um mercado que possui condições muito favoráveis para a prática de transações imobiliárias, bem como o acompanhamento

destas por meio de estudos e pesquisas acerca dos comportamentos e fenômenos nele presentes (CANDIDO, 2012).

Diante do exposto, ressalta-se ainda que hoje, estão listadas na BOVESPA, cerca de 19 empresas com atuação na indústria da construção, metade delas tem origem na cidade de São Paulo (BM&FBOVESPA, 2016). O Sindicato da Indústria da Construção do Estado, figura como o maior sindicato patronal desse segmento em toda América Latina, tendo associadas cerca de 1374 empresas (SECOVI-SP, 2016a). Todas essas informações mostram um pouco da relevância e grandiosidade do mercado imobiliário dessa região. Tal posição de destaque atrai a atenção de renomadas instituições públicas e privadas que realizam estudos e apresentam indicadores desse mercado como referência para o país. Algumas dessas informações disponibilizadas com integridade e confiança serão elementos cruciais para o desenvolvimento desse trabalho.

Uma dessas importantes informações é o indicador FIPE/ZAP, mensurado pela Fundação Instituto de Pesquisas Econômicas (FIPE), a partir deste, pode ser observado o comportamento dos preços dos imóveis para a cidade de São Paulo. Corroborando, o Sindicato das Empresas de Compra e Venda de Imóveis em São Paulo (SECOVI-SP) disponibiliza os dados acerca da oferta e demanda de imóveis para esse mercado. Já o Sindicato da Indústria da Construção (SINDUSCON SP) é responsável por apurar o custo unitário básico do metro quadrado de construção (CUB), indicador de custo amplamente utilizado como referência. O somatório desses fatores reforça a escolha desse município como objeto de estudo desse trabalho.

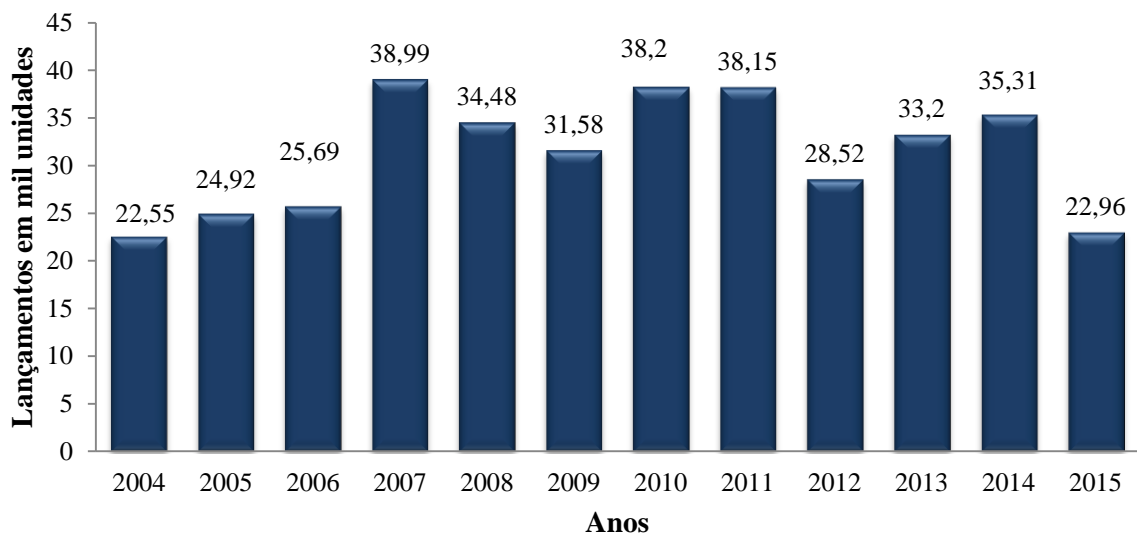
Os dados apresentados por essas instituições mostram que entre 2006 e 2010, houve um acentuado aumento no número de lançamentos residenciais, superando todos os recordes anteriores desde a década de 1980. Segundo dados da Empresa Brasileira de Estudos do Patrimônio (EMBRAESP, 2013), os picos desse momento de elevação intensa da produção imobiliária ocorreram em 2007 e 2010 quando, respectivamente, foram lançadas aproximadamente 62 e 67 mil unidades residenciais. A área central do município de São Paulo também fez parte desse boom imobiliário.

Entre 1991 e 2006, de acordo com Silva (2006), não foram constatadas alterações significativas na dinâmica imobiliária nos distritos da área central, em relação aos lançamentos de empreendimentos residenciais e comerciais novos. Até então, ainda de acordo com Shimbo (2013), nenhum dos instrumentos urbanísticos voltados à atração de empreendimentos imobiliários privados na área central teve o efeito esperado em relação aos objetivos esperados. Entretanto, entre 2007 e 2010, houve um ciclo de elevação de oferta

desses empreendimentos e mudanças quanto aos tipos e preços dos produtos imobiliários nessa área.

Já para Cândido (2012) a partir de 2004, com estímulos governamental e cenário internacional favorável pode ser percebida uma mudança no rumo do mercado imobiliário paulistano. Como reflexo, os empresários do setor voltaram a investir e proporcionaram um forte aumento no número de lançamentos de empreendimentos imobiliários. Enquanto entre 2004 e 2006 a média de lançamentos ficou em 24.385 unidades, no período compreendido entre 2007 e 2011 pode ser observada uma média de 36.280 unidades, uma expansão de 49% no número de lançamentos. A Figura 2 abaixo mostra a quantidade de unidades lançadas anualmente na região metropolitana de São Paulo.

Figura 2 - Unidades lançadas na cidade de São Paulo entre Jan/2004 e Dez/2015



Fonte: Elaborado pelo autor com dados provenientes do SECOVI-SP (2016b).

É importante comentar que em 2012 houve uma queda no número de imóveis lançados, cabendo ressaltar que 80% do número de lançamento do mesmo ano foram no quarto trimestre tendendo para um crescimento para os próximos anos - a redução da atividade econômica de forma geral contribuiu para os resultados tímidos da economia brasileira e, conseqüentemente, do mercado imobiliário nacional, neste ano (SECOVI-SP, 2015b) - tendência mostrada nos número dos dois anos seguintes, 2013 e 2014, quando houve aumento considerável.

Dados do SECOVI-SP (2016b) revelam que no ano de 2014, foram lançadas 35,31 mil unidades em São Paulo, esse número representou um aumento de 7% em relação a 2013. Nos

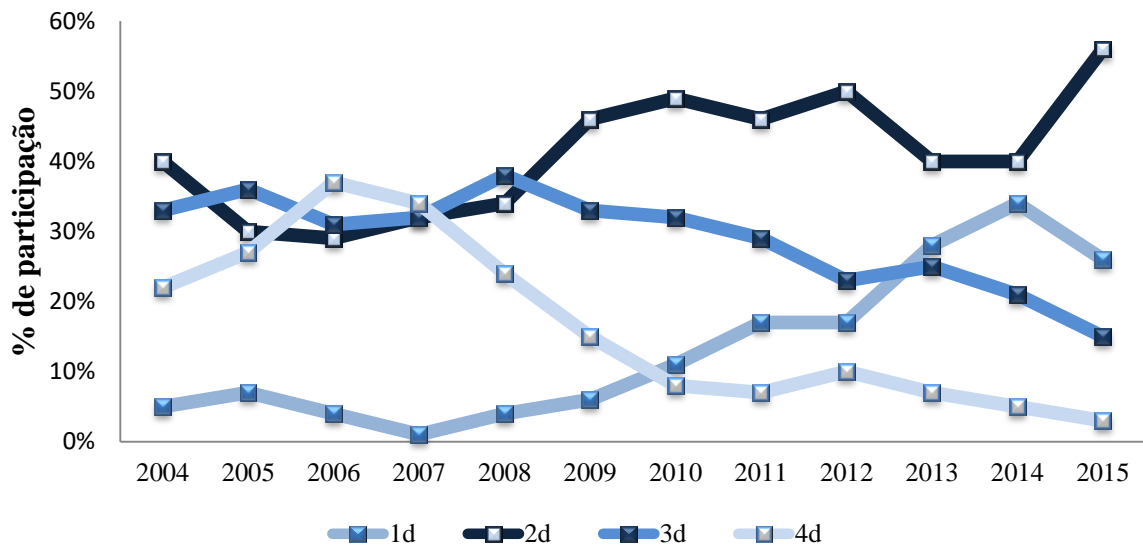
meses de novembro e dezembro, os lançamentos tiveram um melhor resultado, 11 mil unidades lançadas. Os dois últimos meses de 2014 corresponderam a 34,6% do total lançado durante o ano. Essa porcentagem é normal, pois as incorporadoras de capital aberto aproveitam o incremento na renda dos trabalhadores com o 13º salário, e também a necessidade de cumprir metas, para lançarem mais unidades nesses últimos meses (SECOVI-SP, 2015c).

De acordo com dados da SECOVI-SP (2015c), a zona Oeste foi a região com a maior quantidade de lançamentos, com 8.112 unidades, correspondente a 26% do total da cidade, seguida pela Zona Sul, com 7.599 unidades, Leste, com 6.799 unidades, Centro, com 5.058 unidades e Norte, com 4.111 unidades. Já no ano de 2015 é visível a queda de unidades lançadas, para o SECOVI-SP (2015a) essa queda reflete o ambiente desafiador que o país atravessa, em razão das crises política, econômica e institucional. No terceiro trimestre de 2015, o PIB com ajuste sazonal encolheu 1,7% em relação ao segundo trimestre e 4,4% em relação ao mesmo trimestre do ano de 2014 (SECOVI-SP, 2015c).

No ano de 2015, aproximadamente 800 mil trabalhadores com carteira assinada perderam o emprego, segundo os dados do Cadastro Geral de Empregados e Desempregados (CAGED), um dos setores mais atingidos foi o da construção civil, que demitiu mais de 255 mil trabalhadores em 2015 e que, no acumulado de 12 meses (novembro/14 a outubro/15), fechou mais de 442 mil vagas, conjuntamente com esses fatores negativos, a inflação fechou o ano de 2015 em 10,67 % (IBGE, 2016). Essa conjunção de elementos negativos derrubou não somente os lançamentos, mas também as ofertas de unidades na capital paulista.

A oferta de imóveis é outro fator interessante a observar, focando na mudança no perfil das unidades ofertadas no setor no período da recente expansão do mercado. Enquanto 1% das unidades ofertadas em 2007 era de imóveis de um dormitório (1d), em agosto de 2014 esse número foi de surpreendentes 34%. De 2010 para 2011 houve um crescimento de 54% no volume de lançamentos com imóveis de 1d (SECOVI-SP, 2016b). Em contrapartida, os imóveis de quatro dormitórios (4d) que em 2006 representavam 37% das unidades ofertadas, em 2014 chegou a um patamar de 5% de todos os imóveis ofertados no mercado paulistano. Importante salientar ainda a brusca queda de ofertas de imóveis de três dormitórios (3d) e a crescente da oferta de imóveis com dois dormitórios (2d). A Figura 3 mostra a evolução da oferta de imóveis disponíveis no mercado para cada tipo de unidade.

Figura 3 - Unidades residenciais disponíveis para venda na cidade de São Paulo



Fonte: Elaborado pelo autor com dados provenientes do SECOVI-SP (2016b).

A mudança no perfil das unidades ofertadas pode estar relacionada primeiramente com o lançamento do PMCMV em 2007, onde unidade de 1d e 2d constituíram-se como principal tipo de produto. Em outra análise, a mudança no perfil de unidades lançadas no mercado também se relaciona a fatores demográficos como aumento no número de divórcios, maior número de jovens solteiros, aumento da renda individual, entre outros (DENT Jr., 2010).

O efeito do programa estatal de incentivo ao financiamento de imóveis voltado para o público de baixa renda também pode ser visto ao analisarmos as regiões da capital paulista onde foram mais marcantes a aceleração no número de lançamentos. As regiões de maior crescimento entre 2007 e 2011 foram a zona leste e o centro, em ambas os lançamentos buscaram ofertar unidades com o perfil anteriormente citado.

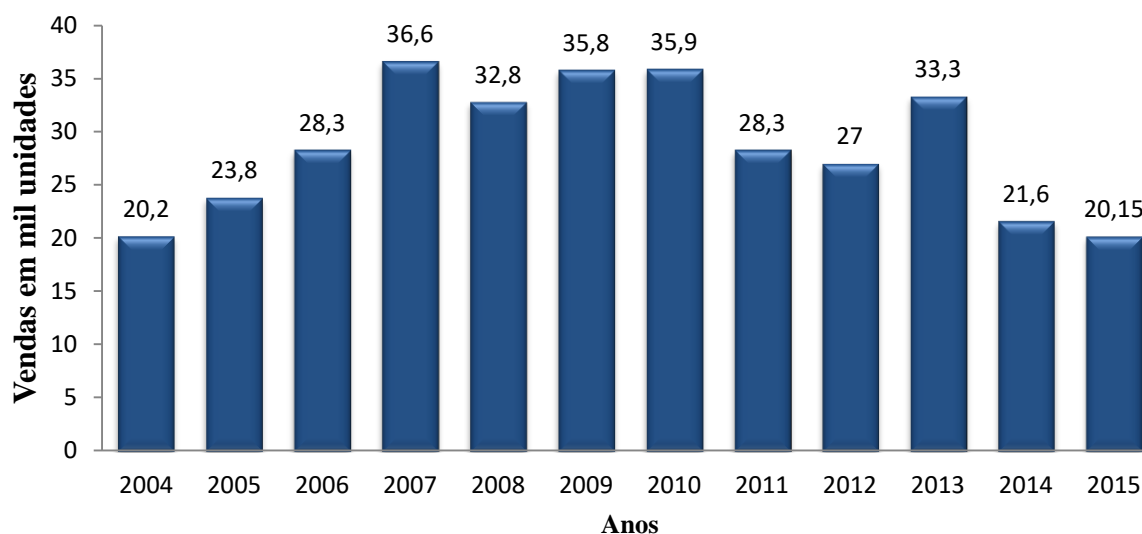
O aquecimento do setor que o levou a relembrar patamares semelhantes ao da década de 90 pode ainda em muito ser atribuído às melhorias nas condições de financiamento, queda na taxa de juros, redução do desemprego, aumento da renda, e de forma geral, maior confiança da sociedade na economia brasileira.

É importante reforçar que o aquecimento do mercado imobiliário não se deu apenas nos empreendimentos do tipo residencial, os lançamentos de conjuntos comerciais acompanharam o ritmo de crescimento, dando um salto mais significativo em 2010 e 2011, onde as empresas do setor lançaram 5,8 mil conjuntos em 2010, 60% superior ao ano anterior, e 7,3 mil conjuntos em 2011 (SECOVI-SP, 2016b). O aumento no número de lançamentos

reflete a confiança dos empresários e investidores no setor, mas também é resultado do aumento na demanda por esse tipo de bem.

No que tange a demanda, o aumento na procura por imóveis e as melhores condições para a realização do sonho da casa própria mostraram-se uma realidade nos últimos anos. De acordo com o SECOVI-SP (2016b), o ano de 2007 foi o ano que mais imóveis foram comercializados, foram vendidos 36.600 imóveis, esse valor foi puxado pela implantação do PMCMV. Já no ano de 2008, teve uma pequena queda, porém nos dois anos seguintes, 2009 e 2010 houve aumento puxado pela oferta de crédito. Já nos anos seguintes teve queda, explicado pelo o que chama de ajuste de mercado. No ano de 2015, a queda é evidente e a explicação é a mesma para a queda de lançamento: em razão das crises política, econômica e institucional, bem como, demissões na construção civil e alta taxa de inflação. A Figura 4 mostra a evolução das vendas de imóveis residências apenas na cidade de São Paulo.

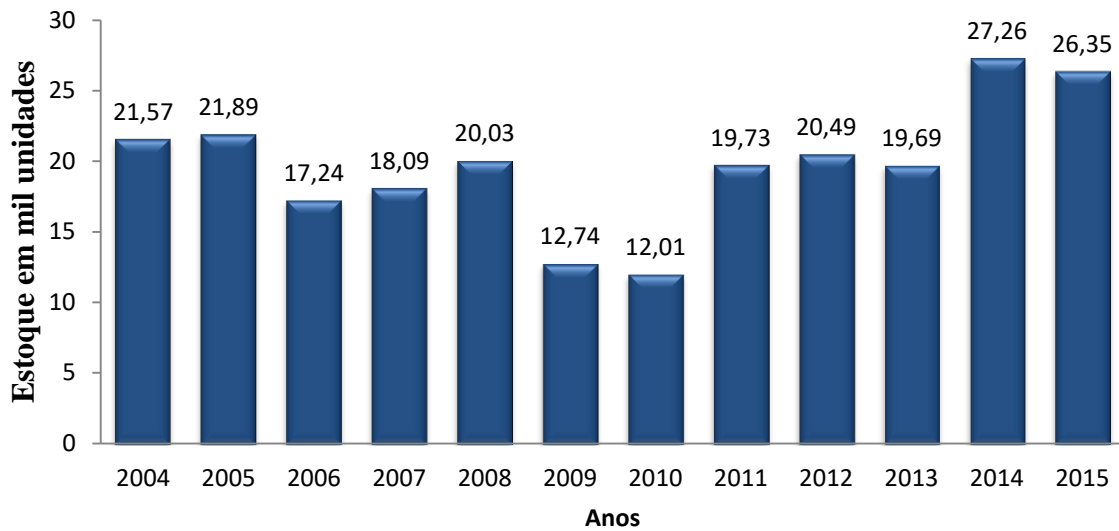
Figura 4 - Unidades residenciais vendidas na cidade de São Paulo



Fonte: Elaborado pelo autor com dados provenientes do SECOVI-SP (2016b).

O sucesso de vendas possibilitou a muitos empresários durante algum tempo conseguir esgotar estoques, vender unidades ainda em fase de projeto, comercializar um empreendimento inteiro em apenas um final de semana. Um sucesso para os negócios imobiliários e para os investidores do setor (CANDIDO, 2012). Mas o que era um sucesso, nos últimos dois anos vem se tornando problema, o número de unidades estocadas na cidade de São Paulo nos anos de 2014 e 2015 foram recordes (SECOVI-SP, 2016b). Na Figura 5 pode-se observar a evolução das unidades estocadas entre 2004 e 2015.

Figura 5 - Evolução das unidades em estoque entre jan/2004 e dez/2015



Fonte: Elaborado pelo autor com dados provenientes do SECOVI-SP (2016b).

De 2010 para 2014, o estoque na cidade, principal mercado habitacional do País, cresceu 126% e as vendas acompanharam o pessimismo, uma queda de 40%. O cenário atual, de escassez de poupança e dificuldades de liberação de crédito pelos bancos, faz com que as pessoas pensem muito antes de tomar a decisão de fazer um financiamento que impacta de 20% a 30% do orçamento ao longo de 20 anos ou mais.

Com muito estoque, aumenta os custos operacionais das incorporadoras, as quais não podem baixar os preços dos imóveis novos, justamente pelos custos provenientes dos mesmos, sendo assim, se as incorporadoras utilizassem métodos de previsão de demanda adequados poderiam prever esse comportamento atual e assim teriam capacidade de estabilizar os lançamentos imóveis. Dentro desse contexto na próxima seção será iniciada a discussão sobre a previsão de demanda.

## 2.2 PREVISÃO DE DEMANDA

Para a introdução do estudo sobre previsão de demanda precisa-se inicialmente entender o conceito de demanda, que nada mais é do que a procura por determinado bem ou serviço (TUBINO, 2004; MOREIRA, 2011). Para Kotler e Armstrong (2009), as demandas são desejos por produtos específicos, respaldados pela habilidade e pela disposição de comprá-los. No entanto, afirmam ainda que os desejos tornam-se demandas apenas quando

apoiados pelo poder de compra. Para Moreira (2011), planejar é uma atividade comum a qualquer tipo de empresa, independentemente do tamanho, ou do ramo que atua, e a previsão de demanda, é a base para elaboração deste planejamento.

De acordo com Gaither e Frazier (2004), prever ou estimar demandas futuras de produtos e/ou serviços e os recursos necessários para produzi-los é o primeiro passo da etapa do planejamento, sendo assim, o ponto de partida para a elaboração das demais previsões da empresa está na estimativa das vendas futuras. Dentro disso, Ritzman e Krajewski (2008) citam que a previsão é a avaliação de acontecimentos futuros, utilizada para fins de planejamento. Os autores afirmam ainda que as previsões são necessárias para auxiliar na tomada de decisão, como a determinação dos recursos necessários, na programação dos recursos existentes e na aquisição de recursos adicionais.

Neste contexto, as ações das organizações dependem de tomadas de decisões, as quais se baseiam em oportunidades de mercado, em fatores contextuais, e no desenvolvimento interno de recursos financeiros, humanos, produtivos e tecnológicos (KOH; LEE, 2015). As previsões de demanda de produtos e serviços auxiliam as tomadas de decisões fornecendo as informações básicas para planejamento e controle de todas as áreas funcionais das organizações, incluindo logística, marketing, produção e finanças (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998; BALLOU, 2001; ALBANEZ, 2015).

É primordial que dentro das organizações a projeção dos níveis de demanda proporciona elementos para o planejamento e controle das áreas envolvidas (BALLOU, 2006). Os gestores demandam de previsões de curto e longo prazo tanto para tomar decisões estratégicas quanto para responder às questões mais contíguas da empresa (GAITHER; FRAIZER, 2005; HOFER; EISL; MAYR, 2015). Dentro disso, Martins e Laugeni (2005) definem previsão de demanda como um processo sistemático para a determinação de informações futuras, baseadas em modelos estatísticos, matemáticos ou econométricos, bem como em modelos subjetivos apoiados em uma metodologia de trabalho. Dentro dos modelos estatísticos, os modelos quantitativos mais utilizados são os que utilizam séries temporais.

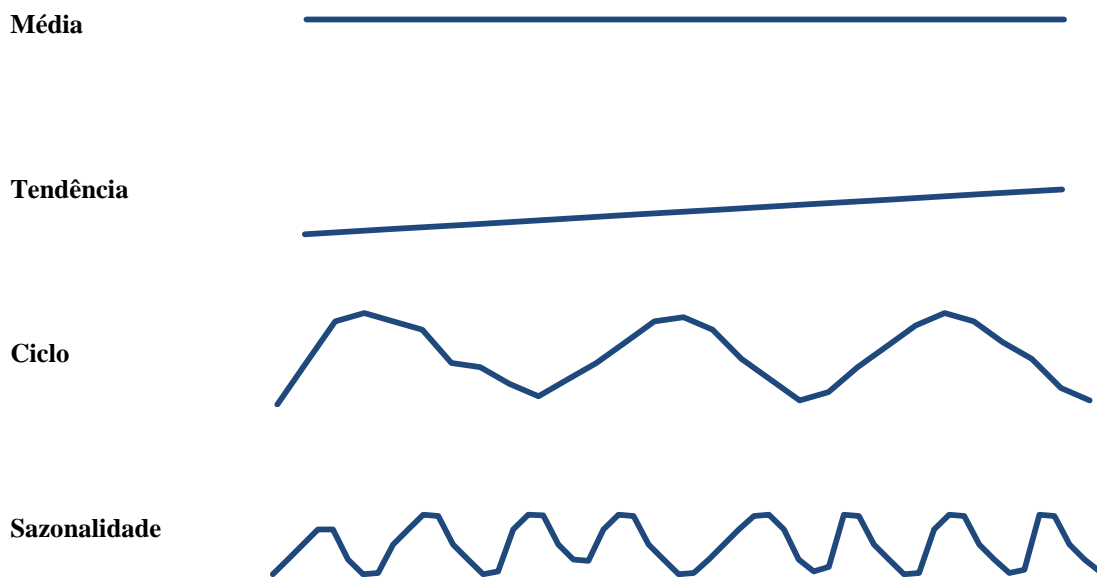
Para Morettin e Tolo (2006), uma série temporal é um processo estocástico gerador de observações no tempo de uma determinada variável a qual representa medições sucessivas de algum fenômeno de interesse. O principal objetivo da análise de séries temporais é investigar o mecanismo gerador dos dados, descrever seu comportamento por meio da construção de gráficos para verificação da existência de tendência, ciclos e variações sazonais (MOORE et al., 2006; LEE; TONG, 2011). Por fim, objetiva-se fazer previsões de valores futuros que servem para tomar decisões. O processo de obtenção dos valores previstos é muito mais que a



posição simplista de estimar valores, pois é necessária a construção de um modelo adequado do sistema a ser tratado. Os métodos de previsão, em geral, baseiam-se na ideia de que observações passadas trazem informações (memória) sobre o padrão do comportamento da série temporal (PELLEGRINI; FOGLIATTO, 2001; SLACK; CHAMBERS; JOHNSTON, 2009).

A previsão de demanda utilizando métodos quantitativos pode ser feita por vários modelos matemáticos. O emprego de cada modelo depende basicamente do comportamento da série temporal que se deseja analisar. Uma série temporal pode exibir até quatro características diferentes em seu comportamento: média, sazonalidade, ciclo e tendência (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998; MORETTIN; TOLOI, 2006). Estas características estão exemplificadas na Figura 6.

Figura 6 - Características de uma série temporal



Fonte: Adaptado de Makridakis, Wheelwright e Hyndman (1998).

Em geral para detecção de tendência e sazonalidade, utiliza-se a análise visual por meio da observação do gráfico da série original, do mesmo modo é possível realizar a decomposição da série em suas características e então encontrar padrões. Entretanto, para afirmar as hipóteses de que a série possui alguma característica, utilizam-se os testes de Wald-Wolfowitz, o teste de Cox-Stuart e o teste de Mann-Kendall. Já para sazonalidade os testes

mais utilizados são o teste de Kruskal-Wallis e o teste de Friedman. Mais detalhe sobre esses testes podem ser observados em Morettin e Toloi (2006).

### 2.2.1 Modelos de Previsão de Demanda

Nesta seção são apresentados os principais modelos utilizados como métodos quantitativos para previsão de demanda, métodos de combinação e por fim as medidas de acurácia para avaliar e selecionar os métodos. Essa seção está dividida em quatro partes, de modo, que apresenta-se na primeira parte os modelos de Suavização Exponencial, posteriormente os modelos de Box-Jenkins, em uma terceira parte é exposto os modelos de Heterocedasticidade Condicional, em sequência são explanados os modelos Redes Neurais Artificiais e por fim os modelos de combinações de previsões são apresentados.

#### 2.2.1.1 Modelos de Suavização Exponencial

Em sua maioria os métodos de previsão baseiam-se na ideia de que observações passadas contêm informações sobre o padrão de comportamento da série temporal. O propósito dos métodos é distinguir o padrão de qualquer ruído que possa estar contido nas observações e então usar esse padrão para prever valores futuros da série (ZHANG, 2003; LEMOS, 2006; ZOU et al., 2007).

Uma classe dos métodos de previsão que aborda as causas de flutuações em séries temporais, é a das suavizações. Técnicas específicas desse tipo assume que os valores extremos da série representam a aleatoriedade e, assim, por meio da suavização desses extremos pode-se identificar o padrão básico (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998; MORETTIN; TOLOI, 2006). A grande popularidade atribuída aos métodos de suavização é devida à simplicidade, à eficiência computacional e à sua razoável precisão (PELLEGRINI, 2000). Nas próximas seções são apresentados os modelos para séries localmente constantes, os adequados para séries que apresentam tendência e para séries sazonais, bem como os com correção de espaço de estado.

##### 2.2.1.1.1 Modelos de Médias Móveis

A Média Móvel caracteriza-se em uma técnica simples de suavização exponencial de séries temporais, calculando uma média aritmética (localmente constante) das  $r$  observações

mais recentes da série, sempre desprezando a observação mais antiga e incluindo a observação mais recente, sendo apresentada na Equação 1 (MORETTIN; TOLOI, 2006).

$$M_t = \frac{Z_t + Z_{t-1} + \dots + Z_{t-r+1}}{r} \quad (1)$$

A previsão de valores futuros é dada pela última média móvel calculada, isto é  $\hat{Z}_t(h) = M_t$ . As desvantagens deste método de previsão estão relacionadas ao fato do método não trabalhar muito bem com séries com tendência e sazonalidade, pois a previsão para um novo período envolve sempre a adição de novos dados e a desconsideração de dados anteriores (DAVIS; AQUILANO; CHASE, 2001; KRAJEWSKI; RITZMAN, 2002; LUSTOSA et al., 2008).

Uma alternativa a este tipo de falha no método é a média móvel ponderada onde ao invés das observações entrarem e saírem abruptamente do intervalo de períodos usado na média, elas são gradualmente ponderadas dentro da média (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998; KRAJEWSKI; RITZMAN, 2002; LUSTOSA et al., 2008). A desvantagem na utilização da média móvel ponderada é a necessidade de conhecimento para determinar os pesos a serem utilizados (DAVIS; AQUILANO; CHASE, 2001).

#### 2.2.1.1.2 Modelo de Suavização Exponencial Simples

O modelo de Suavização Exponencial Simples (SES) ajusta-se melhor às séries de padrão horizontal, isto é, àquelas que não apresentam tendência nem sazonalidade e é apropriado para previsões de curto prazo (PELLEGRINI, 2000; ALLEMÃO, 2004, GONÇALVES, 2007).

O método obtém a previsão para um período futuro ajustando a previsão do período atual com o erro de previsão. A previsão do período  $t+1$  é igual à previsão do período  $t$  mais o ajuste para o erro que ocorreu na previsão do período  $t$ , conforme apresentado na Equação 2. A forma geral do método de SES é apresentada na Equação 3, onde é alocado um parâmetro de suavização  $\alpha$  aos valores mais recentes de  $Y_t$ , e um peso  $1-\alpha$  para previsões mais recentes (WINTERS, 1960; MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998; HOLT, 2004). A forma expandida da Equação 3 é apresentada na Equação 4, onde a previsão está expressa em função de todos os dados da série temporal (WINTERS, 1960).

$$F_{t+1} = F_t + \alpha(Y_t - F_t) \quad (2)$$

$$F_{t+1} = \alpha Y_t + (1 - \alpha)F_t \quad (3)$$

$$F_{t+1} = \alpha \sum_{i=0}^n (1 - \alpha)^i Y_{t-i} + (1 - \alpha)^t F_1 \quad (4)$$

onde  $F_{t+1}$  é a previsão para o período  $t+1$ ,  $F_t$  é a previsão para o período  $t$ ,  $Y_t$  é a demanda realizada no período  $t$ ,  $n$  é o tamanho da série temporal e  $\alpha$  é a constante de suavização com valor entre 0 e 1.

Quanto menor for o valor de  $\alpha$  mais estáveis serão as previsões finais, uma vez que a utilização de baixo valor de  $\alpha$  implica que pesos maiores serão dados às observações passadas e, conseqüentemente, qualquer flutuação aleatória exercerá um peso menor no cálculo de previsão (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998; MORETTIN; TOLOI, 2006).

Para inicializar o método é necessário o valor de  $F_1$ . Tendo este, pode-se proceder de duas maneiras: considerar  $F_1$  igual a  $Y_1$ ; ou utilizar um valor obtido com a média das primeiras observações da série temporal. Um problema do método é a escolha de um  $\alpha$  ótimo, isto pode ser feito por meio da minimização de erros. Assim, elegem-se determinados valores de  $\alpha$ , aplica-se o método e o erro é obtido para diferentes valores de  $\alpha$ . O parâmetro de suavização ( $\alpha$ ) ótimo será aquele com o qual se obtém a melhor acurácia nas previsões (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998; LEMOS, 2006).

Morettin e Tolo (2006) citam que dentre as vantagens de utilizar o modelo de SES está: (a) fácil entendimento; (b) baixo custo de aplicação; e (c) flexibilidade pela variação de  $\alpha$ . Em contrapartida a desvantagem consiste em justamente na complexidade em determinar o valor de  $\alpha$ . O método de SES assume que o padrão de demanda histórica apresenta somente componentes de nível e ruído. Se componentes de tendência e sazonalidade existem nos dados históricos, a Suavização Exponencial Simples não é o método de extrapolação mais indicado (MENTZER; GOMES, 1989; MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998; ALLEMÃO, 2004).

### 2.2.1.1.3 Modelo de Suavização Exponencial de Holt

O método de SES visto na seção anterior, quando aplicado a uma série que apresenta tendência (positiva ou negativa), fornece previsões que subestimam (ou superestimam) continuamente os valores reais. Para evitar esse erro sistemático, um dos métodos aplicáveis é o de Suavização Exponencial de Holt (SEH). Esse método é similar, em princípio, ao Exponencial Simples, a diferença é que em vez de suavizar só o nível, ele utiliza uma nova constante de suavização para “modelar” a tendência da série (MONTGOMERY; JOHNSON; GARDINER, 1990; MORETTIN; TOLOI, 2006).

O modelo de SEH também é conhecido por Suavização Dupla, esse modelo expande o método de Suavização Exponencial Simples para previsões com dados que apresentam tendência linear, mas que não apresentam elementos de sazonalidade (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998, SAMOHYL; SOUZA; MIRANDA, 2008). A previsão com SEH é obtida com o uso de duas constantes de suavização,  $\alpha$  e  $\beta$  (com valores entre 0 e 1, e não relacionados), e das Equações 5, 6 e 7 (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998; TAYLOR, 2003; RASMUSSEN, 2004).

$$\text{Nível:} \quad L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + b_{t-1}) \quad (5)$$

$$\text{Tendência:} \quad b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1} \quad (6)$$

$$\text{Previsão:} \quad F_{t+m} = L_t + b_t m \quad (7)$$

onde  $F_{t+m}$  é a previsão para o período  $t+m$ ,  $m$  é o horizonte de previsão,  $L_t$  é estimativa do nível da série temporal no período  $t$ ,  $b_t$  é a estimativa de tendência da série temporal para o período  $t$  e  $\alpha$  e  $\beta$  são as constantes de suavização.

Na Equação 5 é apresentada o cálculo da média ponderada entre a demanda real e a nova base, a qual começa a incorporar uma quantia de aumento ou redução da demanda expressa na variável tendência. A Equação 6 realiza a suavização da tendência, calculada a partir da variação da base nos dois últimos períodos e a estimativa anterior. A Equação 7 traz a previsão de demanda para  $m$  períodos a frente, utilizando progressão linear (MORETTIN; TOLOI, 2006; LUSTOSA et al., 2008).

Morettin e Tolo (2006) discorrem as vantagens e desvantagens da utilização, como sendo semelhantes aos dos métodos anteriormente estudados, acrescentando desvantagem da complexidade de determinação da constante  $\beta$ . O método de Holt é indicado para os casos onde a série não apresenta sazonalidade, apenas indica a existência de tendência (SAMOHYL; SOUZA; MIRANDA, 2008; KLINDOKMAI et al., 2014).

#### 2.2.1.1.4 Modelo de Suavização Exponencial Sazonal de Holt-Winters

Morettin e Tolo (2006) citam que para séries temporais que apresentam um padrão de comportamento mais complexo, existem outras formas de suavização, tais como os modelos de Suavização Exponencial de Holt-Winters (SEHW) e o método de suavização exponencial geral (ou suavização direta).

O método de SEHW é utilizado em situações em que as séries temporais apresentam padrão de demanda com tendência linear e sazonalidade. O método aplica equações de suavização para estimar o nível, a tendência e a sazonalidade da série temporal analisada no processo de previsão (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998; ELLERO; PELLEGRINI, 2014; KAYA; DEMIREL, 2015).

O método é dividido em forma multiplicativa e aditiva, essas formas segundo Winters (1960) dependem da forma como é modelada a sazonalidade. A forma aditiva é adequada para séries temporais cuja amplitude da sazonalidade é independente do nível de demanda, já a forma multiplicativa é sugerida para séries temporais em que a amplitude da sazonalidade varia com o nível da demanda (WINTERS, 1960; HOLT, 2004; RASMUSSEN, 2004).

As equações básicas do método multiplicativo de SEHW podem ser verificadas nas Equações 8, 9, 10 e 11 (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998; HOLT, 2004; RASMUSSEN, 2004).

$$\text{Nível:} \quad L_t = \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + b_{t-1}) \quad (8)$$

$$\text{Tendência:} \quad b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1} \quad (9)$$

$$\text{Sazonalidade:} \quad S_t = \gamma \left( \frac{Y_t}{L_t} \right) + (1 - \gamma)S_{t-s} \quad (10)$$

$$\text{Previsão:} \quad F_{t+m} = (L_t + b_{t-m})S_{t-s+m} \quad (11)$$

onde  $s$  é o número de períodos por ciclo sazonal,  $S_t$  é a estimativa do componente sazonal da série temporal no período  $t$  e  $\alpha$ ,  $\beta$  e  $\gamma$  são as constantes de suavização (com valores entre 0 e 1, e não relacionados).

A Equação 8 ajusta  $L_t$  para a tendência do período anterior, pela adição de  $b_{t-1}$  ao último valor suavizado de nível  $L_{t-1}$ . O primeiro termo da Equação 8 é dividido por um termo sazonal ( $S_{t-s}$ ) para eliminar as flutuações sazonais no cálculo do nível  $L_t$ . A Equação 9 atualiza a tendência realizando a diferença entre os últimos dois valores suavizados de nível ( $L_t$  e  $L_{t-1}$ ).

A Equação 10 estima o componente sazonal pela ponderação, por meio de uma constante de suavização  $\gamma$ , da razão de  $Y_t$  e  $L_t$ , (que corresponde a sazonalidade do período  $t$ ) com a sazonalidade  $S_{t-s}$  (correspondente a sazonalidade do período analisado do ciclo sazonal anterior). A Equação 11 é utilizada para obter a previsão de demanda, sendo obtida pela multiplicação da estimativa de um componente sazonal ( $S_{t-s+m}$ ) à previsão do método linear de Holt (Equação 7) (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998; MORETTIN; TOLOI, 2006).

O método aditivo de SEHW trata a componente sazonal de forma aditiva, este ajusta-se mais a séries com tendência e sazonalidade aditiva, isto é, aquelas em que o efeito sazonal não é função do nível médio corrente da série temporal e pode ser simplesmente adicionado ou subtraído de uma previsão que dependa apenas de nível e tendência (KOEHLER et al., 2001; RASMUSSEN, 2004; MORETTIN; TOLOI, 2006).

Corroborando Archibald e Koehler (2003) explicam que o modelo SEHW aditivo é utilizado quando a amplitude da variação sazonal permanece constante e o multiplicativo quando a amplitude da sazonalidade aumenta com o tempo. As Equações 12, 13, 14 e 15 apresentadas, são utilizadas para o processo da previsão do método de SEHW aditivo.

$$\text{Nível:} \quad L_t = \alpha(Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + b_{t-1}) \quad (12)$$

$$\text{Tendência:} \quad b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1} \quad (13)$$

$$\text{Sazonalidade:} \quad S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-s} \quad (14)$$

$$\text{Previsão:} \quad F_{t+m} = (L_t + b_{t-m})S_{t-s+m} \quad (15)$$

É possível observar que a Equação 13 é idêntica à Equação 9, a qual se refere ao modelo multiplicativo de Winters. A única diferença, entre as demais expressões e as

correspondentes ao modelo multiplicativo de Winters, é que os índices sazonais agora são somados e subtraídos ao invés de serem tomados em produtos e divisões (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998; MORETTIN; TOLOI, 2006; RAGSDALE, 2009).

Morettin e Toloi (2006) ressaltam que o método de previsão SEHW apresentam várias vantagens como: fácil entendimento, aplicação não dispendiosa, adequada para série com padrão de comportamento mais geral. Os mesmos autores acrescentam que as desvantagens que são: dificuldades de determinar os valores mais apropriados das constantes de suavização e/ou impossibilidade de estudar as propriedades estatísticas, tais como média e variância da previsão e, conseqüentemente, a construção de um intervalo de confiança.

#### 2.2.1.1.5 Métodos de Suavização Exponencial com correção de espaço de tempo

Os quatro métodos citados anteriormente (MMS, SES, SEH e SEHW) são considerados a base da SE, a qual teve e têm avanços até os dias atuais. Existem outros métodos que derivam destes, proporcionando novas classificações para os métodos de SE. Pegels (1969) classificou em nove os modelos de SE, segundo a tendência e sazonalidade (aditiva ou multiplicativa).

A principal modificação de Pegels foi a introdução dos modelos com tendência multiplicativa. Gardner e McKenzie (1985) desenvolveram um modelo com tendência aditiva amortecida, com a principal função de corrigir a previsão, que em muitas situações não é linear. O último avanço significativo em termos de modelagem foi realizado em Taylor (2003), quando ele propôs modelos com tendência multiplicativa amortecida. Todos estes modelos foram classificados em Hyndman et al. (2002) quanto à tendência e sazonalidade e depois adaptado em Taylor (2003) para incluir tendência multiplicativa amortecida. O Quadro 1 apresenta esta classificação.

Quadro 1 - Resumo dos métodos de suavização exponencial proposta por Taylor

Tendência	Sazonalidade		
	Nenhuma (N)	Aditiva (A)	Multiplicativa (M)
Nenhuma (N)	NN	NA	NM
Aditiva (A)	AN	AA	AM
Aditiva Amortecida (Ad)	AdN	AdA	AdM
Multiplicativa (M)	MN	MA	MM
Multiplicativa Amortecida (Md)	MdN	MdA	MdM

Fonte: Taylor (2003).



O modelo NN representa uma série sem tendência e sem sazonalidade e é a notação da Suavização Exponencial Simples (SES). O modelo a AN representa uma série com tendência aditiva e sem sazonalidade, sendo este o modelo de Holt. E os modelos AA e AM representam os modelos de Holt-Winters na forma da sazonalidade aditiva e multiplicativa respectivamente. A formulação padrão dos modelos de suavização exponencial apresentados no Quadro 1 pode ser observada em Gardner (2006).

Porém há outra representação dos modelos de SE, Ord et al. (1997) propuseram a representação dos modelos de SE por meio de modelos de espaço de estado com uma única fonte de erro. A vantagem desta representação é que várias propriedades estatísticas podem ser estudadas, como a estimativa da média e da variância das previsões e a estabilidade, sendo assim gerando melhores previsões. A escolha de uma única fonte de erro permitiu formular com simplicidade os modelos de suavização exponencial, lineares e não lineares. Os modelos não lineares são todos que possuem a tendência ou a sazonalidade multiplicativa.

Em Hyndman et al. (2002) é proposto a representação dos modelos de suavização sob duas condições de erro, aditivo e multiplicativo. O Quadro 2 apresenta a notação desses métodos e no Anexo A apresenta-se a formulação dos modelos de SE com condições de erro.

Quadro 2 - Resumo dos métodos de suavização exponencial com correção de espaço de tempo proposta por Ord et al. (1997)

Tendência	Sazonalidade					
	Nenhuma (N)		Aditiva (A)		Multiplicativa (M)	
	Erro					
	A	M	A	M	A	M
Nenhuma (N)	ANN	MNN	ANA	MNA	ANM	MNM
Aditiva (A)	AAN	MAN	AAA	MAA	AAM	MAM
Aditiva Amortecida (Ad)	AAdN	MAdN	AAdA	MAdA	AAdM	MAdM
Multiplicativa (M)	AMN	MMN	AMA	MMA	AMM	MMM
Multiplicativa Amortecida (Md)	AMdN	MMdN	AMdA	MMdA	AMdM	MMdM

Fonte: Elaborado pelo autor adaptado de Hyndman et al. (2002).

São casos já vistos em Ord et al. (1997) onde se considera as situações com erro homocedástico (variância constante) e erro heterocedástico (variância não constante). Quando o tipo de erro é selecionado a classificação do modelo é realizada com três letras, sendo a primeira o erro, aditivo (A) ou multiplicativo (M), a segunda letra a tendência, aditiva (A), multiplicativa (M), aditiva amortecida (Ad), multiplicativa (Md) e a terceira letra a sazonalidade, aditiva (A) ou multiplicativa (M).

### 2.2.1.2 Modelos de Box-Jenkins

Os modelos de Box-Jenkins, também conhecidos como Modelos Autorregressivos Integrados a Média Móvel, ou simplesmente ARIMA (*Autoregressive Integrated Moving Average*), foram propostos por George Box e Gwilym Jenkins no início dos anos 70 (MADDALA, 2003; MORETTIN; TOLOI, 2006; BOX; JENKINS; REINSEL, 2008). O método de Box-Jenkins é um método de previsão de demanda que utiliza um algoritmo matemático complexo, com termos autorregressivos e de média móvel, para identificar a forma do modelo matemático mais adequado para a série temporal analisada com  $n$  observações (ARCHER, 1980; MORETTIN; TOLOI, 2006; ASLAM, 2016).

O método é apropriado para séries de comprimento médio a longo, com no mínimo 50 observações. Um dos processos fundamentais na metodologia de Box-Jenkins é transformar uma série não estacionária em uma estacionária, cuja análise é mais simples, pois se obtém a estabilidade dos parâmetros (MENTZER; GOMES, 1989; SOUZA, 2006).

Os modelos de Box-Jenkins partem da ideia de que os valores de uma série temporal são altamente dependentes, ou seja, cada valor pode ser explicado por valores prévios da série. Os modelos ARIMA representam a classe mais geral de modelos para a análise de séries temporais. Alguns conceitos devem ser considerados para o entendimento dos modelos Box-Jenkins, tais conceitos são apresentados na sequência (MORETTIN; TOLOI, 2006).

#### 2.2.1.2.1 Conceitos básicos para percepção dos modelos Box-Jenkins

Preliminarmente, alguns conceitos devem ser apresentados para familiaridade e facilitação do entendimento dos modelos de Box-Jenkins.

- Modelos Estocásticos e Determinísticos

Para Pellegrini (2000), a representação de fenômenos físicos mostrada numa série temporal pode ser feita utilizando uma modelagem matemática, podendo ser agrupados e descritos por meio de equações matemáticas. Nas palavras de Giovannelli e Proietti (2016) pode-se utilizar modelagem matemática, para prever o valor de variáveis de interesse em qualquer momento no tempo, caso as variáveis sejam dependentes do tempo. Sempre que uma previsão exata for possível, os modelos são ditos determinísticos. No entanto, Morettin e Tolo (2006) citam que muitos fenômenos não são de natureza determinística, devido à

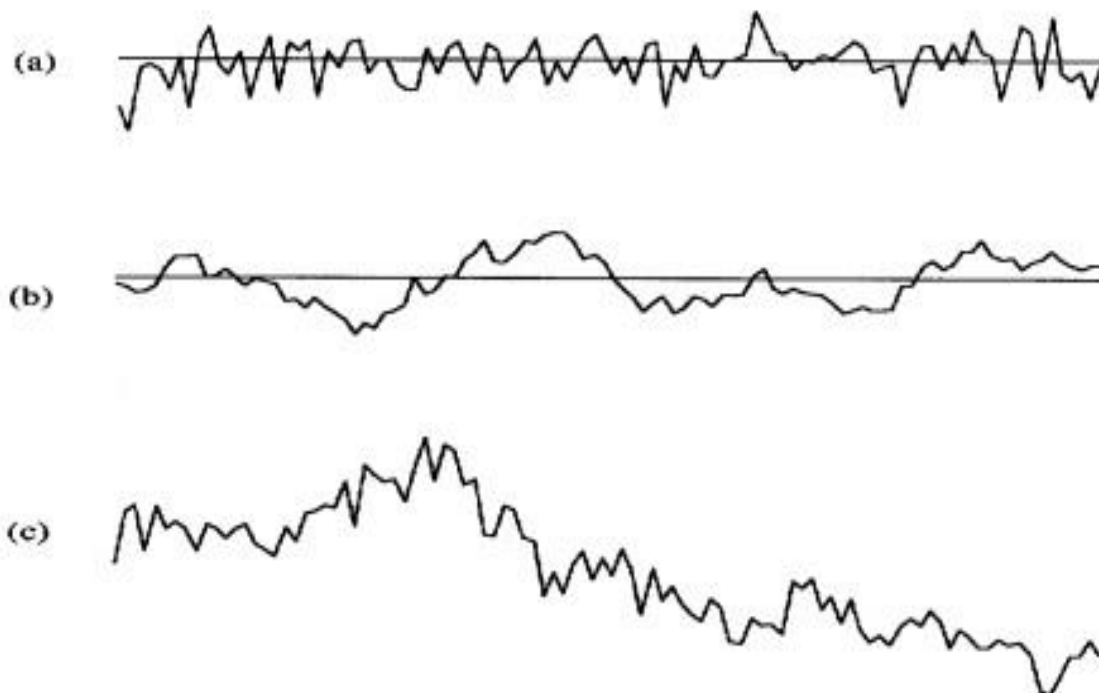
presença aleatória de fatores desconhecidos, nestes casos, a previsão do valor futuro está sujeita a um cálculo de probabilidade. Modelos matemáticos desenvolvidos para analisar tais sistemas são ditos estocásticos.

Um processo estocástico é caracterizado por uma família de variáveis aleatórias que descrevem a evolução de algum fenômeno de interesse. Processos estocásticos que caracterizam os estudos de séries temporais descrevem a evolução temporal de um fenômeno de interesse (PELLEGRINI; FOGLIATTO, 2001).

- Modelos Estocásticos Estacionários e Não-Estacionários

Uma importante classe de modelos estocásticos utilizados na representação de séries temporais são os modelos estacionários. Tais modelos pressupõem um processo sob equilíbrio, onde as variáveis se mantêm a um nível constante médio (BOX; JENKINS; REINSEL, 2008). Porém, muitas séries temporais são melhores representadas por modelos não estacionários. Séries estacionárias e não estacionárias podem ser visualizadas graficamente na Figura 7.

Figura 7 - Série estacionária com ruído aleatório (a); Série estacionária com ruídos autocorrelacionados; Série não estacionária (c)



Os gráficos (a) e (b) na Figura 7 mostram séries temporais exibindo variação estacionária. Tais séries variam de maneira estável no tempo, sobre um valor de média fixo. O gráfico (c) mostra uma série temporal não estacionária, a qual não se desloca no tempo sobre uma média fixa.

A série da Figura 7(a) é uma série de ruído aleatório. Em tais séries, as diferenças entre as observações e a média são estatisticamente independentes, seguindo alguma distribuição de probabilidade. A série da Figura 7(b) também é estacionária, mas apresenta ruídos autocorrelacionados. Nesse caso, diferenças entre observações e a média não são estatisticamente independentes entre si. A série da Figura 7(c) demonstra uma variação não estacionária, essas séries são encontradas com frequência em aplicações na indústria, em estudos de negócios e economia, bem como em estudos envolvendo séries temporais no mercado imobiliário.

Além da visualização gráfica da série, para ter certeza da estacionariedade da série, é interessante realizar testes de presenças de raízes unitárias. Um destes testes é o *Augmented Dickey-Fuller* (ADF). Este teste possui as seguintes hipóteses H0: tem raiz unitária (não é estacionária); e H1: não tem raiz unitária (é estacionária). Mais informações sobre este teste, verificar em Brooks e Tsolacos (2014).

- Autocorrelação

A autocorrelação é utilizada para delinear a correlação entre dois valores da mesma série temporal, em diferentes períodos de tempo, sendo fundamental para a análise de séries temporais. Assim, um coeficiente de autocorrelação  $\rho_1$  mede a correlação entre dois valores adjacentes na série, e a autocorrelação, neste caso, é dita autocorrelação de *lag* (ou defasagem). De maneira geral, o coeficiente de autocorrelação  $\rho_k$  mede a correlação entre observações distantes  $k$  períodos de tempo (MORETTIN; TOLOI, 2006). A autocorrelação de *lag*  $k$  é medida pelo coeficiente  $\rho_k$ , definido por (BOX; JENKINS; REINSEL, 2008) na Equação 16.

$$\rho_k = \frac{E[(z_t - \mu)(z_{t-k} - \mu)]}{\sigma_z^2} \quad (16)$$

Uma estimativa do coeficiente de autocorrelação populacional  $\rho_k$  na Equação 16 é dado pelo coeficiente de autocorrelação amostral, que é definido na Equação 17.

$$r_k = \frac{\sum_{t=k+1}^n (z_t - \bar{z})(z_{t-k} - \bar{z})}{\sum_{t=1}^n (z_t - \bar{z})^2}, \text{ com } k = 0, 1, 2, \dots, n, \quad (17)$$

$$\text{Onde } \bar{z} = \frac{1}{n} \sum_{t=1}^n z_t$$

Na prática para se alcançar boa estimativa do coeficiente de autocorrelação, deve-se dispor de pelo menos 50 observações da variável  $z$ . O número de autocorrelações de *lags* diferentes que se calcula para a análise da série temporal deve ser de  $n/4$ , onde  $n$  é o número total de observações na série.

- Autocorrelação parcial

O coeficiente de autocorrelação parcial de ordem  $k$ , designado por  $\alpha_k$ , pode ser calculado estimando-se os coeficientes da regressão múltipla de  $x_t$  em função de  $x_{t-1}, x_{t-2}, \dots, x_{t-k}$ , como apresentado na Equação 18.

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_k x_{t-k} \quad (18)$$

Destarte, autocorrelações parciais são utilizadas para mensurar o grau de associação entre as observações  $x_t$  e  $x_{t-k}$ , quando os efeitos das outras observações, defasadas no tempo em 1, 2, ...,  $k-1$  períodos, são removidos. As autocorrelações parciais  $\alpha_1, \alpha_2, \dots, \alpha_k$ , tomadas em conjunto, constituem a função de autocorrelação parcial ou PACF (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998).

Expostos os conceitos requisitados para compreender os modelos de Box-Jenkins é importante citar que o método possibilita combinações de modelos, e com a ajuda de análises estatísticas pode determinar o modelo combinado mais apropriado para uma dada situação. O método de Box-Jenkins modela a função de autocorrelação de uma série temporal estacionária com o mínimo de parâmetros possíveis, utilizando uma combinação de termos de autorregressão (AR), integração (I) e média móvel (MA). A notação do modelo geral é ARIMA  $(p, d, q)(P, D, Q)S$  onde (WALKER; MCCLELLAND, 1991; BUSINGER; READ, 1999; BOX; JENKINS; REINSEL, 2008):

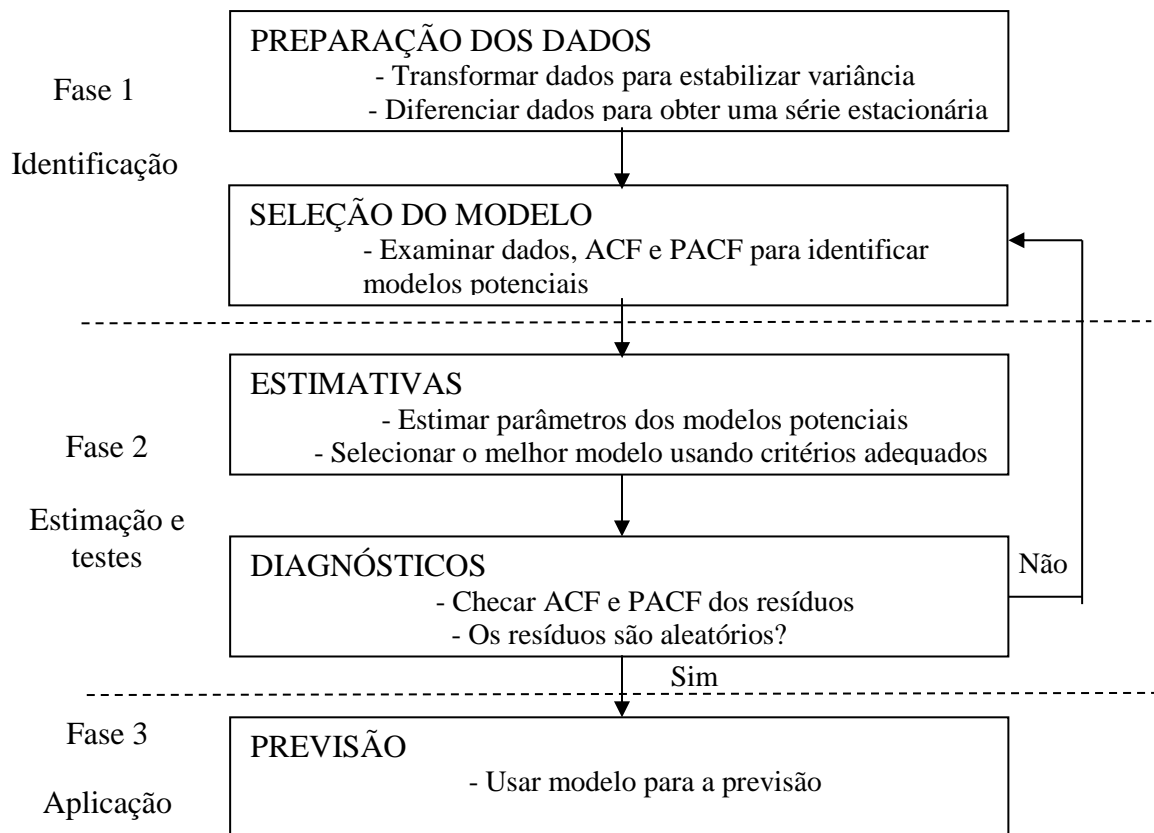
- (AR) – parcela autorregressiva que modela a dependência de um valor atual sobre valores passados;
- (I) – parcela de integração que remove a tendência da série temporal, tornando a série estacionária;
- (MA) – parcela de média móvel, a qual assume que valores atuais são dependentes de erros de previsão de períodos passados;
- $p$  – ordem máxima dos parâmetros de autorregressão simples;
- $d$  – número de diferenciações não-sazonais aplicadas para tornar a série temporal estacionária;
- $q$  – ordem máxima dos parâmetros de média móvel simples;
- $P$  – ordem máxima dos parâmetros de auto-regressão sazonal;
- $D$  – número de diferenciações sazonais aplicadas para tornar a série temporal estacionária;
- $Q$  – ordem máxima dos parâmetros de média móvel sazonal; e
- $s$  – número de períodos por ciclo sazonal.

A operacionalização do método Box-Jenkins segue três passos para a modelagem de séries temporais (Identificação, Estimação e Teste, e Aplicação), passos estes apresentados na Figura 8.

Para Morettin e Toloí (2006), esse ciclo é utilizado para identificar, não somente um método, mas alguns modelos que serão estimados e verificados. Como o propósito, para essa pesquisa, é a previsão, escolher-se-á entre os modelos ajustados o melhor, por exemplo, no sentido de fornecer o menor erro médio de previsão, em geral são várias as medidas de acurácia, que são apresentadas na seção 2.2.2.

A fase crítica do procedimento da Figura 8 é a identificação. É possível que vários pesquisadores identifiquem modelos diferentes para a mesma série temporal (MORETTIN; TOLOI, 2006; LEMOS, 2006; HENKE; WICHERN, 2009). Em geral, os modelos postulados são parcimoniosos, pois contém um número pequeno de parâmetros e as previsões obtidas são bastante precisas, comparando-se favoravelmente com os demais métodos de previsão (MORETTIN; TOLOI, 2006).

Figura 8 - Sequência de execução do método de Box-Jenkins



Fonte: Adaptado de Makridakis, Wheelwright e Hyndman (1998).

A denominação dos modelos, mais especificamente são  $AR(p)$  (Autorregressivo de ordem  $p$ ) e  $MA(q)$  (média-móvel de ordem  $q$ ), que podem ser integrados, gerando os modelos  $ARIMA(p,d,q)$ , ou seja, composto de  $AR(p)$  e  $MA(q)$  com  $d$  diferenciações. Existem ainda modelos com sazonalidade, esses são utilizados para decompor séries temporais que apresentam comportamento recorrente no tempo, por exemplo, a cada  $s$  períodos de tempo. Esses modelos são denominados por  $SARIMA(p,d,q)(P,D,Q)_s$ , onde as letras maiúsculas denotam a parte sazonal do modelo e são análogas a  $p,d,q$  sendo que  $s$  funciona como descrito acima. Os modelos expostos serão apresentados na sequência (MORETTIN; TOLOI, 2006; BOX; JENKINS; REINSEL, 2008).

#### 2.2.1.2.2 Modelos Autorregressivos

Um modelo estocástico útil na representação de um grande número de séries temporais é o modelo autorregressivo. Neste modelo, o valor corrente do processo é expresso como uma

combinação linear finita de valores prévios do processo e um ruído aleatório  $a_t$  (MORETTIN; TOLOI, 2006; PELLEGRINI, 2010).

Definem-se os valores observados de um processo em espaços de tempo igualmente divididos  $t, t-1, t-2, \dots$  por  $z_t, z_{t-1}, z_{t-2}, \dots$

Definem-se também  $\tilde{z}_t, \tilde{z}_{t-1}, \tilde{z}_{t-2}, \dots$  como sendo desvios da média  $\mu$ , ou seja,

$$\tilde{z}_t = z_t - \mu, \tilde{z}_{t-1} = z_{t-1} - \mu, \tilde{z}_{t-2} = z_{t-2} - \mu, \dots$$

Então, tem-se a Equação 19 representa um processo autorregressivo de ordem  $p$ , ou simplesmente AR( $p$ ).

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \dots + \phi_p \tilde{z}_{t-p} + a_t \quad (19)$$

A razão para o nome autorregressivo é pelo fato de um modelo linear relacionando uma variável dependente  $z$  a um grupo de variáveis independentes quaisquer e a um termo de erro  $a$ , ser geralmente referido como um modelo de regressão. Na Equação 19, a variável  $z$  é regredida em valores prévios da própria variável; por essa razão, o modelo é denominado autorregressivo (BOX; JENKINS; REINSEL, 2008).

Os coeficientes autorregressivos  $\phi_1, \phi_2, \dots, \phi_p$ , são parâmetros que descrevem como um valor corrente  $z_t$  relaciona-se com valores passados  $z_{t-1}, z_{t-2}, \dots, z_{t-p}$ . O modelo AR( $p$ ) contém  $p+2$  parâmetros desconhecidos ( $\mu, \phi_1, \phi_2, \dots, \phi_p, \sigma_a^2$ ), os quais podem ser estimados a partir dos valores observados na série temporal. Realça-se que  $\sigma_a^2$  é a variância do processo de ruído aleatório  $a_t$ .

Processos autorregressivos podem ser estacionários ou não estacionários (BOX; JENKINS; REINSEL, 2008). A premissa necessária para a estacionariedade é que o operador autorregressivo  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ , considerado como sendo um polinômio em  $B$  de grau  $p$ , tenha todas as suas raízes  $\phi(B) = 0$  maiores que 1, em valores absolutos, ou seja, todas as raízes devem estar fora do círculo unitário.

O processo autorregressivo possui dois importantes casos especiais: os processos de primeira e segunda ordem. Se  $p = 1$ , tem-se um processo autorregressivo de primeira ordem, designado por AR(1) é descrito na Equação 20.

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + a_t \quad (20)$$



Este processo também é conhecido como *processo de Markov*, em um processo de Markov, para saber-se o valor assumido pela variável de interesse num instante  $t$  qualquer, necessita-se somente a informação sobre o valor assumido pela mesma em  $t-1$  (ROSS, 1993). Para o processo AR(1) ser estacionário, a raiz de  $\phi B = 1 - \phi_1 B = 0$  deve estar fora do círculo unitário. Isto equivale a dizer que  $|\phi_1| < 1$ , para que a estacionariedade se verifique.

A função de autocorrelação do processo é dada por  $p_k = \phi_1 p_{k-1}$ , com  $k > 0$ , ou  $p_k = \phi_1^k$ , com  $k \geq 0$ , já que  $p_0 = 1$ . Assim, a função de autocorrelação extingue-se exponencialmente quando  $\phi_1$  é positivo. Analogamente, quando  $\phi_1$  é negativo, a ACF extingue-se exponencialmente com alternância de sinal. Quando  $p = 2$ , tem-se um processo autorregressivo de segunda ordem, designado por AR(2) e descrito na Equação 21.

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + a_t \quad (21)$$

Novamente, para o processo ser estacionário, as raízes da equação  $\phi B = 1 - \phi_1 B - \phi_2 B^2 = 0$  devem estar fora do círculo unitário. Isto implica em parâmetros  $\phi_1$  e  $\phi_2$  que satisfaçam as seguintes condições:

- a)  $\phi_1 - \phi_2 < 1$
- b)  $\phi_2 - \phi_1 < 1$  e
- c)  $-1 < \phi_2 < 1$

A função de autocorrelação do processo AR(2) é dada por  $p_k = \phi_1 p_{k-1} + \phi_2 p_{k-2}$ , com  $k > 0$ . Para  $k = 1$  e  $2$ , têm-se respectivamente as funções de autocorrelações apresentadas nas Equações 22 e 23.

$$p_1 = \frac{\phi_1}{1 - \phi_2} \quad (22)$$

$$p_2 = \phi_2 + \frac{\phi_1^2}{1 - \phi_2} \quad (23)$$

As Equações 22 e 23 são denominadas *equações de Yule – Walker* (BOX; JENKINS; REINSEL, 2008). A ACF de ordem 2 é complexa. Se  $\phi_1^2 + 4\phi_2 \geq 0$ , a ACF é uma mistura de distribuições exponenciais decrescentes. Quando  $\phi_1^2 + 4\phi_2 < 0$ , a função de autocorrelação

extingue-se de maneira senoidal. De uma maneira geral, a função de autocorrelação para um processo estacionário autorregressivo consiste de uma mistura de distribuições exponenciais com ondas senoidais decrescentes (BOX; JENKINS; REINSEL, 2008).

### 2.2.1.2.3 Modelos de Médias Móveis

Nos modelos de média móvel,  $\tilde{z}_t$ , que representa a observação  $z_t$  subtraída da média  $\mu$ , depende linearmente de um número finito  $q$  de valores prévios do ruído aleatório  $a_t$ . Assim, na Equação 24 é possível visualizar a equação de médias móveis (MA) de ordem  $q$ .

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (24)$$

O nome, média móvel pode levar à equívocos de interpretação, já que os pesos  $1, -\theta_1, -\theta_2, \dots, -\theta_q$ , não somam, necessariamente, a unidade nem precisam ser, necessariamente, positivos (MONTGOMERY; JOHNSON; GARDINER, 1990).

O coeficiente de média móvel  $\theta$  de ordem  $q$  pode ser expresso usando a definição do operador  $B$ ,  $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ , simplificando matematicamente para  $\tilde{z}_t = \theta(B)a_t$ , expressão a qual contém  $q+2$  parâmetros desconhecidos ( $\mu, \theta_1, \theta_2, \dots, \theta_q, \sigma_a^2$ ) a partir dos valores observados na série temporal.

Uma vez que a série  $\Psi(B) = \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$  é finita, nenhuma restrição é necessária sobre os parâmetros do processo de média móvel para assegurar estacionariedade. A função de autocorrelação de um processo MA( $q$ ) pode ser visualizada na Equação 25.

$$p_k = \frac{-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q}{1 + \theta^2 + \theta_q^2} \text{ com } k = 1, 2, \dots, q \text{ e } p_k = 0 \text{ quando } k > q \quad (25)$$

Para um caso particular de um processo MA(1), é apresentada a Equação 26, com o processo sendo estacionário para qualquer valor de  $\theta_1$ .

$$\tilde{z}_t = (1 - \theta_1 B)a_t \quad (26)$$

A função de autocorrelação do processo MA (1) é apresentada na Equação 27.

$$p_k = \frac{-\theta_1}{1 - \theta_1^2} \text{ quando } k = 1 \text{ e } p_k = 0 \text{ quando } k \geq 2 \quad (27)$$

Outro caso particular de interesse é o processo de média móvel MA(2) o qual é estacionário para qualquer valor de  $\theta_1$  e  $\theta_2$ , representado na Equação 28.

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \quad (28)$$

A ACF do processo MA(2) Para  $k = 1$  e  $2$ , tem-se respectivamente as funções de autocorrelações apresentadas nas Equações 29 e 30. Ressalta-se que  $p_k = 0$ , para  $k \geq 3$ .

$$p_1 = \frac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2} \quad (29)$$

$$p_2 = \frac{\theta}{1 + \theta_1^2 + \theta_2^2} \quad (30)$$

Processos autorregressivos e de média móvel apresentam uma propriedade de interesse, denominada invertibilidade (PELLEGRINI, 2000; MORETTIN; TOLOI, 2006). Considere, por exemplo, um processo MA(1), representado pela Equação 26, reescrito isolando  $a_t$ , podendo ser visualizado na Equação 31.

$$a_t = (1 - \phi_1 B)^{-1} \tilde{z}_t \quad (31)$$

Se  $|\theta_1| < 1$ , a Equação 31 pode ser escrita na forma apresentada na Equação 32.

$$a_t = (1 + \theta_1 B^1 + \theta_1 B^2 + \dots) \tilde{z}_t \quad (32)$$

Thomas Jr. e Finney (1992) citam que isto se deve ao fato de que a expressão  $(1 - x)^{-1}$ , para casos onde  $|x| < 1$ , pode ser representada por uma série infinita de forma:  $1 + x + x^2 + x^3 + \dots + x^n + \dots$ . Assim, a Equação 32 representa um processo autorregressivo de ordem infinita. Então, o processo MA(1) é *invertido* para se obter um processo AR( $\infty$ ). A premissa  $|\theta_1| < 1$  é chamada de condição de invertibilidade para um processo MA(1) (MONTGOMERY; JOHNSON; GARDINER, 1990).

Em geral, para um processo MA( $q$ ) ser invertido em um AR( $\infty$ ), as raízes do polinômio  $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q = 0$  devem estar fora do círculo unitário. No processo MA(2), em particular, essa condição implica em:

- a)  $\phi_1 - \phi_2 < 1$
- b)  $\phi_2 - \phi_1 < 1$  e
- c)  $-1 < \phi_2 < 1$

As condições de invertibilidade para os coeficientes em um processo MA( $q$ ) são idênticas às condições de estacionariedade para um processo AR( $p$ ). Assim, (i) o processo MA( $q$ ) é estacionário para todos os valores de  $\Theta$ , mas é inversível apenas se as raízes da equação  $\theta(B) = 0$  resultarem fora do círculo unitário; e, (ii) o processo AR( $p$ ) é estacionário apenas se as raízes de  $\phi(B) = 0$  resultarem fora do círculo unitário, mas é inversível para todos os valores de  $\phi$  (MONTGOMERY; JOHNSON; GARDINER, 1990; MORETTIN; TOLOI, 2006).

#### 2.2.1.2.4 Modelos mistos Autorregressivos e de Médias móveis

Algumas vezes, séries temporais são melhores modeladas com a inclusão de termos autorregressivos e de média móvel. O resultado é um modelo misto autorregressivo - média móvel de ordem ( $p, q$ ), que pode ser observado na Equação 33.

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \dots + \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (33)$$

Rearranjando os termos da Equação 33, utilizando a notação do operador de defasagem B, tem-se na Equação 34, o modelo que pode ser abreviado para ARMA ( $p, q$ ).

$$\phi(B)\tilde{z}_t = \theta(B) a_t \quad (34)$$

O modelo possui  $p + q + 2$  parâmetros desconhecidos ( $\mu, \phi_1, \dots, \phi_p; \theta_1, \dots, \theta_q, \sigma_a^2$ ), que podem ser estimados a partir dos valores observados na série temporal. Na prática, os valores de  $p$  e  $q$  são geralmente menores que 2 para séries temporais estacionárias (BOX; JENKINS; REINSEL, 2008).

As condições de estacionariedade e invertibilidade estabelecidas para os processos  $AR(p)$  e  $MA(q)$  se mantêm nos modelos  $ARMA(p, q)$ . Ou seja, um modelo  $ARMA(p, q)$  é estacionário se as raízes do polinômio  $\Phi(B) = 0$  estiverem fora do círculo unitário, e inversível se as raízes de  $\Theta(B) = 0$  resultarem fora do círculo unitário. Um caso especial de interesse prático dos modelos mistos autorregressivo – média móvel é o processo  $ARMA(1, 1)$ , dado pela Equação 35, ressalta-se que esse processo é estacionário se  $|\Phi_1| < 1$  e inversível se  $|\Theta_1| < 1$ .

$$\tilde{z}_t - \Phi_1 \tilde{z}_{t-1} = a_t - \Theta_1 a_{t-1} \quad (35)$$

A ACF de um processo  $ARMA(1,1)$  é apresentada na Equação 36.

$$p_1 = \frac{(1 - \Phi_1 \Theta_1)(\Phi_1 - \Theta_1)}{1 - \Theta_1^2 - 2\Phi_1 \Theta_1}, \quad e \quad p_k = \Phi_1 p_{k-1} \quad \text{quando } k \geq 2 \quad (36)$$

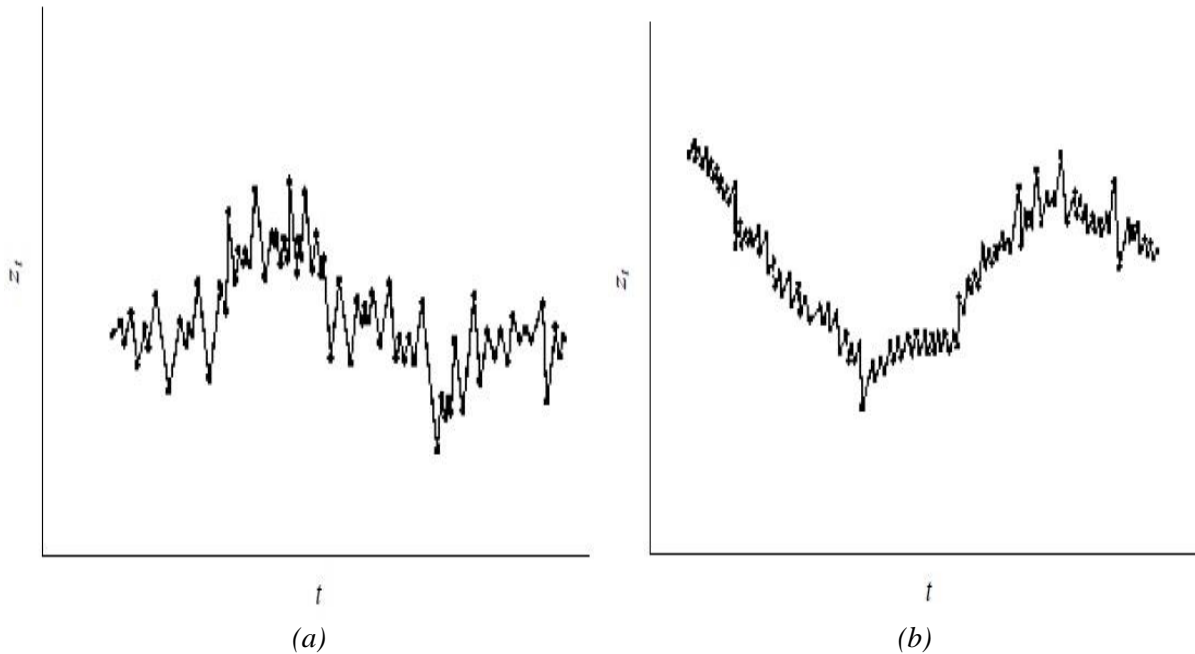
A partir da Equação 36 compreende-se que a componente de média móvel faz parte apenas da determinação de  $p_1$ . Por conseguinte, a ACF apresenta um pequeno decréscimo entre  $p_0$  e  $p_1$ , decrescendo exponencialmente a partir de  $p_1$ , em contraste com o modelo  $AR(1)$ , que decresce exponencialmente a partir de  $p_0$ ,  $p_1$  é positivo sempre que  $\Phi_1$  for maior que  $\Theta_1$ , e negativo em caso contrário.

#### 2.2.1.2.5 Modelos não estacionários

Muitas séries temporais não possuem uma média constante. Isto significa que, em nenhum dado intervalo de tempo, as observações da série se comportam como as observações de um intervalo de tempo distinto. Tais séries são chamadas de não estacionárias na média. Da mesma forma, é possível uma série temporal exibir não estacionariedade na média e na declividade (PELLEGRINI, 2000, MORETTIN; TOLOI, 2006, TAYLOR; SNYDER, 2012). Um exemplo de série não estacionária na média vem apresentado na Figura 9(a); um exemplo de série não estacionária na média e na declividade vem apresentado na Figura 9(b).

Séries temporais não estacionárias podem exibir independente da média local (ou média e declividade locais), um comportamento geral homogêneo, com a ocorrência de tendências que se repetem (MORETTIN; TOLOI, 2006).

Figura 9 - Série temporal não estacionária na média (a); Série temporal não estacionária na média e na declividade (b)



Fonte: Montgomery, Johnson e Gardiner (1990, p. 88).

Séries não estacionárias podem geralmente ser representadas por um operador autorregressivo generalizado  $\phi(B)$ , no qual uma ou mais raízes do polinômio  $\phi(B)$  são iguais a 1 em módulo. Em particular, se existirem  $d$  raízes unitárias, o operador  $\phi(B)$  assumirá a forma apresentada na Equação 37 (BOX; JENKINS; REINSEL, 2008; MORETTIN; TOLOI, 2006).

$$\phi(B) = \phi(B)(1 - B)^d \quad (37)$$

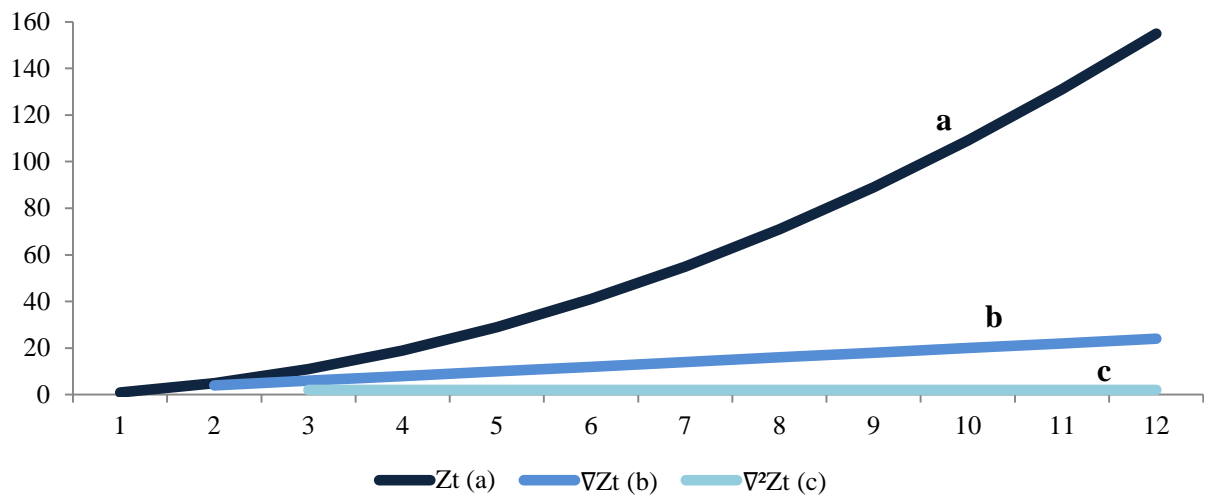
Na Equação 37  $\phi(B)$  é um operador estacionário. Em contrapartida, um modelo que apresenta comportamento homogêneo não estacionário apresenta a seguinte forma, apresentada na Equação 38.

$$\phi(B)w_t = \phi(B)a_t, \quad \text{onde } w_t = \nabla^d z_t = z_t - z_{t-d} \quad (38)$$

Assim, um comportamento homogêneo não estacionário pode ser representado por um processo estacionário, com  $d$  níveis de diferenciação. Na prática,  $d$  pode ser 0, 1, ou, no

máximo, 2. Uma boa representação do efeito da diferenciação sobre uma série temporal não estacionária vem dada pela Figura 10.

Figura 10 - Redução da não estacionariedade de uma série temporal após sucessivas diferenciações



Fonte: Adaptado de Montgomery, Johnson e Gardiner (1990).

A série temporal da Figura 10 (a) exibe comportamento não estacionário na média e na declividade. Após a primeira diferenciação ( $z_t - z_{t-1}$ ), mostrada na Figura 10 (b), a série passa a apresentar não estacionariedade apenas na média. Após a segunda diferenciação [ $(z_t - z_{t-1}) - (z_{t-1} - z_{t-2}) = z_t - 2z_{t-1} + z_{t-2}$ ], com o resultado apresentado na Figura 10 (c), a série torna-se estacionária. A série representada na Figura 10 (a) é um exemplo de série homogeneamente não estacionária: após aplicarem-se diferenciações, esta, torna-se estacionária.

O processo definido pela Equação 38 produz um eficiente modelo para descrever séries temporais estacionárias e não estacionárias. Esse modelo é chamado de processo autorregressivo integrado a média móvel (ARIMA) de ordem  $(p, d, q)$ , onde  $p$  corresponde ao componente autorregressivo,  $d$  ao número de diferenciações e  $q$  ao componente de média móvel. O processo é representado pela Equação 39 (BOX; JENKINS; REINSEL, 2008).

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}, \quad \text{com } w_t = \nabla^d z_t \quad (39)$$

Na Equação 39, quando  $d = 0$ , substituindo-se  $w_t$  por  $z_t - \mu$ , obtém-se o modelo misto estacionário apresentado na Equação 33.

A palavra “Integrado” no modelo ARIMA tem o sentido de “somado”, já que o complemento da Equação 38 pode ser escrito como  $z_t = S^d w_t$ , onde  $S = \nabla^{-1} = (1 - B)^{-1}$  é o operador de soma, definido na Equação 40.

$$S w_t = \sum_{j=0}^{\infty} w_{t-j} = w_t + w_{t-1} + w_{t-2} + \dots \quad (40)$$

Assim, o processo geral autorregressivo integrado a média móvel ARIMA pode ser gerado somando-se ou “integrando-se” o processo estacionário ARMA  $w_t$ ,  $d$  vezes. A seguir, são apresentados alguns casos especiais do modelo ARIMA (BOX; JENKINS; REINSEL, 2008).

- Modelo ARIMA (0, 1, 1)

A formulação desse modelo é apresentada na Equação 41.

$$\nabla z_t = a_t - \theta_1 a_{t-1} = (1 - \theta_1 B) a_t \quad (41)$$

onde  $p = 0$ ,  $d = 1$ ,  $q = 1$ ,  $\phi(B) = 1$ ,  $\theta(B) = 1 - \theta_1 B$ . Este modelo pode ser descrito abreviadamente por IMA (1,1), uma vez que não possui componente autorregressivo (MONTGOMERY; JOHNSON; GARDINER, 1990; MORETTIN; TOLOI, 2006).

- Modelo ARIMA (0, 2, 2)

A formulação desse modelo é apresentada na Equação 42. Onde  $p = 0$ ,  $d = 2$ ,  $q = 2$ ,  $\phi(B) = 1$ ,  $\theta(B) = 1 - \theta_1 B - \theta_2 B^2$ .

$$\nabla^2 z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} = (1 - \theta_1 B - \theta_2 B^2) a_t \quad (42)$$

- Modelo ARIMA (1, 1, 1)

A formulação desse modelo é apresentada na Equação 43.



$$(1 - \phi_1 B) \nabla z_t = (1 - \theta_1 B) a_t \quad (43)$$

onde  $p = 1$ ,  $d = 1$ ,  $q = 1$ ,  $\phi(B) = 1 - \phi_1 B$ ,  $\theta(B) = 1 - \theta_1 B$ .

#### 2.2.1.2.6 Modelos Sazonais

Grande parte das séries temporais, principalmente na indústria, apresentam variações sazonais (MONTGOMERY; JOHNSON; GARDINER, 1990; PINDYCK; RUBINFELD, 2010; GUJARATI; PORTER, 2011). Isto ocorre quando a série exibe uma característica periódica que se repete a cada  $s$  intervalos de tempo. Por exemplo, em séries compostas por observações mensais e sazonalidade anual,  $s$  é igual a 12.

Define-se  $\nabla_s = (1 - B^{12})$  como sendo o operador de diferença sazonal. Assim na Equação 44 é possível observar a formulação da primeira diferenciação sazonal.

$$\nabla_s z_t = (1 - B^{12}) z_t = z_t - z_{t-s} \quad (44)$$

Em geral,  $D$  diferenciações sazonais podem ser requeridas para produzir uma série estacionária. Neste caso, o operador de diferenciação sazonal de ordem  $D$  é  $\nabla_s^D = (1 - B^{12})^D$ . Assim, a forma geral do modelo sazonal autorregressivo integrado a média móvel de ordem  $(P, D, Q)$  é apresentado na Equação 45 (BOX; JENKINS; REINSEL, 2008).

$$\Phi(B^{12}) \nabla_s^D z_t = \theta(B^{12}) \alpha_t \quad (45)$$

onde  $\Phi(B^{12})$  e  $\theta(B^{12})$  são polinômios em  $B^{12}$  de graus  $P$  e  $Q$ , respectivamente, que satisfazem as condições de estacionariedade e invertibilidade. A representação da ordem  $(P, D, Q)$  é feita em letras maiúsculas, para diferenciá-la da representação feita nos modelos não sazonais.

No modelo da Equação 45, os componentes de erro  $\alpha$  estão geralmente correlacionados. Assim,  $\alpha_t$  estaria correlacionado com  $\alpha_{t-1}$  e  $\alpha_{t-2}$ , etc. Para tratar tal relacionamento, introduz-se um segundo modelo, apresentado na Equação 46.

$$\phi(B) \nabla^d \alpha_t = \theta(B) \alpha_t \quad (46)$$

Onde  $\alpha_t$  é um processo de ruído aleatório e  $\Phi(B)$  e  $\Theta(B)$  são polinômios em  $B$  de graus  $p$  e  $q$ , respectivamente, que satisfazem as condições de estacionariedade e invertibilidade. Substituindo-se a Equação 45 na Equação 46, obtém-se um modelo multiplicativo geral, apresentado na Equação 47, chamado de processo multiplicativo de ordem  $(p, d, q) \times (P, D, Q)$ s.

$$\phi_p(B)\Phi_p(B^{12})\nabla^d\nabla_s^D z_t = \theta_q(B)\Theta_q(B^{12}) \alpha_t \quad (47)$$

#### 2.2.1.2.7 Identificação do Modelo

A identificação de um modelo ARIMA é feita por meio da análise dos dados que compõem a série temporal. As principais ferramentas usadas no processo de identificação do modelo são a função de autocorrelação do inglês *Autocorrelation Function* (ACF) e a função de autocorrelação parcial do inglês *Partial Autocorrelation Function* (PACF) (MORETTIN; TOLOI, 2006; ARMSTRONG; FILDES, 2006). A escolha do modelo mais apropriado para descrever uma série temporal não é uma tarefa trivial, dado que existe uma grande variedade de modelos ARIMA a serem considerados como candidatos. Na prática, a identificação do melhor modelo pode ser auxiliada pela sequência de passos abaixo (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998; MORETTIN; TOLOI, 2006):

- a) Considerando o gráfico da série temporal, analisa-se seu comportamento no tempo. Em algumas situações, é necessário fazer transformações nos dados (por exemplo, uma transformação logarítmica ou uma diferença), com a finalidade de estabilizar a variância da série em estudo.
- b) Uma vez estabilizada a variância, caso tal procedimento seja necessário, verificar a estacionariedade da série. Este procedimento é feito em duas etapas: análise da série temporal e dos gráficos das ACF e PACF e testes de raízes unitárias. Quando a série exibe dados em torno de uma média constante e os gráficos das ACF e das PACF apresentam autocorrelações que tendem a zero rapidamente, tem-se a indicação de que a série é estacionária. Se algum destes requisitos não for observado, há evidências que a série é, possivelmente, do tipo não estacionária. Para afirmar, propõe realizar as três abordagens.
- c) Séries não estacionárias devem ser transformadas utilizando diferenciação. Para séries não sazonais, faz-se a diferenciação das observações ( $z'_t = z_t - z_{t-1}$ ); para

séries sazonais, faz-se a diferenciação sazonal das observações ( $z'_t = z_t - z_{t-s}$ ), que leva em consideração o intervalo sazonal  $s$ . Persistindo a não estacionaridade, nova diferenciação é necessária. Geralmente, a série torna-se estacionária após, no máximo, duas diferenciações.

- d) Uma vez a série estando estacionária, os valores de  $D$  e  $d$  são conhecidos. Resta determinar os componentes normais,  $p$  e  $q$ , e os componentes sazonais,  $P$  e  $Q$ . Se a ACF extingue-se rapidamente e a PACF trunca abruptamente após o  $l^{\text{ésimo}}$  lag, então  $p = l$ . Da mesma maneira, para séries sazonais, o valor de  $P$  será igual ao número de lags significativos múltiplos de  $s$  na PACF. Entendem-se como lags significativos aqueles que ultrapassarem os limites de  $\pm 2S(r_k)$  para a ACF, e  $\pm 2S(\hat{\phi}_{kk})$  para a PACF, onde é apresentado na Equação 48.

$$S(r_k) \approx n^{\frac{1}{2}} [1 + 2(r_1^2 + r_2^2 + \dots + r_q^2)]^{\frac{1}{2}}, \quad S(\hat{\phi}_{kk}) \approx n^{-\frac{1}{2}} \quad (48)$$

Se a PACF extingue-se rapidamente, e a ACF trunca abruptamente após o  $l^{\text{ésimo}}$  lag, então  $q = l$ . Para modelos sazonais, o valor de  $Q$  será igual ao número de lags significativos múltiplos de  $s$  na ACF.

Quando ambas as ACF e PACF extinguem-se rapidamente, um modelo misto pode ser necessário. Tais modelos são de difícil identificação, devendo-se usar um processo por tentativas que inicie testando valores baixos de  $P$ ,  $Q$ ,  $p$  e  $q$ .

#### 2.2.1.2.8 Estimativa dos parâmetros e verificação dos modelos

Uma vez identificado o modelo, seus parâmetros devem ser estimados. O método dos mínimos quadrados pode ser usado na identificação dos parâmetros de modelos ARIMA (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998). Todavia, para os componentes MA, não existe uma fórmula simples para determinação das estimativas dos parâmetros.

Para a estimação dos parâmetros, um método comumente utilizado é o da máxima verossimilhança. Desse modo, o estimador será o valor de parâmetro que maximiza a função de verossimilhança, sendo  $Z$  uma variável aleatória com distribuição de probabilidade  $f(z, \theta)$ , onde  $\theta$  é um parâmetro desconhecido a ser estimado. Sejam  $z_1, z_2, \dots, z_n$  valores observados numa amostra aleatória de tamanho  $n$ . Então, a função de verossimilhança da amostra é apresentada na Equação 49 (MONTGOMERY; RUNGER, 1994).

$$L(\theta) = f(z_1, \theta) \times f(z_2, \theta) \times \dots \times f(z_n, \theta) \quad (49)$$

O estimador de máxima verossimilhança de  $\theta$  é o valor de  $\theta$  que maximiza a função de verossimilhança  $L(\theta)$ . Tal estimador é usualmente determinado diferenciando-se  $L(\theta)$ , igualando o resultado a zero e resolvendo a expressão resultante para  $\theta$ . Uma vez obtido um modelo ajustado para a série temporal, deve-se determinar sua adequação e necessidade de melhoria. Um método lógico para verificação do modelo utiliza o cálculo dos resíduos ( $e_t = z_t - \hat{z}_t$ ). Deve-se estimar e examinar a ACF dos resíduos (MONTGOMERY; JOHNSON; GARDINER, 1990). Se o modelo obtido for apropriado, a ACF da amostra dos resíduos  $r_e(k)$  não deve apresentar *lags* significativos para nenhum valor de  $k$ , neste caso definidos como sendo os maiores que  $n^{-1/2}$ . Quando este for o caso, os valores observados na série temporal  $z_t$  foram transformados em um processo de ruído aleatório  $e_t$ .

#### 2.2.1.2.9 Previsão

Uma vez determinado o melhor modelo para a série temporal em estudo, pode-se usá-lo para gerar previsões de observações futuras, como é o propósito dessa pesquisa. Partindo-se do período atual  $t$ , e supondo que se deseja prever a série em um período futuro,  $\hat{z}_{t+k}$  representa a previsão para um período  $t+k$  feita em  $t$ .

A previsão para o período  $t+k$  é normalmente construída a partir de sucessivas previsões para os períodos  $t+1, t+2, \dots, t+k-1$  (MONTGOMERY; JOHNSON; GARDINER, 1990). Neste procedimento, o valor de  $z_{t+j}$ , o qual não se conhece no tempo  $t$ , é substituído pela sua previsão  $\hat{z}_{t+j}$ . O valor de  $a_{t+j}$ , o qual também não se conhece no tempo  $t$ , é substituído por zero, e  $a_{t-j} = z_{t-j} - \hat{z}_{t-j}$ . No início do processo de previsão, deve-se assumir que  $a_{t-j} = 0$  para  $t-j \leq 0$ .

#### 2.2.1.3 Modelos de Heterocedasticidade Condicional

A grande maioria da literatura de séries temporais não somente assume que as séries estacionárias (ou que podem ser transformadas em séries estacionárias utilizando alguma transformação simples), mas que estas também são processos lineares, de forma que podem ser representadas por modelos lineares (FERREIRA, 2006; CHEN et al., 2016; LYDIA et al., 2016).

Entretanto, não há nenhuma razão que generaliza a suposição de linearidade. Na realidade, é de se esperar que devido à natureza não-linear do meio que forma o cenário de acontecimentos da série temporal (mundo real), é bem mais natural se supor a existência de estruturas não-lineares (CLEMENTS; FRANCES; SWANSON, 2004). Algumas séries temporais, principalmente séries financeiras, apresentam a variância condicional evoluindo no tempo, desse modo, os modelos lineares do tipo ARIMA não são adequados para descrever esse tipo de comportamento (MORETTIN; TOLOI, 2006).

Há uma variedade muito grande de modelos não-lineares disponíveis na literatura, mas essa pesquisa abordará o mais conhecido deles, a classe de modelos ARCH (“*autoregressive conditional heterocedasticity*”), introduzida por Engle (1982) e suas extensões. Para detalhes de outros modelos, veja Tong (1990) e Tsay (2002).

Diante do exposto, uma serie de dados pode ser não linear de, pelo menos, duas formas: com relação a sua variância e/ou com relação a sua média. Quando a série é não linear com relação a variância, se utiliza a técnica de modelos autorregressivos com heterocedasticidade condicional (ARCH). Entretanto, os modelos ARCH contem apenas a parte autorregressiva. Em 1986, Bollerslev generalizou, adicionando uma parte de médias móveis. Estes são os modelos ARCH generalizados, ou GARCH, do original em inglês. Nesta classe de modelos existem ainda os ARCH-M (onde a média depende da variância condicional), IGARCH (modelo GARCH integrado), EGARCH (um modelo GARCH exponencial) entre outros que não serão abordados nessa pesquisa (MORETTIN; TOLOI, 2006; ENDERS, 1995).

O modelo ARCH de ordem  $q$  é definido como um modelo AR( $q$ ) onde  $\{r_t\}$  é condicionado à informação avaliada no tempo  $t-1$ , com variância, apresentada na Equação 50.

$$h_t = a_0 + \sum_{i=1}^q a_i r_{t-i}^2, \text{ com } a_0 > 0 \text{ e } a_i \geq 0 \quad (50)$$

Desta forma, a formulação do modelo ARCH é apresentada na Equação 51. Neste modelo a não-linearidade vem da variância condicional de  $\{r_t\}$ .

$$z_t = r_t \left[ a_0 + \sum_{i=1}^q a_i r_{t-i}^2 \right]^{\frac{1}{2}} \quad (51)$$

Para verificar se o modelo estimado é adequado, pode-se calcular a estatística Q de Ljung-Box para a sequência de  $z_t$ , ou calcula-se os coeficientes de assimetria e curtose estimados e/ou fazer um gráfico QxQ para avaliar a suposição de normalidade.

Desde o trabalho de Engle (1982), muitas variações do modelo ARCH têm sido propostas. Uma importante expansão do modelo ARCH é o modelo heterocedástico condicional autoregressivos geral (GARCH), sugerido por Bollerslev (1986), onde é permitido que a variância condicional  $h_t$  também seja uma função de si própria, isto é, uma função de  $h_{t-1}, \dots, h_{t-p}$  - modelo GARCH ( $p, q$ ). Um modelo GARCH é definido na Equação 52.

$$h_t = a_0 + \sum_{i=1}^r a_i X_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}, \text{ onde } X_t = \sqrt{h_t} \varepsilon_t \text{ com } a_0 > 0 \text{ e } a_i \text{ e } \beta_j \geq 0 \quad (52)$$

Para identificar um modelo GARCH, poderia ser utilizado o teste de Ljung-Box (Q) e também analisar-se o correlograma de resíduos (ENDERS, 1995; MORETTIN; TOLOI, 2006). Para maiores detalhes de identificação de modelos ARCH, GARCH e outras extensões dessa classe ver Morettin e Tolo (2006).

Os modelos ARCH e GARCH já foram aplicados com sucesso na economia nacional (SATO, 2013; CALDEIRA et al., 2014; FREITAS; SAFADI, 2015) e internacional (KRISTJANPOLLER; MINUTOLO, 2015; POKHILCHUK; SAVELEV, 2016; HUANG; LIU; WANG, 2016) contudo existem algumas desvantagens do ponto de vista técnico. Uma destas desvantagens vem da mistura de operações de adição e multiplicação, resultando em uma parametrização pouco natural e não intuitiva, levando a grandes complicações para se provar algum resultado aplicado a estes modelos.

Um segundo ponto vem do fato de que apenas momentos estatísticos de baixas ordens aparentam existir. De forma prática, o quarto momento já se apresenta como uma grandeza infinita, implicando que variâncias de algumas estatísticas importantes possam ser infinitas ou muito grandes. Outro aspecto negativo para estes modelos é que a suposição que a distribuição de  $\{r_t\}$  assintota à distribuição normal é fraca para amostras finitas.

Clements, Franses e Swanson (2004) e Coelho, Santos e Costa Jr. (2008) citam que desempenho de previsão para os modelos não-lineares estão abaixo do esperado, devido ao nível de complexidade adicionado ao problema de previsão com uma abordagem não-linear, quando comparado com uma abordagem linear, não há um ganho de desempenho significativo na previsão como poderia ser esperado.

De Gooijer e Kumar (1992), com o seu artigo de revisão a respeito dos modelos não-lineares, não conseguiram evidências claras em favor dos modelos não-lineares, quando comparados aos modelos lineares em termos de desempenho de previsão. Tal situação encontra-se em mudança quando observados estudos que realizam comparações entre métodos lineares e não lineares (ARRUDA; FERREIRA; CASTELAR, 2011; FERNANDEZ, 2014; ERAKER; WANG, 2015), sendo observado uma melhor acurácia para os modelos não-lineares.

Sendo assim, é um conceito global aceitar que o ambiente gerador das séries temporais é não-linear de forma geral, e portanto o problema pode ser simplesmente o fato que os atuais modelos não-lineares não consigam descrever a realidade de forma tão melhor que as simples aproximações lineares, sugerindo que uma nova abordagem deve ser desenvolvida.

Observa-se em particular nas ciências econômicas, como nas teorias macroeconômicas e financeiras, que um grande repositório de modelos não-lineares está em voga. Por exemplo, a maior parte dos modelos de ciclo real de negócios (como o objeto de estudo dessa pesquisa, setor imobiliário) são altamente não-lineares, como também os modelos de finanças (CLEMENTS; FRANCES; SWANSON, 2004). O uso destes modelos não-lineares na economia e finanças é totalmente consistente com o uso dos modelos lineares aplicados na prática, uma vez que modelos lineares podem ser vistos como uma aproximação dos fenômenos não-lineares de interesse.

Desta forma, na perspectiva de previsão, existe uma grande motivação para a pesquisa de modelos não-lineares, como destaca a edição especial do *International Journal of Forecasting* de abril/junho de 2004 - *Forecasting Economic and Financial Time Series Using Nonlinear Methods*, onde é esperado que a sofisticação dos testes, seleção e estimação de modelos não-lineares venha a aperfeiçoar o desempenho na previsão. Diante dessa motivação de pesquisa, foi de interesse abordar os métodos não-lineares nesse estudo e assim farão parte do projeto.

Destarte, não é surpreendente que modelos não-lineares nos últimos anos tenham tido grande atenção na literatura, até mesmo com a inclusão de novos paradigmas aplicados ao problema de previsão, como é o caso das redes neurais artificiais, outra classe de modelos que pelo fato do estudo dos modelos não-lineares serão abordados no próximo capítulo, para também serem testados na série temporal em estudo.

#### 2.2.1.4 Modelos de Redes Neurais Artificiais

Na Seção 2.2.1.4 serão apresentados os modelos de Redes Neurais Artificiais (RNAs). Em razão da extensão e complexidade desses modelos, serão abordados unicamente aspectos fundamentais para o conhecimento e utilização destes modelos no presente estudo. Para tanto, esta seção está dividida em duas outras seções. A primeira apresentará os aspectos conceituais envolvendo as RNAs, já a segunda apresentará a previsão utilizando RNAs.

##### 2.2.1.4.1 Aspectos conceituais sobre Redes Neurais Artificiais

Uma Rede Neural Artificial é um processador maciçamente paralelamente distribuído de unidades de processamento simples, que possuem a propensão natural para armazenar conhecimento experimental e torná-lo disponível para o uso (HAYKIN, 1999). Desta forma, trata-se de um conjunto de entrelaçamentos, sendo que cada conexão ou sinapse apresenta um peso associado para o ajuste dos dados de entrada aos dados de saída (SOBREIRO; ARAÚJO; NAGANO, 2009). As RNAs são treinadas por intermédio de algumas regras de treinamento que ajustam os pesos das conexões conforme os dados disponíveis são minimizados por uma função de erro apropriada (ABURTO; WEBER, 2007).

O cérebro humano é um sistema de processamento de informações altamente complexo, não-linear e paralelo. As RNAs visam modelar como o cérebro humano realiza determinados processamentos (chamadas redes neurais naturais) e, sendo assim podem ser projetadas e implementadas por intermédio de componentes eletrônicos como por programação computacional (HAYKIN, 1999; FLORES, 2009).

Essa não-linearidade de séries temporais faz com que tais modelos sejam utilizados também para a previsão de demanda, onde os modelos lineares (como por exemplo: modelos de SE e ARIMA) não capturam tal padrão. Zhang (2003) reforça que o poder destes modelos como previsores, deve-se ao processamento paralelo das informações provenientes dos dados.

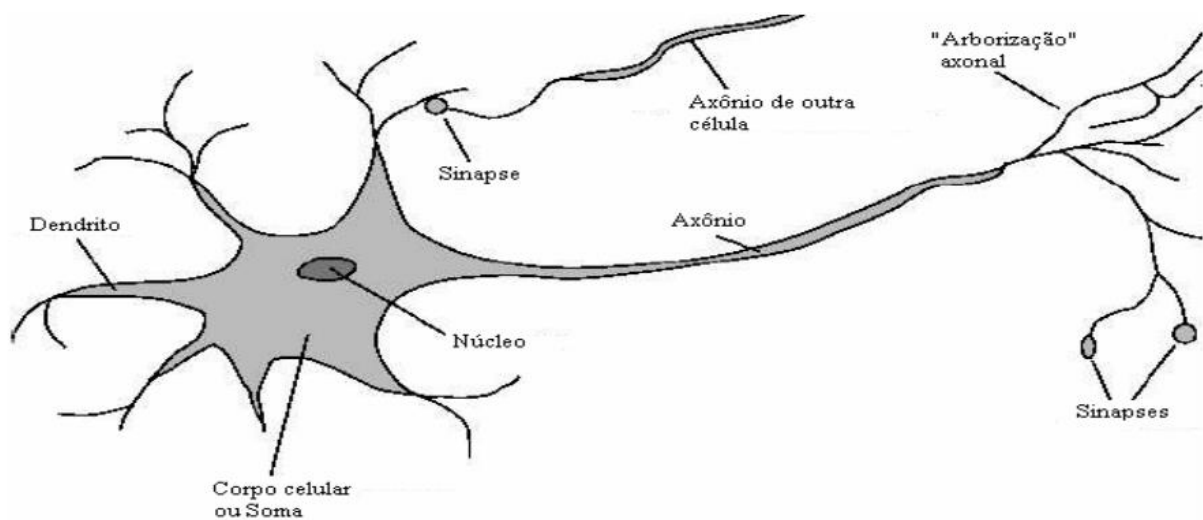
A motivação para a descoberta das RNAs são as Redes Neurais Naturais (RNNs), sendo que as RNAs visam reproduzir as funções das RNNs, buscando implementar seu comportamento funcional e sua dinâmica. Porém, o funcionamento das RNNs ainda não é totalmente conhecido pelo homem, no entanto, o comportamento individual do neurônio é bem entendido do ponto de vista funcional, fato este que se baseia a teoria sobre RNAs (BRAGA; CARVALHO; LUDERMIR, 2011). Para ilustrar tal fenômeno, pesquisadores



apresentam um neurônio biológico explicando suas respectivas funções, o qual pode ser visualizado na Figura 11.

Os neurônios biológicos são divididos em três seções: (i) o corpo celular; (ii) os dendritos; e, (iii) o axônio. A função dos dendritos é receber as informações, ou impulsos nervosos, oriundas de outros neurônios e conduzi-las até o corpo celular. No corpo celular a informação é processada, gerando novos impulsos. O axônio tem por função transmitir estes impulsos até outros dendritos dos neurônios seguintes e, neste ponto de contato surge aquilo que se chama de sinapse (BRAGA; CARVALHO; LUDERMIR, 2011).

Figura 11 - Neurônio biológico



Fonte: Araújo (2005, p. 26).

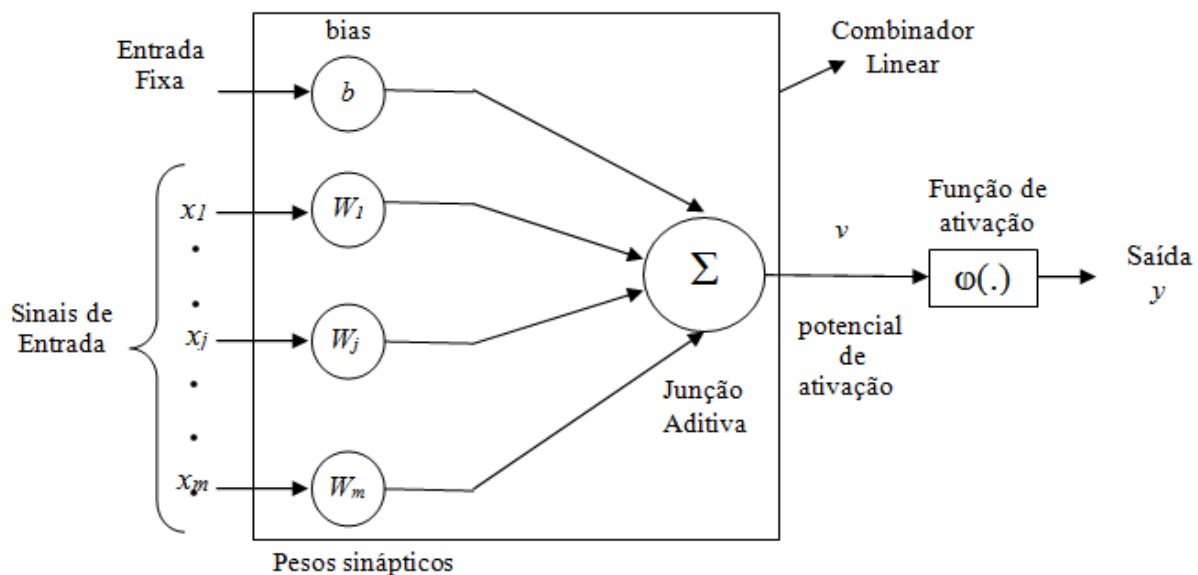
Já os neurônios artificiais são compostos pelos seguintes elementos: (i) um conjunto sinapses ou elos de conexão; (ii) um somador dos sinais de entrada; e, (iii) uma função de ativação. As sinapses são pesos ou forças que serão multiplicadas por um determinado sinal de entrada ( $x_j$ ) sendo que os pesos sinápticos podem assumir tanto valores positivos quanto negativos em determinado intervalo. O elemento somador é aquele que realizará a combinação linear de todos os sinais de entrada ( $x_j$ ) com seus pesos sinápticos, tratando-se, portanto, de uma soma ponderada.

A função de ativação é um elemento utilizado para restringir a amplitude da saída de um neurônio (HAYKIN, 1999). É perceptível que a definição de neurônio artificial dada por Haykin (1999) é semelhante àquela dada por Braga, Carvalho e Ludermir (2011) para o

neurônio biológico, inclusive, seguindo a mesma lógica de explicação. Para ilustrar, a Figura 12 apresenta um modelo de neurônio artificial.

O modelo de neurônio artificial assim proposto é associado computacionalmente a uma estrutura matricial, com as entradas  $x_1, x_2, x_3, \dots, x_m$  e uma saída  $y$ . Cada entrada tem seu peso associado  $w_1, w_2, w_3, \dots, w_m$  e que pondera quantitativamente cada entrada em relação à saída. O produto da entrada pelos pesos determina a primeira operação matemática de um neurônio artificial, juntamente com um limiar de ativação do neurônio, o *bias*. O *bias* tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação, dependendo se ele é positivo ou negativo (ARAÚJO, 2005; HAYKIN, 1999).

Figura 12 - Modelo de neurônio artificial



Fonte: Adaptado de Haykin (1999, p. 38).

Araújo (2005) apresenta o conjunto de Equações 53, 54 e 55 que descrevem um neurônio artificial e ressalta que o resultado dessa soma  $v$  é a variável independente de uma função de ativação  $\phi(\cdot)$ , que faz correspondência do somatório anterior com a saída  $y$ , atuando de modo a limitar a amplitude do sinal de saída.

$$u = \sum_{j=1}^m x_j w_j \quad (53)$$

$$v = (u + b) \quad (54)$$

$$y = \varphi(.) \quad (55)$$

Nesta seção abordou-se o conceito de RNA, explicando as suas relações com o sistema biológico, do mesmo modo, apresentou-se a motivação que originou as pesquisas nesta área. Na próxima seção serão apresentados aspectos importantes a serem delineados para um adequado entendimento sobre a previsão de série temporais utilizando RNAs.

#### 2.2.1.4.2 Previsão de séries temporais utilizando Redes Neurais Artificiais

O emprego de RNAs para previsão de séries temporais, segundo Zhang, Patuwo e Hu (1998), ocorreu em 1968 com o trabalho de Hu que realizava previsões referentes ao clima. Conforme Allemão (2004), geralmente, as RNAs são aplicadas a problemas relacionados à clusterização, à previsão de séries temporais e ao reconhecimento. O autor complementa que para o exercício de previsão, utilizam-se as redes conhecidas como aproximadoras de funções como a MLP e redes Elman.

Recentemente, as RNAs têm sido extensivamente estudadas e usadas em séries temporais para a previsão de demanda (ZHANG, 2003; KHASHEI; BIJARI, 2010) e, uma das vantagens mais significativas das RNAs é a sua capacidade para aproximar funções não-lineares (como muitas vezes são as séries temporais) com alto grau de acuracidade (KHASHEI; BIJARI, 2010; ZHANG, 2003), situação esta não verificada nos modelos ARIMA (LEE; TONG, 2011).

Para Khashei e Bijari (2010), o poder destes modelos para a modelagem de séries temporais provém do processamento paralelo realizado pela RNA. Desta forma, Aburto e Weber (2007) fazem uma comparação entre os modelos ARIMA e os modelos MLP de RNAs, apresentado no Quadro 3.

Uma RNA largamente utilizada em pesquisas que envolvem modelagem e previsão por intermédio de séries temporais é a *feedforward single layer* ou *feedforward* de uma única camada escondida (como é o caso de uma rede MLP de única camada escondida).

Quadro 3 - Comparação entre modelos ARIMA e MLP

ARIMA	MLP
Modelo linear: assume, <i>a priori</i> , o comportamento da série temporal	Modelo não-linear: mais graus de liberdade para o Modelo
Modelagem requer estacionariedade da série	Não requer análise da série temporal
Requer interação com o usuário	Requer menos interação com o usuário
O modelo proporciona <i>insight</i> e informação através dos parâmetros	Dificuldade para interpretar o modelo
Não há superajustamento aos dados de treinamento	Superajustamento é possível

Fonte: Aburto e Weber (2007, p. 140).

Tal modelo é descrito por três camadas (a de entrada, a escondida e a de saída) e, a relação entre as saídas da rede e suas entradas pode ser descrita matematicamente pela Equação 56 (ZHANG, 2003; KHASHEI; BIJARI, 2010).

$$y_t = w_0 + \sum_{j=1}^q w_j g \left( w_{0j} + \sum_{i=1}^p w_{ij} y_{t-i} \right) + \varepsilon_t \quad (56)$$

Onde  $w_j$  ( $j = 0, 1, 2, \dots, q$ ) e  $w_{ij}$  ( $i = 0, 1, 2, \dots, p; j = 1, 2, 3, \dots, q$ ) são os parâmetros do modelo geralmente chamados de *connection weights*;  $p$  é o número de neurônios na camada de entrada e o  $q$  é o número de neurônios na camada escondida (KHASHEI; BIJARI, 2010; ZHANG, 2003). A função de transferência de regra utilizada na camada escondida é a função logística (ZHANG, 2003), apresentada na Equação 57. Na camada de saída, a função mais utilizada é a linear, uma vez que funções não-lineares podem trazer distorções à capacidade preditiva da RNA (KHASHEI; BIJARI, 2010).

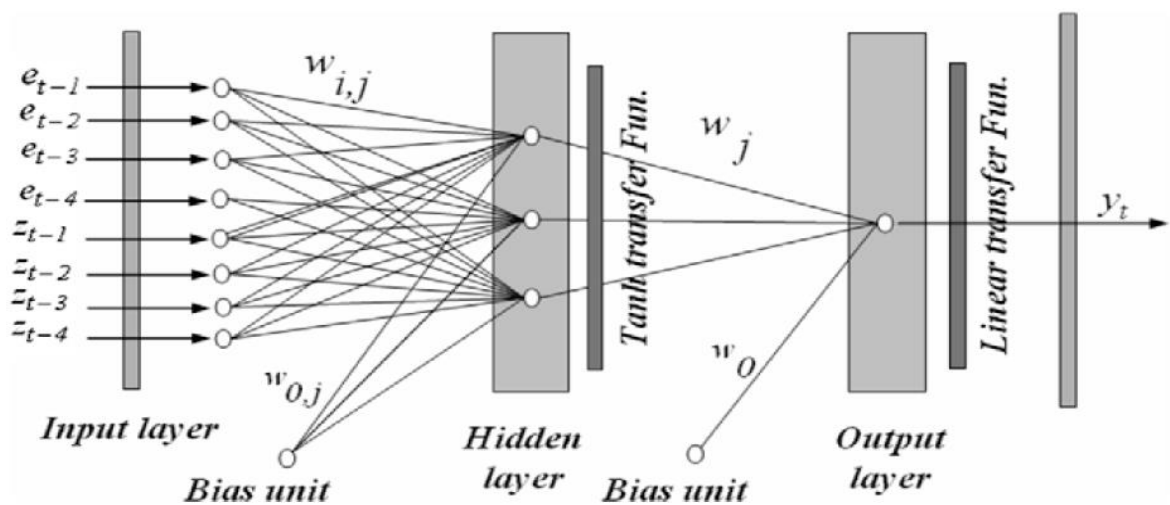
$$g(x) = \frac{1}{1 + \exp(-x)} \quad (57)$$

O modelo de RNAs apresentado na Equação 56 modela a função não-linear presente no processo gerador da série, com base nas observações passadas, conforme apresentado na sequência pela Equação 58 desta para a previsão de seus valores futuros. Geralmente utiliza-se um neurônio na camada de saída e, este tipo de modelagem chama-se *one step ahead* (ZHANG, 2003; ZOU et al., 2007; KASHEI; BIJARI, 2010).

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}, w) \quad (58)$$

Onde  $w$  é o vetor de todos os parâmetros e  $f$  é a função determinada pela estrutura e conexões sinápticas da RNA. Desta forma, o presente modelo torna-se similar a um modelo não-linear autorregressivo (ZHANG, 2003; KHASHEI; BIJARI, 2010). A topologia de uma rede MLP é apresentada em Khashei e Bijari (2010), conforme é visualizado na Figura 13.

Figura 13 - Modelo estrutural de uma RNA para a previsão de séries temporais



Fonte: Khashei e Bijari (2010, p. 5).

Quanto à topologia de rede MLP, Aburto e Weber (2007) ressaltam que esta é a mais utilizada dentre os modelos de RNAs. A estrutura de redes da MLP tem sido utilizada em diversas pesquisas com RNAs para a previsão de demanda. O papel das redes *feedforward* de múltiplas camadas (como as redes MLP) é o seguinte:

[...] transformar, sucessivamente, o problema descrito pelo conjunto de dados no espaço de entrada em uma representação tratável para a camada de saída da rede. Através de transformações sucessivas nos espaços de representação de cada camada intermediária o problema pode, então, ser tratado pela camada de saída. Assim, por exemplo, um problema não-linearmente separável, resolvido por uma rede de duas camadas, é transformado em um problema linearmente separável pela camada intermediária, criando uma nova disposição interna à rede para os dados de entrada. A partir dessa nova disposição, linearmente separável, a camada de saída pode resolver o problema descrito no espaço de entrada (BRAGA; CARVALHO; LUDERMIR, 2011, p. 67).

Conforme Araújo (2005), a topologia de rede MLP é uma das mais importantes redes da classe de múltiplas camadas alimentadas adiante. Aburto e Weber (2007) ressaltam que a

topologia MLP também é a mais popular dentre os modelos de RNAs. Araújo (2005) cita as principais características destes modelos:

- a) cada neurônio da rede possui uma função de ativação não-linear (função logística ou sigmóide);
- b) a rede contém uma ou mais camadas de neurônios ocultos; e
- c) rede exibe um alto grau de conectividade, determinado por suas sinapses.

A utilização da topologia MLP para a previsão de séries temporais requer a especificação de dois parâmetros. O primeiro parâmetro ( $k$ ) indica o horizonte de tempo que será utilizado como *input* para a predição da série temporal. O segundo parâmetro ( $s$ ) especifica o número de períodos futuros aos qual a rede irá realizar a previsão (ABURTO; WEBER, 2007). Ansuji et al. (1996) expõem que uma rede MLP ajusta-se ela mesma como um modelo de previsão de séries temporais da seguinte forma:

- a) As unidades de entrada são constituídas de informações relevantes a respeito da série temporal;
- b) Os pesos são os parâmetros dos modelos e serão estimados a partir do treinamento da rede que levará em conta o par de entradas e seus respectivos valores de saída (valores reais da série de tempo);
- c) As camadas escondidas são as conexões entre as camadas de entrada e as camadas de saída. Trata-se de uma questão importante para a seleção de melhores conjuntos de pesos, desde que a não-linearidade do modelo possa ser localizada na função de ativação das unidades escondidas;
- d) A camada de saída é constituída de somente uma unidade e traz a informação necessária para a previsão;
- e) A rede é treinada para estimar os parâmetros do modelo. Depois do treinamento, as previsões para os períodos seguintes no futuro podem ser geradas na camada de saída; e
- f) Os parâmetros de ativação são valores fornecidos pelo usuário que em geral são constantes. Tais parâmetros são: taxa de aprendizagem, *momentum* e tamanho do intervalo de variação dos pesos.

Uma vez definida a estrutura da rede, esta estará pronta para o processo de treinamento, processo este que ajustará os pesos internos da rede conforme uma função de custo vai sendo minimizada. A função de custo é um critério que busca a minimização dos erros quadrados da rede, apresentada na Equação 59, sendo que a minimização é realizada por intermédio de um eficiente algoritmo de treinamento não-linear, o *back-propagation*, demonstrado na Equação 60 e, para acelerar o processo de treinamento, evitando também a instabilidade do algoritmo, introduz-se um termo chamado *momentum*  $\delta$ , exibido na Equação 61 (KOVÁCS, 2006; ABURTO; WEBER, 2007; KHASHEI; BIJARI, 2010).

$$E = \frac{1}{N} \sum_{n=1}^N (e_i)^2 = \frac{1}{N} \sum_{n=1}^N \left( y_t - \left( w_0 + \sum_{j=1}^Q w_j g \left( w_{0j} + \sum_{i=1}^P w_{ij} y_{t-i} \right) \right) \right)^2 \quad (59)$$

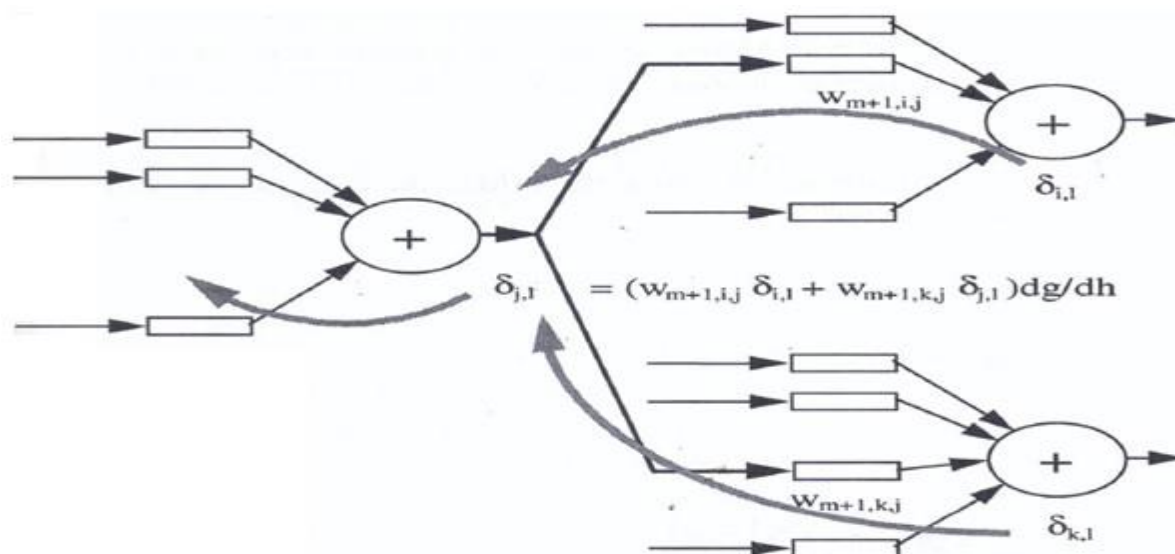
$$\Delta w_{ij} = -n \frac{\partial E}{\partial w_{ij}} \quad (60)$$

$$\Delta w_{ij}(t+1) = -n \frac{\partial E}{\partial w_{ij}} + \delta w_{ij}(t) \quad (61)$$

O *back-propagation* é considerado um algoritmo de treinamento do tipo supervisionado, pois se utiliza das informações dos padrões de entrada fornecidos à rede e a sua respectiva saída desejada e que, por meio de um mecanismo de correção de erros (gradiente descendente), ajusta os pesos da rede aos padrões de entrada fornecidos na etapa de treinamento.

Assim, o treinamento é realizado em duas etapas: (i) a fase *forward* e (ii) a fase *backward*. Na primeira fase, são apresentados conjuntos de dados à rede e esta propaga o sinal até a sua camada de saída. A segunda fase realiza a correção dos erros, alterando os pesos internos da rede, conforme a saída desejada (BRAGA; CARVALHO; LUDERMIR, 2011). A Figura 14 ilustra o processo de retro propagação do erro em uma rede do tipo MLP.

Figura 14 - A retro propagação do erro em uma rede MLP



Fonte: Kovács (2006, p. 79).

#### 2.2.1.5 Combinação de previsão

Em séries temporais, os métodos de Suavização Exponencial e a metodologia de Box-Jenkins são as duas classes de modelos mais difundidas para previsões de séries temporais (MORETTIN; TOLOI, 2006). No entanto, estes métodos possuem a limitação de serem técnicas lineares, isto é, inadequadas para muitos problemas reais. Sob esta ótica, Kashefi e Bijari (2010) sugerem que as Redes Neurais Artificiais podem ser uma eficaz opção aos métodos lineares, devido a sua capacidade de modelagem não-linear como foi visto na seção anterior.

Entretanto, desde Bates e Granger (1969) as técnicas de previsões deixaram de ser estudadas apenas de forma individual. Para incorporar diferentes previsões originadas de diferentes técnicas, tem-se utilizado a combinação de previsões. A combinação de previsões é apresentada como uma alternativa satisfatória diante de previsões individuais (HIBON; EVGENIOU, 2005). Estudos apresentam resultados em que as diversas combinações de previsões avaliadas apresentaram acurácia superior à das previsões individuais (MAKRIDAKIS; HIBON, 2000; ANDRAWIS; ATIYA; EL-SHISHINY, 2011; MARTINS; WERNER, 2014; KATRIS, DASKALAKI, 2015; VORTELINOS, 2017). Diversos métodos de combinação de previsões ganharam destaque por seu desempenho.

O método de combinação inicialmente proposto por Bates e Granger (1969) consiste na obtenção do peso ótimo e combinação linear de duas previsões e é conhecida como método



de Variância Mínima (VAM). A obtenção do peso ótimo é baseada nas variâncias dos erros das previsões individuais, o cálculo dos pesos pode ser observado na Equação 62.

$$k_p = \frac{(\sum_{i=1}^N e_{pi}^2)^{-1}}{\sum_{p=1}^p (\sum_{i=1}^n e_{pi}^2)^{-1}} \quad (62)$$

Onde  $k_p$  é o peso de cada  $p$  método para a combinação linear da previsão e  $e_{pi} = (y_i - \hat{y}_{pi})$  é o erro da previsão do método  $p$  no período  $i$ . Tendo encontrado o peso de cada método, a Equação 63 apresenta como é realizada a previsão por meio da combinação linear. Onde  $\hat{C}_i$  é a previsão combinada no período  $i$ .

$$\hat{C}_i = \sum_{p=1}^p k_p \hat{y}_{pi} \quad (63)$$

A Média Aritmética (MT) é uma técnica de combinação de fácil utilização e que tende apresentar desempenho satisfatório com relação a técnicas mais robustas. Algumas pesquisas tem comparado o desempenho da combinação da média aritmética e concluído que esta é uma boa alternativa quando se almeja a utilização de um método simples (STOCK; WATSON, 2004). A combinação das previsões de  $n$  modelos por meio da MT é dada na Equação 64.

$$\hat{C}_i = \frac{\hat{x}_i^{(1)} + \hat{x}_i^{(2)} + \dots + \hat{x}_i^{(n)}}{n} \quad (64)$$

onde  $\hat{C}_i$  é a combinação das previsões no período  $i$  obtidas por  $\hat{x}_t^{(1)}$ , até  $\hat{x}_t^{(n)}$ , sendo a previsão pelo método 1 no período  $i$ , somada até a previsão do  $n$  método.

Alguns estudos englobaram outros métodos de combinação, além da MT, como a Média Geométrica (MG) e Média Harmônica (MH), tendo em vista que estas são mais sensíveis a valores de previsão menores (PATTON; SCHEPPARD, 2009; ANDRAWIS; ATIYA; EL-SHISHINY, 2011). Nas Equações 65 e 66 são dadas as combinações das previsões por meio dos métodos citados, respectivamente.

$$\hat{C}_i = \left( \hat{x}_i^{(1)} \hat{x}_i^{(2)} \dots \hat{x}_i^{(n)} \right)^{\frac{1}{n}} \quad (65)$$

$$\hat{C}_i = \frac{n \hat{x}_i^{(1)} \hat{x}_i^{(2)} \dots \hat{x}_i^{(n)}}{\hat{x}_i^{(1)} + \hat{x}_i^{(2)} + \dots + \hat{x}_i^{(n)}} \quad (66)$$

Outro método comumente utilizado é o da classe dos modelos de regressão, conforme estudado por Granger e Ramanathan (1984). Segundo os autores, a principal vantagem deste tipo de método na combinação das previsões é que mesmo previsões enviesadas geram combinações não enviesadas, caracterizando-se como uma alternativa aos métodos de otimização, principalmente. Na Equação 67 é dada a técnica de combinação do tipo Regressão Linear (RL) abordada neste estudo.

$$\hat{C}_i = w_0 + w_1 \hat{x}_i^{(1)} + w_2 \hat{x}_i^{(2)} + \dots + w_n \hat{x}_i^{(n)} \quad (67)$$

Os valores dos coeficientes ( $w$ ) da equação de regressão podem ser estimados aplicando o método dos mínimos quadrados, o qual busca encontrar os coeficientes que minimizem a soma dos quadrados dos erros (MONTGOMERY et al., 2006). Após a estimação dos coeficientes do modelo de regressão para determinar os valores de  $\hat{C}$ , testa-se a significância do modelo de regressão analisando o Teste F, sendo necessário atender os pressupostos da regressão juntamente com os pressupostos dos resíduos. Maiores detalhes são apresentados em Charnet et al. (2008).

Outro método utilizado para combinar previsões gerando bons resultados é a Análise de Componentes Principais (ACP) (SKITTIDES; FRUH, 2014). O método consiste na transformação do conjunto original de variáveis em outro, os componentes principais, de dimensões equivalentes, porém com a propriedade de que cada componente retém porcentagem da variância original e que as variâncias são decrescentes do primeiro ao último componente principal, ou seja, o primeiro componente possui a maior variância, o segundo possui a segunda maior, e assim por diante.

Portanto, cada componente principal é combinação linear das variáveis originais, independente entre si e estimado com o propósito de reter, em ordem de estimação, o máximo da informação, em termos de variação total, contida nos dados iniciais. A combinação de previsões via ACP, consiste na soma das previsões individuais obtidas, ponderadas por um peso distinto para cada previsão individual. Outros detalhes podem ser observados em Vicini (2005) e Martins (2014).

É interessante salientar que é possível realizar hibridizações com os modelos individuais, desse modo, autores afirmam os modelos híbridos possibilitam a solução de problemas de previsão com resultados melhores em termos de acurácia, além de focarem diferentes níveis de agregação e horizonte temporais das previsões (CHANG; LIU; FAN, 2009; FARUK, 2009; HICHAM; MOHAMED; ABDELLAH, 2012; BABU; REDDY, 2014; PRAHLADA; DEKA, 2015; GUO; LIU; SUN, 2016; SHEN et al., 2017).

## 2.2.2 Medidas de Acurácia

Esta seção voltar-se-á as atenções para tratar de um tópico relevante no campo de estudo das previsões quantitativas, a mensuração do grau de adequabilidade e utilidade de um método de previsão a um conjunto de dados, a série temporal. Haja vista a importância desse tópico, no contexto da qualidade de veracidade desejadas neste trabalho, foi necessária, então, que se definisse uma estratégia de avaliação de modo conseguir um diagnóstico confiável e realista dos resultados.

Primeiramente, há a necessidade de se declarar que não se priorizará verificar o grau de aderência dos modelos aos dados utilizados no processo de calibração (quantificação dos erros de ajuste aos dados do conjunto inicialização) e sim com o poder real de previsão de cada modelo aos dados preditos. Entretanto, os cálculos para ajuste da série serão apresentados nos resultados, porém a escolha do melhor modelo será feita em relação aos erros de previsão e não de ajuste. A seguir, apresentam-se as medidas e estatísticas utilizadas durante o processo de avaliação e seleção dos modelos.

### 2.2.2.1 Medidas estatísticas padrão

Dependendo do comportamento da série temporal que se deseja analisar, vários modelos podem ser empregados na previsão de seus valores futuros (MAKIDRAKIS; WHEELWRIGHT; HYNDMAN, 1998). A escolha do modelo mais apropriado, no entanto, é baseada a partir da análise dos erros gerados por cada modelo ( $e_t = Z_t - \hat{Z}_t$ ). Uma vez que o cálculo dos erros pode resultar em valores positivos ou negativos, o que pode, em algumas situações, zerar o seu somatório, diferentes manipulações da estatística de erro ( $e_t$ ) devem ser empregadas. Estas diferentes manipulações constituem-se em critérios para escolha de modelos mais apropriados às séries temporais.

Para selecionar modelos com diferentes números de parâmetros, os critérios *Akaike Information Criterion* (AIC) e *Bayesian Information Criterion* (BIC) podem ser utilizados, pois modelos com muitos parâmetros ou não estacionários obtêm maior penalidade (DURBIN; KOOPMAN, 2008). Souza (2006) cita não serem raros os casos em que não se consegue identificar um único modelo, e sim vários modelos candidatos a gerador da série em estudo. Isso porque, trabalhando com a ACF e a PACF, fica difícil, muitas vezes, decidir se elas são decrescentes ou se são truncadas. Esses critérios são geralmente utilizados para selecionar o melhor modelo de Box-Jenkins, em vez de estabelecer  $p$  e  $q$  precisamente, estimam-se os modelos correspondentes a vários pares  $(p, q)$  e escolhe-se aquela especificação que apresentar o menor valor para AIC ou para BIC.

Na Equação 68 e 69 são apresentadas o cálculo dos critérios AIC e BIC, respectivamente.

$$AIC = \ln \hat{\sigma}_\varepsilon^2 + \frac{2(p + q)}{n} \quad (68)$$

$$BIC = \ln \hat{\sigma}_\varepsilon^2 + \frac{(p + q)\ln(n)}{n} \quad (69)$$

A crítica que se faz, a esses critérios, é que podem conduzir a modelos super especificados, ou seja, modelos com valores de  $p$  e/ou  $q$  maiores do que o correto. Assim sendo, os critérios AIC e BIC devem ser usados como procedimento complementar, e não alternativo aquele baseado na ACF e na PACF (SOUZA, 2006).

Após a identificação, verificação, estimação e previsão dos modelos é preciso comparar quais modelos fornecem previsões mais próximas dos valores reais observados. Os métodos de comparação dos valores previstos com os valores observados são os que caracterizam com menos erros a capacidade preditiva dos modelos aplicados (SIQUEIRA, 2002).

Para Tubino (2004) uma forma de monitorar o desempenho do modelo é por meio da verificação do comportamento do indicador de erro de previsão é *Mean Absolute Percentage Error* (MAPE), o qual considera desvios relativos ou percentuais no cálculo do erro. Na Equação 70 é possível verificar a fórmula para calcular o MAPE.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{e_t}{z_t} \times 100 \right| \quad (70)$$

Outro indicador de erro amplamente utilizado é *Mean Absolute Error* (MAE), que pode ser visualizado na Equação 71.

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t| \quad (71)$$

Outra métrica utilizada é *Root Mean Square Error* (RMSE), sua formulação é apresentada na Equação 72.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \quad (72)$$

No meio das medidas mencionadas para se mensurar a acurácia de previsão, a mais eficiente e utilizada é a MAPE (BOPP, 1985; LAWRENCE; EDMUNDSON; O'CONNOR, 1986; RINGUEST; TANG, 1987; ARMSTRONG; COLLOPY, 1992; KAHN, 1998). Esta afirmação deve-se ao fato de que as medidas RMSE e MAE são influenciadas pela escala dos números a serem trabalhado, sendo assim, não permitem que se proceda a análises comparativas acerca de diferentes séries temporais em diferentes intervalos de tempo. Porém, a MAPE carrega consigo a desvantagem de que, quando a série temporal contém valores iguais a zero, torna-se impossível o uso de sua fórmula.

O nível de acurácia requerido no resultado da previsão é talvez o mais importante pré-requisito para avaliação de métodos de previsão, pois o custo de melhorar a acurácia de uma previsão pode exceder os benefícios gerados pelo método (MENTZER; GOMES, 1989; LEMOS, 2006).

#### 2.2.2.2 Estatística U de Theil

Nenhuma das medidas citadas anteriormente dá uma boa base de comparação ordinal quanto aos ganhos de acuracidade auferidas pelo uso de determinado método de previsão em detrimento de outro. Por exemplo, será que pelo fato de se observar um valor de cinco unidades para a estatística MAE, ou um percentual de 3,2% para a estatística MAPE já

permite dizer que um determinado método é um bom previsor para um processo qualquer? Dentro deste contexto a estatística U de Theil é adequada, pois permite uma comparação relativa entre métodos de previsão formais. Matematicamente, a estatística U de Theil é definida na Equação 73 (THEIL, 1966; MAKIDRAKIS; WHEELWRIGHT; HYNDMAN, 1998).

$$U_{Theil} = \sqrt{\frac{\sum_{t=1}^n (z_t - \hat{z}_t)^2}{\sum_{t=1}^n (z_t - z_{t-1})^2}} \quad (73)$$

Quanto mais próximo o valor de  $U_{Theil}$  for de zero, maior será o ajustamento da série prevista em relação a série original. Em contrapartida, valores mais próximos ou maior que a unidade indicam que o modelo não conseguiu gerar boas previsões, sendo assim, uma previsão ingênua é melhor que a gerada pelo modelo (MAKIDRAKIS; WHEELWRIGHT; HYNDMAN, 1998).

### 2.3 PREVISÃO DE DEMANDA NO MERCADO IMOBILIÁRIO

Na seção 2.3 é apresentada de forma de integralizar a seção 2.1 e 2.2, explanando trabalhos desenvolvidos relacionados à previsão de demanda no mercado imobiliário.

Pagourtzi et al. (2008) demonstraram a capacidade de previsão do sistema Pythia<sup>1</sup>, comparando técnicas de previsão de séries temporais, que foram utilizadas para prever os empréstimos hipotecários no Reino Unido, a fim de mostrar como o Pythia pode ser útil para um banco. Os autores concluem que o método de Winters foi o que produziu melhor previsão, devido ao fato da existência de sazonalidade nos dados.

Ho e Addae-Dapaah (2014) em seu artigo ajudaram a compreender o ciclo imobiliário para escritórios por meio da aplicação de um modelo de Vetor Autorregressivo (VAR). Os autores estudaram as cidades de Hong Kong, Kuala Lumpur e Singapura, encontrando quatro resultados. Um deles, é que o ciclo imobiliário é geralmente diferente do ciclo de negócios subjacentes nessas cidades. Dois, o ciclo imobiliário é mais excessivo nas áreas de construção e desenvolvimento do que nos aluguéis e vacância. Três, o ciclo de vacância tende a conduzir o ciclo de aluguel. E quatro, novas construções tendem a atingir o pico quando o ciclo de

---

<sup>1</sup> PYTHIA é uma plataforma de previsão de negócios projetada e desenvolvida na Unidade de Sistemas de Previsão da Universidade Técnica Nacional de Atenas.

vacância também está em pico. Os autores concluem que é possível analisar e estimar o ciclo imobiliário por meio de modelos VAR.

Dentro do mesmo assunto, Miller (2014), em seu estudo revela que a demanda por espaço (área) por trabalhador em escritórios vai diminuir no futuro (e que isso já está acontecendo), cerca de 30% em 20 anos. Assim, as incorporadoras terão que se adequar para oferecer outro tipo de imóvel para as empresas.

Assimakopoulus, Hatzichristos e French (2003) objetivaram fornecer uma visão geral dos métodos, disponíveis na literatura, para realizar avaliação real dos preços de imóveis. Salientando que os métodos de avaliação podem ser tradicionais e avançados, chegaram à conclusão que dentre os tradicionais, o método de regressão, comparando, custo, receita e lucro é o mais utilizado, dentre os métodos avançados, foram encontrados RNAs, modelos de preços hedônicos, métodos de análise espacial, lógica *fuzzy* e modelos ARIMA. Pode-se observar que dois dos métodos avançados citados nesse trabalho serão utilizados na atual pesquisa para a previsão de demanda.

A expectativa do envelhecimento da população aumenta a necessidade de modelos precisos e previsões confiáveis de demanda habitacional (GIBLER et al., 2009; GIBLER; CLEMENTS, 2011). Gibler e Clements (2011) objetivaram analisar se o único modelo de previsão publicado, o qual prevê o movimento real da habitação dos americanos mais velhos em situação de aposentadoria. Este modelo baseia-se em uma regressão logística, sendo usado para testar as causas que influenciam na escolha da habitação na aposentadoria usando dados de 2002 e 2004. Os testes aplicados mostraram que são necessárias mais pesquisas para desenvolver modelo, para poderem estimar com mais precisão a demanda do mercado da habitação nos países em envelhecimento.

Adicionalmente, Csesko e Reed (2009) citam que as incorporadoras não irão conseguir suprir a demanda por imóveis para aposentados, devido ao fato da falta de estudos sobre essa demanda. Semelhante, Ball e Nanda (2013) realizaram um estudo sobre a demanda habitacional para pessoas aposentadas, tendo como objetivo relatar a necessidade de estimar a demanda futura de habitação especializada para aposentados utilizando dados para a Inglaterra. Eles concluíram que há sim a necessidade de criação de métodos especializados para prever a demanda para esse setor, já que a demanda por habitações para pessoas aposentadas está sendo grande.

De La Paz (2014) desenvolveu uma equação para estimar a nova elasticidade da oferta de habitação na Espanha. O objetivo do trabalho foi avaliar o papel da oferta de habitação na evolução dos preços. Os resultados mostraram que o choque da crise de crédito teve vários

graus de gravidade em regiões espanholas, reduzindo drasticamente a construção de casas por causa da alta sensibilidade a alterações nos preços. Essa estimação da elasticidade pode ser usada para prever respostas às mudanças nos preços das casas. Os resultados contribuem para a compreensão do mecanismo de equilíbrio no mercado da habitação em todas as regiões. A autora enfatiza que este é o primeiro artigo que analisa a oferta de habitação, calcula elasticidades de oferta e mede o impacto da crise de crédito sobre o mercado imobiliário do lado da oferta na Espanha.

Huang, Wang e Gai (2011) analisaram os principais fatores que influenciam a demanda do mercado imobiliário, a pesquisa estabeleceu um modelo de análise de demanda imobiliária fundada na base da Teoria dos Sistemas Cinza para demonstrar os fatores que influenciam o mercado imobiliário da China e, finalmente, descobre que fatores são mais importantes. A tendência da demanda por imóveis nos três anos subsequentes é analisada usando a combinação de método dos mínimos quadrados e método de regressão. Os autores concluem que a renda per capita da população e o preço dos imóveis são os dois fatores preponderantes para modificar a demanda.

Liu e Li (2009) propuseram uma equação de regressão múltipla para analisar os fatores que afetam as vendas de imóveis em Pequim após a realização dos Jogos Olímpicos de 2008. O rendimento per capita, tamanho da população e o preço médio da habitação comercial foram escolhidos como variáveis independentes, os autores concluíram que o modelo serviria para prever a demanda do mercado imobiliário em Pequim sendo útil para a tomada de decisão do governo e empresas imobiliárias. Corroborando, Yao (2009) cita que depois dos Jogos Olímpicos os preços dos imóveis em Pequim subiram exponencialmente e naquele ano seria necessária uma regulação governamental para uma melhora na política macroeconômica.

Dentro do mesmo assunto, Hyung e Jung (2013) apresentaram um novo modelo de simulação desenvolvido para considerar vários ambientes de longo prazo para a previsão de demanda habitacional. Os autores realçaram que o aumento das famílias unipessoais e a sociedade em envelhecimento modifica a demanda. O modelo foi aplicado ao mercado imobiliário da cidade portuária sul coreana, Pusan, com o objetivo de estimar a preferência de preço baseando-se na oferta de habitação e do número de agregados familiares. O método testado teve resultado satisfatório, porém ainda aquém do que os autores consideravam ideal.

Em relação a estudos nacionais são raras as pesquisas relacionadas à previsão de demanda no mercado imobiliário, pode-se concluir que são praticamente inexistentes, Varandas Júnior e Alencar (2009) objetivaram testar a aplicabilidade dos modelos de previsão



de demanda de DW (1992) e também o modelo DW melhorado por Cowel (2002) em economias em desenvolvimento. Concluindo que a aplicação destes modelos exige uma série histórica de dados não disponível no Brasil, métodos como os dois abordados tornam-se impraticável no mercado imobiliário nacional. Adicional a isso, o estudo torna-se relevante para a interpretação dos movimentos do setor e para o planejamento dos agentes envolvidos neste mercado.

Outros estudos como o de Meyer (2008), analisa o mercado imobiliário residencial de São Paulo, mensurando a demanda de imóveis disponíveis, cruzando com a oferta de demanda por categoria de rendas das famílias, já o trabalho de Varandas Júnior (2010) aborda uma revisão bibliográfica sobre o tema demanda habitacional e suas interfaces com o ambiente econômico, a partir da análise da produção habitacional privada na cidade de São Paulo de 1998 até 2008 frente às variáveis condicionantes selecionadas: formação de domicílios, renda, oferta de recursos, custos de financiamento, preço e taxa de atratividades sendo interpretadas as fontes de distorção que provocaram o deslocamento entre a formação de domicílios e a produção habitacional no período; também foram importantes pesquisas na área de demanda imobiliária, porém nenhum propõe um modelo de predição da demanda.

O Capítulo 2 apresentou a fundamentação teórica pesquisada e desenvolvida para a realização da presente pesquisa, advindo do mercado imobiliário, apresentando métodos de previsão de demanda para séries temporais e encerrando com a exposição de pesquisas relacionadas à previsão de demanda no mercado imobiliário.



### 3 METODOLOGIA

No Capítulo 3 é apresentada a metodologia de pesquisa a qual será utilizada para a condução e realização do estudo, sendo este capítulo dividido em quatro seções: a Seção 3.1 apresenta a classificação da pesquisa, a Seção 3.2 explana sobre a obtenção dos dados; a Seção 3.3 descreve sobre as etapas da pesquisa necessárias para cumprir o objetivo proposto, por fim a Seção 3.4 aborda as delimitações e limitações da pesquisa.

#### 3.1 CLASSIFICAÇÃO DA PESQUISA

Na concepção de Gil (2009) uma pesquisa científica pode ser classificada de acordo com sua: (a) natureza; (b) objetivos; (c) abordagem; e (d) procedimentos técnicos. Dessa forma, utilizou-se a classificação proposta pelo autor para classificar essa pesquisa.

Com relação a sua natureza esse estudo é classificado como aplicado, pois essa pesquisa objetiva gerar conhecimentos para aplicação prática e é conduzida a solução de um problema específico, a previsão de demanda no mercado imobiliário residencial da capital paulista.

Em relação aos objetivos, essa pesquisa é classificada como exploratória e explicativa, pois visa proporcionar maior intimidade com o problema a fim de torná-lo mais explícito e possibilitar a construção de resultados satisfatórios, bem como se utiliza de modelagem matemática para analisar as causas dos fatos do mercado imobiliário de São Paulo.

De acordo com sua abordagem essa pesquisa é classificada como quantitativa, visto que este trabalho tem como base a modelagem computacional que utiliza dados numéricos e coletados de um problema real (vendas de imóveis residenciais no mercado imobiliário paulistano) e assim para modelagem dos dados é utilizado softwares adequados, por intermédio de modelos estatísticos gerando resultados numéricos os quais auxiliam profissionais e organizações da área na tomada de decisão gerencial.

Por fim, os procedimentos técnicos dessa pesquisa podem ser classificados como pesquisa bibliográfica e experimental. Para se familiarizar e apresentar o assunto de previsão de demanda no mercado imobiliário utilizou-se de pesquisa bibliográfica em livros, artigos de periódicos, dissertações e teses e assim utilizar-se-á da pesquisa experimental para testar os métodos de previsão de demanda, bem como combina-los e posteriormente compara-los a partir de medidas de acurácia.

### 3.2 OBTENÇÃO DE DADOS

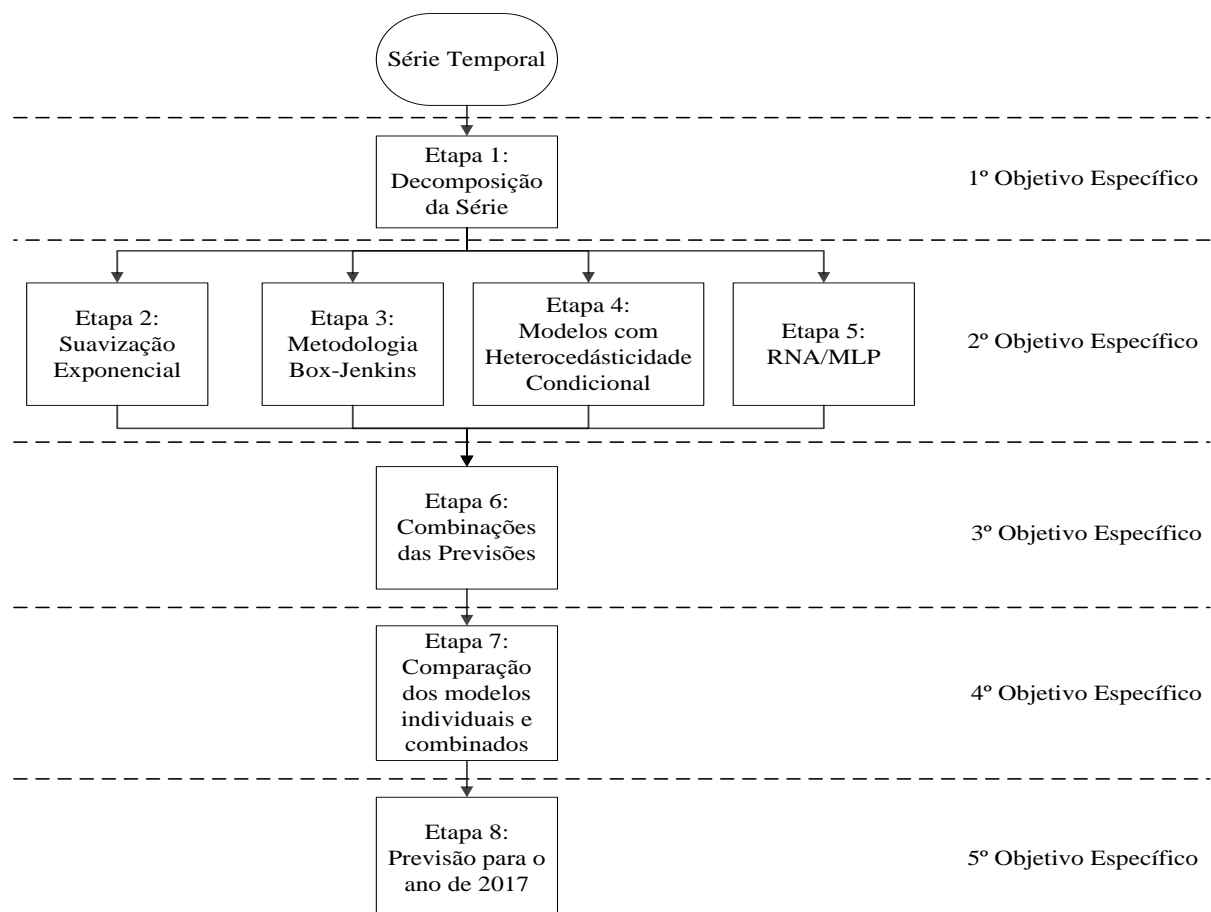
Para uma previsão de demanda quantitativa é necessário a obtenção de dados, que no caso dessa pesquisa, seria a série temporal de unidades vendidas no mercado imobiliário da cidade de São Paulo.

Os dados que são utilizados nessa pesquisa estão disponíveis na página da internet da Câmara Brasileira da Indústria da Construção (CBIC), obtido por meio de estudos do SECOVI-SP e EMBRAESP, por serem dados previamente tabelados são considerados dados secundários. Os dados compreendem o período de janeiro de 2004 até junho de 2016, correspondendo a 150 observações.

### 3.3 ETAPAS DA PESQUISA

Para atingir os objetivos específicos e, por conseguinte o geral dessa pesquisa, esta foi dividida em oito etapas, as quais podem ser observadas na Figura 15.

Figura 15 - Etapas para o desenvolvimento da pesquisa



Fonte: Elaborado pelo autor.

As seções subsequentes descrevem os passos para a modelagem, previsão e comparação dos resultados, conforme demonstrado na Figura 16.

### 3.3.1 Etapa 1 – Características da série

A primeira etapa da pesquisa refere-se ao primeiro objetivo específico e almeja encontrar as características da série. Antes disso, a série composta de 150 observações foi dividida em duas, com proporção de 75% (112 observações) para modelagem/ajuste dos modelos e outros 25% (38 observações) para medir a acurácia das previsões geradas pelos modelos.

Para realizar a decomposição da série utilizou-se o método de Decomposição Clássica Aditivo, o qual considera que os elementos são somados. A decomposição é composta pela Tendência, série sem a Tendência, Sazonalidade e Índices Sazonais. O software utilizado para essa etapa é o NNQ-Stat.

Para confirmar as evidências encontradas na análise visual dos gráficos gerados pela decomposição, para testar a tendência realiza-se os testes de Wald-Wolfowitz, de Cox-Stuart e o de Mann-Kendall, sob as hipóteses  $H_0$ : Não há tendência e  $H_1$ : Há tendência. Para testar a presença de sazonalidade utilizam-se os testes Kruskal-Wallis e de Friedman, sob as hipóteses  $H_0$ : Não há sazonalidade e  $H_1$ : Há sazonalidade. Salienta que para todos os testes o nível de confiança é de 95%, sendo realizados no software *Action Stat*.

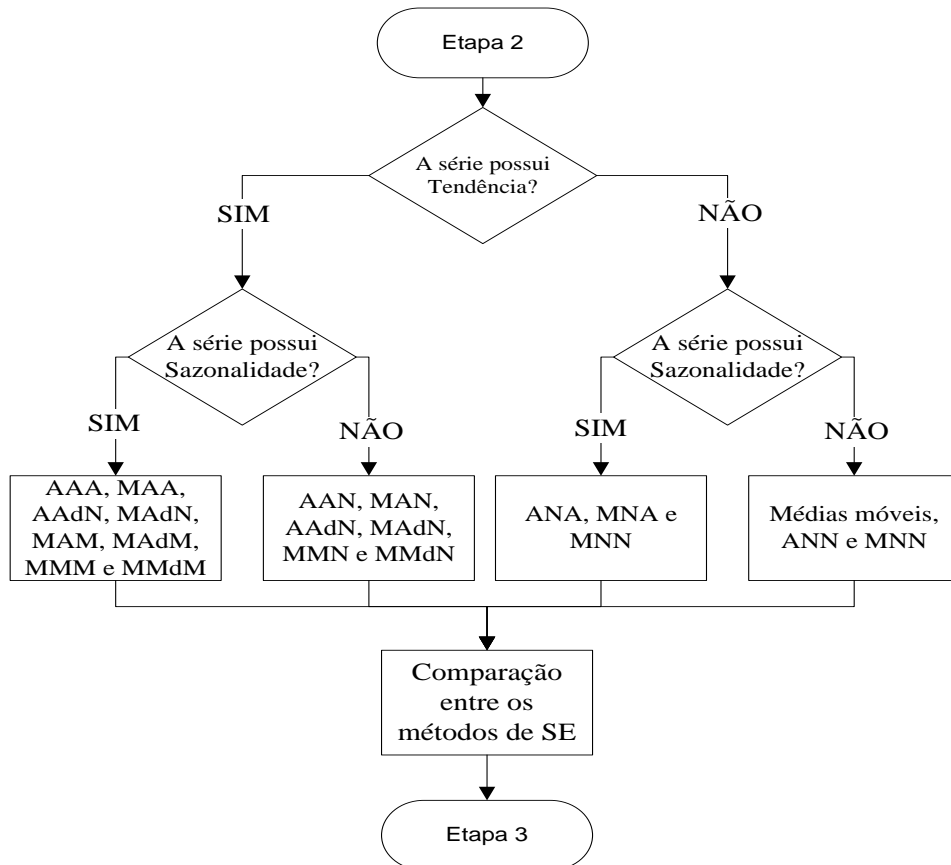
### 3.3.2 Etapa 2 – Suavização Exponencial

A segunda etapa da pesquisa é a modelagem dos métodos de Suavização Exponencial e está associada ao segundo objetivo específico. Para a modelagem utilizou-se o software NNQ-Stat. A metodologia utilizada nessa etapa está apresentada na Figura 16.

Esta etapa é derivada da etapa anterior, visto que é necessário conhecer as características da série para então saber quais são os métodos de Suavização Exponencial adequados para modelagem da série. Se a série não apresentar tendência nem sazonalidade, Médias Móveis, ANN e MNN podem ser testados. Caso a série apresente tendência, mas não sazonalidade, os métodos testados são AAN, MAN, AAdN, MAdN, MMN e MMdN. Se a série não apresentar tendência, mas possuir elementos sazonais, os métodos apropriados são ANA, MNA e MNN. Por fim, se a série tiver elementos sazonais e possui tendência, os métodos indicados são AAA, MAA, AAdN, MAdN, MAM, MAdM, MMM e MMdM. Após

os ajustes e previsão dos modelos apropriados os métodos testados são comparados por meio do MAE, MAPE, RMSE e UTHEIL, essas medidas são calculadas tanto para as séries ajustados quanto para a previsão.

Figura 16 - Metodologia utilizada na Etapa 2 de modelagem dos métodos de Suavização Exponencial



Fonte: Elaborado pelo autor.

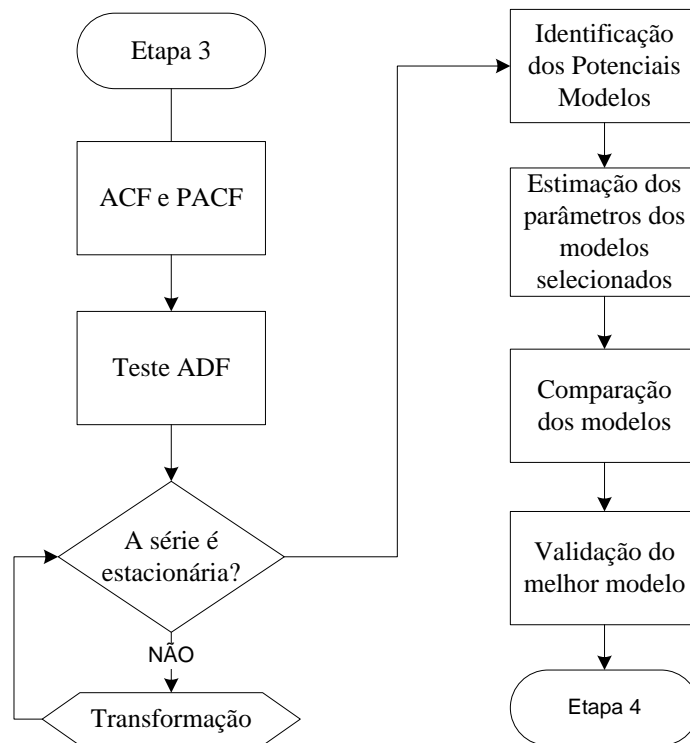
Salienta-se que o software NNQ-Stat não modela os métodos com: (i) Sazonalidade Aditiva com tendência multiplicativa tanto para erro multiplicativo quanto aditivo; (ii) Sazonalidade Aditiva com tendência amortecida tanto para erros multiplicativo quanto aditivo; e (iii) Sazonalidade multiplicativa com correção de erros aditivos. Sendo assim dos 30 métodos apresentados no Quadro 2, somente são modelados 19.

Ressalta-se que todas as previsões realizadas nesse trabalho são *out-of-sample*, ou seja, os 25% da série não entram em nenhum modelo, somente servem para medir a qualidade da previsão.

### 3.3.3 Etapa 3 – Metodologia de Box-Jenkins

A etapa 3 da pesquisa também está atrelada ao segundo objetivo específico e consiste na aplicação da metodologia de Box-Jenkins na obtenção do modelo ARIMA que melhor se adequa e que gere boas previsões para a série temporal em estudo. A modelagem do processo ARIMA será baseada na sequência de execução da metodologia de Box-Jenkins (Figura 8), a qual foi incluída no estudo na forma descrita na Figura 17, a qual mostra os passos dados na obtenção do modelo ARIMA para previsão da série temporal em estudo. Para modelagem desta etapa utiliza-se o software *Gretl 1.9.5 cvs*

Figura 17 - Fluxograma da metodologia para a obtenção do modelo ARIMA para a previsão da série temporal



Fonte: Elaborado pelo autor.

O primeiro passo foi realizar uma análise prévia do comportamento da ACF e da PACF, para encontrar evidências de estacionariedade da série. Em seguida, aplicou-se o teste estatístico de raiz unitária Augmented Dickley-Fuller (ADF) com as seguintes hipóteses:  $H_0$ : a série possui raiz unitária; e  $H_1$ : a série não apresenta raiz unitária (estacionária), para

confirmar a hipótese inicialmente estabelecida acerca da estacionariedade da série. O teste foi aplicado ao nível de significância de 1%, 5% e 10%.

Após, foram identificados, por meio da ACF e da PACF e das diretrizes descritas na seção 2.2.1.2.7, os potenciais modelos ARIMA para a série. Dessa forma por meio dos critérios penalizadores AIC e BIC foram selecionados os melhores modelos, os quais apresentam menores valores dos critérios penalizadores, para estimação dos parâmetros, por fim é realizado a comparação dos modelos selecionados por meio dos critérios MAE, MAPE, RMSE e UTHEIL, cálculos são realizados para avaliação do ajuste e previsão, assim é escolhido o modelo que gera a melhor previsão.

Por fim, ocorre a validação do melhor modelo escolhido. A validação foi realizada por meio da obtenção e análise do gráfico dos resíduos padronizados e pela análise da ACF e PACF dos resíduos obtidos pelo modelo. Se o modelo obteve resíduos com média zero, variância constante e ausência de autocorrelação, este foi validado para ser utilizado na previsão dos valores futuros da série temporal.

### **3.3.4 Etapa 4 – Modelos Heterocedásticos com Variância Condicional**

A etapa 4 corresponde a modelagem dos métodos não-lineares de heterocedasticidade com variância condicional e está também atrelado ao segundo objetivo específico. Para modelagem desta etapa utilizou-se o software Gretl.

Em um primeiro momento realiza-se o teste de presença de heterocedasticidade condicional, sendo a hipótese nula de ausência de heterocedasticidade condicional, com a finalidade de testar se a variância dos resíduos é constante. Caso rejeita-se a hipótese nula modela-se a variância, utilizando os modelos de Heterocedasticidade. O teste utilizado para detectar a heterocedasticidade será o teste ARCH-LM e será aplicado nos resíduos do melhor modelo encontrado na etapa 3.

Tendo verificado a presença de heterocedasticidade, modela-se a variância testando os parâmetros no software e verificando a significância.

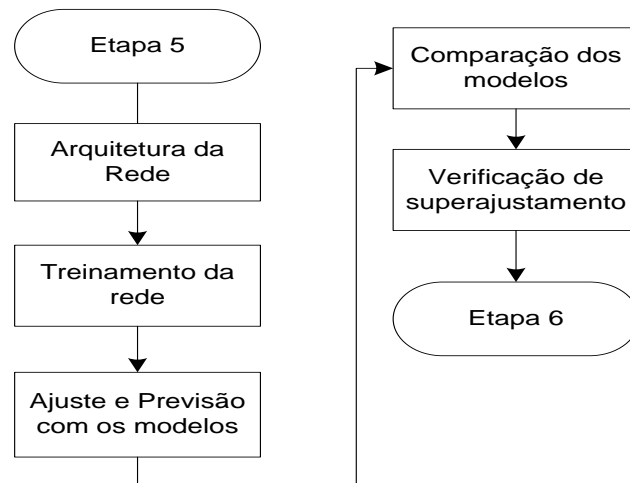
### **3.3.5 Etapa 5 – RNA/MLP**

A etapa 5 também está atrelada ao segundo objetivo específico e objetiva realizar a modelagem e a previsão dos valores da série por meio do modelo MLP/RNA. Na Figura 18 é apresentada a metodologia considerada para a obtenção do modelo MLP. Para modelagem de



Redes Neurais foi utilizado apenas a topologia de *Multilayer Perceptron* (MLP) com uma camada escondida, também conhecida como camada intermediária. Para treinamento e modelagem das RNAs utilizou-se o software *Statistica 10.0*.

Figura 18 - Fluxograma da metodologia utilizada para a obtenção do modelo MLP



Fonte: Elaborado pelo autor.

O primeiro passo nessa etapa é montar a arquitetura da rede, ou seja, quantos neurônios cada camada irá possuir. O número de neurônios na camada de entrada, foi definido com base na característica sazonal observada na série temporal estudada, já para a camada escondida serão testados modelos com 5, 10, 15, 20, 25 e 30 neurônios, salienta-se que por tratar-se de um problema de séries temporais univariado, o número de neurônios na camada de saída será 1.

Depois de decidido a arquitetura da rede, parte-se para o treinamento da rede, nessa etapa utilizou-se dois algoritmos em duas fases. A primeira fase do treinamento foi realizada pelo algoritmo *Backpropagation* e a segunda fase foi realizada pelo algoritmo *Conjugate Gradiente Descent*. Para todos os modelos de RNAs utilizados neste estudo, foi estipulado o número de 20 mil ciclos de treinamento, tendo como objetivo evitar o superajustamento aos dados de treinamento. Após o treinamento modelou-se a série temporal conforme os modelos que serão ajustados, dependendo do número de neurônios em cada camada, sendo também realizada a previsão 38 passos a frente para cada modelo.

Depois de realizada a previsão faz-se a comparação entre todos os modelos, utilizando as métricas RMSE, MAPE, MAE e UTHEIL. Sendo escolhido o melhor modelo, aquele com

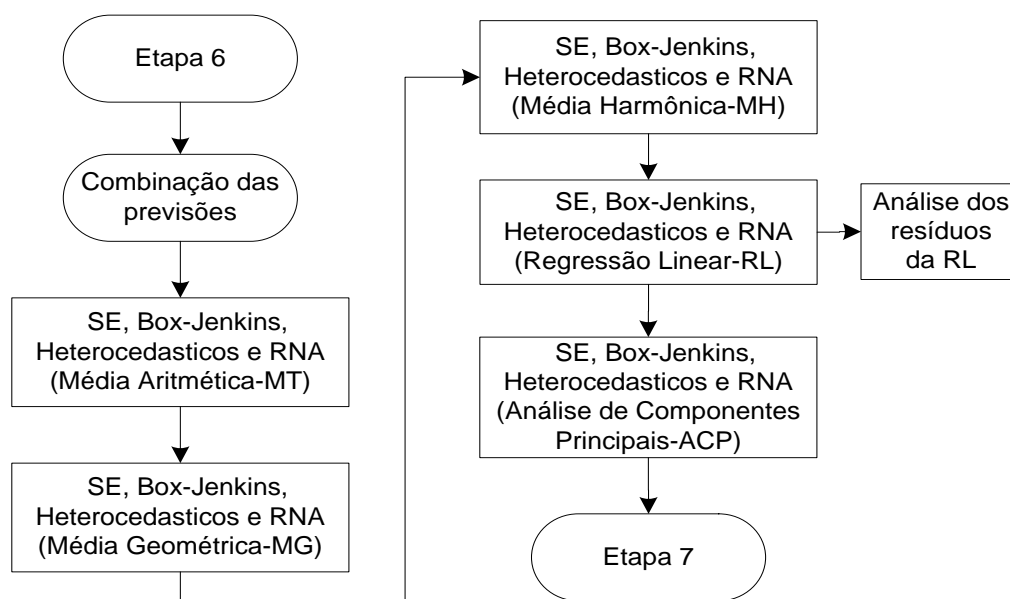
menores valores das estatísticas de acurácia, é necessário verificar se não houve superajustamento. Para a verificação de superajustamento das RNAs aos dados da série, a análise consiste na comparação da variação das medidas de erro do processo de treinamento e de validação das RNAs. Caso constatada divergências acentuadas entre tais medidas, há fortes suposições de que determinada RNA teria memorizado os dados fornecidos para o seu treinamento, perdendo a capacidade de generalização, situação esta que não gera resultados práticos a presente pesquisa.

Salienta-se que para simplificar, adotou-se a nomenclatura MLP (3,10,1), onde 3 é o número de neurônios na camada de entrada que contém os valores defasados referentes a 3 períodos, 10 é o número de neurônios da camada escondida e 1 é o número de neurônios na camada de saída.

### 3.3.6 Etapa 6 – Combinações das previsões

Na etapa 6 são realizadas as combinações das previsões, sendo que esta atrelada ao terceiro objetivo específico. Na Figura 19 são apresentadas as fases para obtenção das combinações das previsões.

Figura 19 - Metodologia para a combinação das previsões



Fonte: Elaborado pelo autor.

O primeiro método de combinação realizado é por meio da Média Aritmética, demonstrado na Equação 64, posteriormente utiliza-se a Equação 65 para combinar as previsões pelo método da Média Geométrica. O método da Média Harmônica, explanado na Equação 66 é utilizado para realizar a combinação dos quatro modelos de previsão individuais. Esses três métodos de combinação são modelados no software Excel.

O quarto método de combinação a ser utilizado nesse estudo é o de Regressão Linear. Assim a variável dependente será a série de Vendas utilizada para validação do modelo e as variáveis independentes serão as previsões geradas pelas quatro classes de modelos: SE, Box-Jenkins, Heterocedasticidade Condicional e RNA. Essa combinação será realizada pelo software *Action Stat*.

Por fim é preciso validar o modelo por meio da análise de resíduos da regressão. Para que o modelo de regressão seja válido é necessário que alguns pressupostos sejam atendidos, visando garantir avaliações não tendenciosas, eficientes e consistentes. Os pressupostos a serem atendidos são (a) os erros tem média zero; (b) os resíduos são homocedásticos (Teste de Breusch Pagan); (c) Os resíduos são independentes (Teste de Durbin-Watson); (d) os erros são normalmente distribuídos (Testes de Kolgomorov-Sminorv, Lilliefors e Shapiro-Wilks); (e) não há problemas de multicolinearidade (VIF – *Variance Influencion Factore*); e (f) não há *outliers* (Teste de Bonferroni). Salienta-se que para todos os testes dessa pesquisa adotou-se 5% para o nível de significância. Por fim, o método de ACP é utilizado para combinação, adotando o fator 1 como combinação linear das variáveis.

Após os modelos combinados, verificar-se-á se é possível, diante dos resultados, a correção dos modelos combinados, bem como verificar a possibilidade de utilizar os resultados de modelos combinados como dados de entrada em outros métodos, os quais os resultados possam gerar previsões acuradas. Desse modo, propondo modelos híbridos e comparando com o melhor modelo combinado e modelos individuais.

### **3.3.7 Etapa 7 – Comparação dos modelos individuais e combinados**

Na penúltima etapa deste estudo, foi realizada uma comparação entre os modelos individuais utilizados para a previsão *versus* suas respectivas combinações e entre as combinações das previsões, com o objetivo de verificar qual das combinações resultou em um melhoramento significativo na capacidade preditiva.

Após, analisou-se o desempenho obtido com as combinações das previsões em relação aos modelos utilizados. Para isso, utilizaram-se as Equações 69, 70, 71 e 72 das medidas de

acurácia MAPE, MAE, RMSE e UTHEIL, respectivamente, desse modo mensurando os resultados obtidos na etapa de modelagem (ajuste) e na etapa de previsão para os modelos individuais, os modelos combinados e o modelo híbrido proposto. Todas as métricas são calculadas no software Excel.

### **3.3.8 Etapa 8 – Previsão**

Após todas as etapas, sendo encontrado o método mais adequado para prever a demanda do mercado imobiliário de São Paulo, este será utilizado para prever a demanda do ano de 2017. Mais do que isso, por meio da previsão do ano de 2017 é realizada uma discussão acerca dos resultados e das tendências do mercado apresentadas pelos economistas, corretores e incorporadoras por meio de reportagens publicadas em alguns meios de comunicação, assim será verificado se o resultado apresentado pelo melhor modelo vai ao encontro destes ou não.

O software para previsão irá depender de qual será o método de melhor acurácia, sendo utilizado o software, o qual o método foi modelado.

## **3.4 DELIMITAÇÕES E LIMITAÇÕES DA PESQUISA**

Na previsão são comumente encontrados dois tipos de métodos, os qualitativos e os quantitativos. Dentro dos métodos quantitativos os mais comuns são aqueles que utilizam séries temporais. Há dois enfoques utilizados na análise de séries temporais, em ambos o objetivo é construir modelos para essas séries. No primeiro enfoque, a análise é feita no domínio temporal e os modelos são modelos paramétricos. Já no segundo enfoque, têm-se os modelos não paramétricos, caracterizados pela análise conduzida no domínio da frequência. Deste modo, o estudo delimita-se em utilizar métodos quantitativos, os quais utilizam séries temporais e são classificados em paramétricos. Métodos qualitativos e modelos não paramétricos não são abordados nessa pesquisa.

Apesar da ascensão do mercado imobiliário na última década, este mercado encontra-se ainda em fase de amadurecimento. A falta de informações resulta na escassez de dados, isso também repercute na dificuldade da obtenção de literaturas acadêmicas sobre o tema. Dentro disso, o mercado imobiliário com o maior número de dados é a capital paulista, São Paulo. Deste modo, o estudo limitar-se-á a essa cidade, portanto os resultados auferidos nessa pesquisa servirão apenas para o seu respectivo mercado. Outra limitação vem do fato dos

dados serem secundários, advindos de coleta do SECOVI-SP, onde o autor não pode garantir a inteira confiabilidade, visto que as informações já foram previamente tabeladas. Por fim, os dados disponíveis compreendem o período a partir do ano de 2004 até junho de 2016, mudanças no mercado antes desse período não poderão ser analisadas.



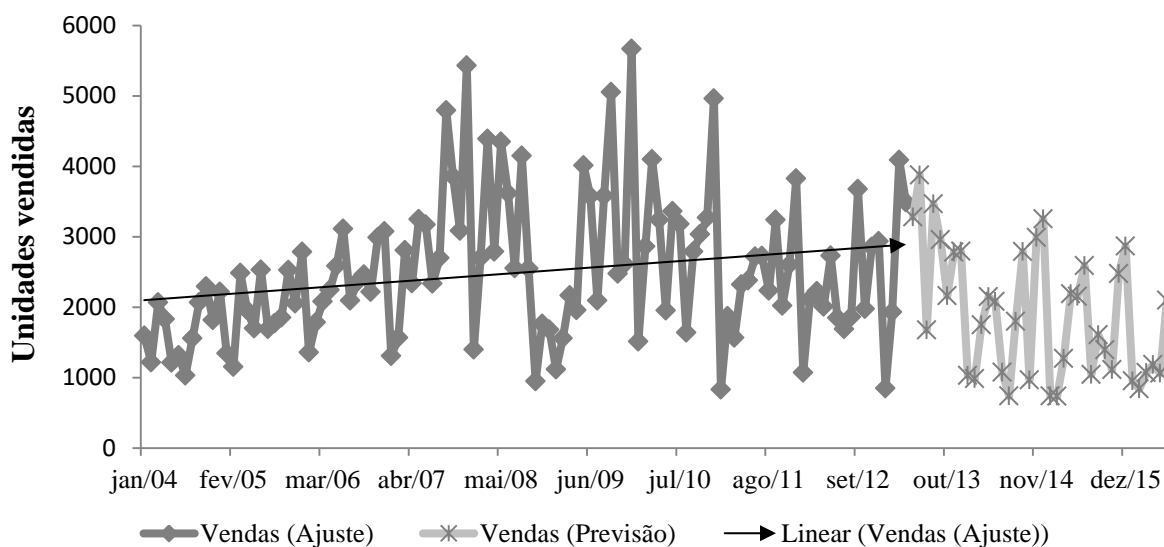
## 4 RESULTADOS E DISCUSSÃO

Neste capítulo são apresentadas as modelagens das séries temporais utilizadas no presente estudo, conforme os modelos descritos no Capítulo 2. Em um primeiro momento, na Seção 4.1 apresenta-se a decomposição da série para diagnóstico das características da mesma. Na sequência, seção 4.2 apresentam-se as modelagens da série temporal de vendas de imóveis residências na cidade de São Paulo. A Seção 4.3 restringe-se à comparação dos melhores modelos. A Seção 4.4 apresenta a combinação de modelos de previsão, propondo um modelo híbrido para previsão da demanda de imóveis na cidade de São Paulo. Por fim, a Seção 4.5 apresenta uma previsão para os próximos meses e perpetra uma discussão sobre as premissas do mercado em 2017.

### 4.1 CARACTERÍSTICAS DA SÉRIE TEMPORAL

A primeira etapa da pesquisa inicia-se com o intuito de compreender o comportamento da série de vendas de imóveis na capital paulista, destarte apresenta-se na Figura 20 o gráfico da série em estudo juntamente com uma linha de tendência, salienta-se que somente 75% da série será utilizada para ajustamento dos modelos de previsão. Evidencia-se que não se encontrou alguma observação espúria nos dados em estudo.

Figura 20 - Vendas de imóveis em São Paulo entre Jan/04 e Jun/2016

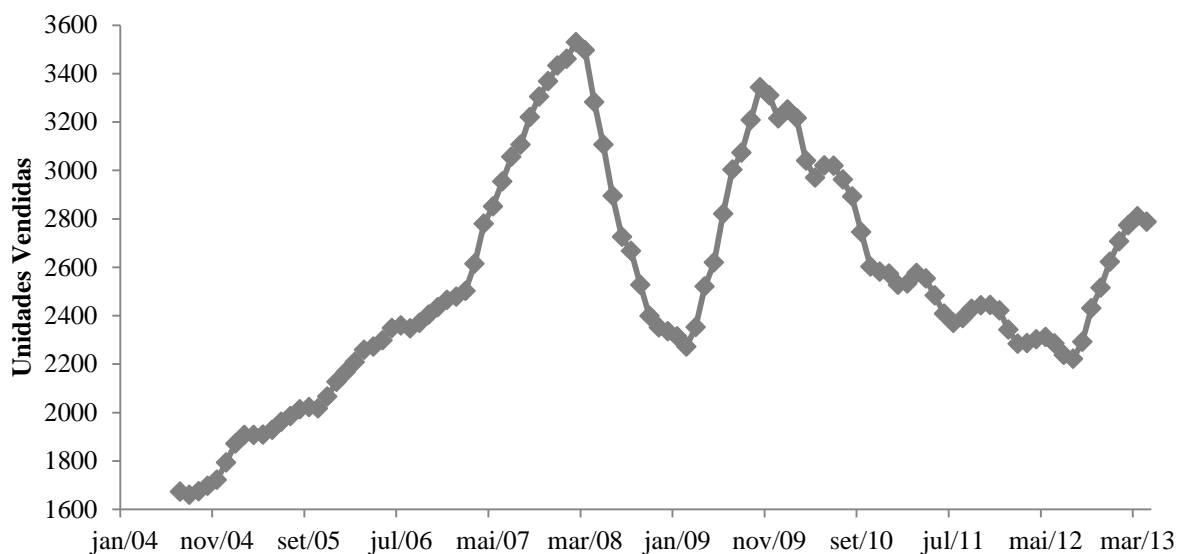


Fonte: Elaborado pelo autor.

Percebe-se por meio da linha de tendência apresentada na Figura 20, que há evidências de que a série tem uma tendência positiva. Conforme se observa na mesma figura, a série temporal apresenta padrões de comportamento diferenciados em diversos períodos da série, especialmente no início (menor variância) em relação ao seu final (maior variância). Desse modo, procurou-se analisar para afirmar a presença de características como tendência, sazonalidade e padrão cíclico. Desse modo, obteve-se a decomposição da série temporal, utilizando o método de Decomposição Clássica por meio do software NNQ/UFSC.

Na Figura 21, tem-se o gráfico da componente tendência. Nota-se que até a observação 50, correspondente ao mês de fevereiro de 2008, há uma tendência positiva, esta pode ser explicada pela nova política de habitação do primeiro governo do presidente Lula, a qual visou ampliar maciçamente os recursos de fundos públicos e semi públicos para os financiamentos habitacionais.

Figura 21 - Decomposição da componente tendência da série temporal de vendas de unidades de imóveis residências na cidade de São Paulo



Fonte: Elaborado pelo autor.

Após esse período, verifica-se uma tendência negativa até fevereiro de 2009, período este de instabilidade da economia mundial diante da crise imobiliária nos EUA. O mercado voltou a apresentar uma tendência positiva até outubro de 2009, após esse momento houve um grande período de tendência negativa, até julho de 2012. Sendo este, um período de ajuste de mercado, diante dos inúmeros lançamentos no mercado nos anos anteriores. Por fim, há evidências de que há uma tendência positiva a partir de julho de 2012, fato o qual, pode ser



explicado pelo aumento do limite de utilização do FGTS para financiamento para aquisição de imóveis.

Para confirmar a presença da componente tendência na série, apresentam-se na Tabela 2 os resultados dos testes de tendência Wald-Wolfwitz, Mann-Kendall e Cox-Stuart, os três possuem como hipóteses  $H_0$ : Não há tendência e  $H_1$ : Há tendência.

Tabela 2 - Testes para detecção de presença de tendência na série temporal

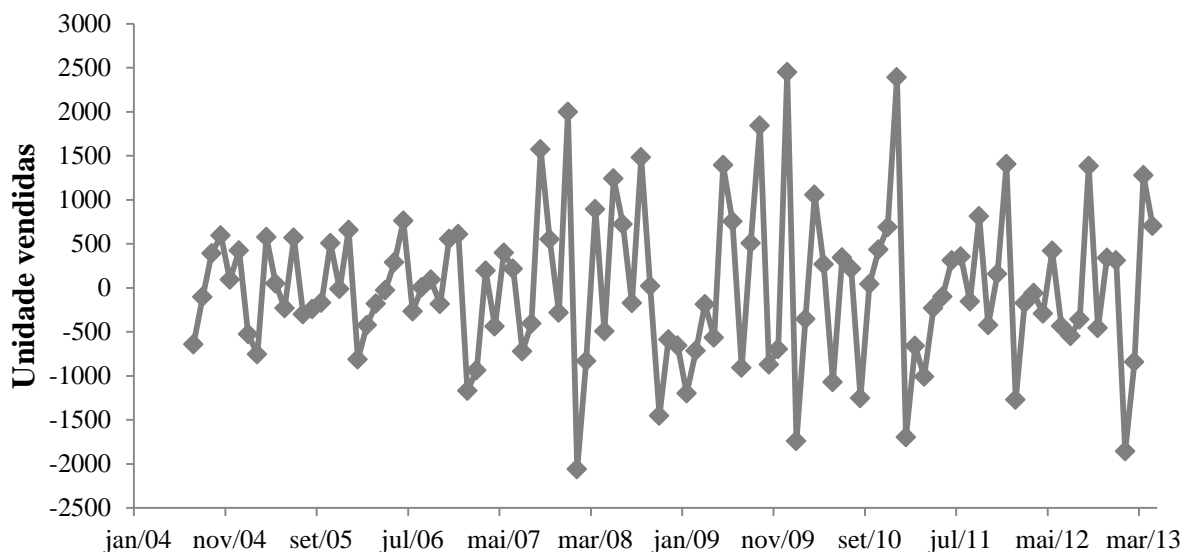
	Wald-Wolfwitz	Mann-Kendall	Cox-Stuart
<b>Estatística</b>	42	0,187258691	29
<b>p-valor</b>	0,00440517	0,003451943	0,0446927

Fonte: Saída do software *Action Stat* adaptada pelo autor.

Observando os resultados dos testes, para os três rejeita-se a hipótese nula, visto que o p-valor é menor que o nível de significância (5%), afirma-se então que a série em estudo apresenta a característica tendência.

Após a decomposição da componente tendência, na Figura 22 apresenta-se o gráfico da série decomposta sem a tendência, verifica-se que a série apresenta vários clusters, tendo evidências de sazonalidade, com uma menor variância até a observação 40 e após uma variância maior, devido a uma maior instabilidade no mercado.

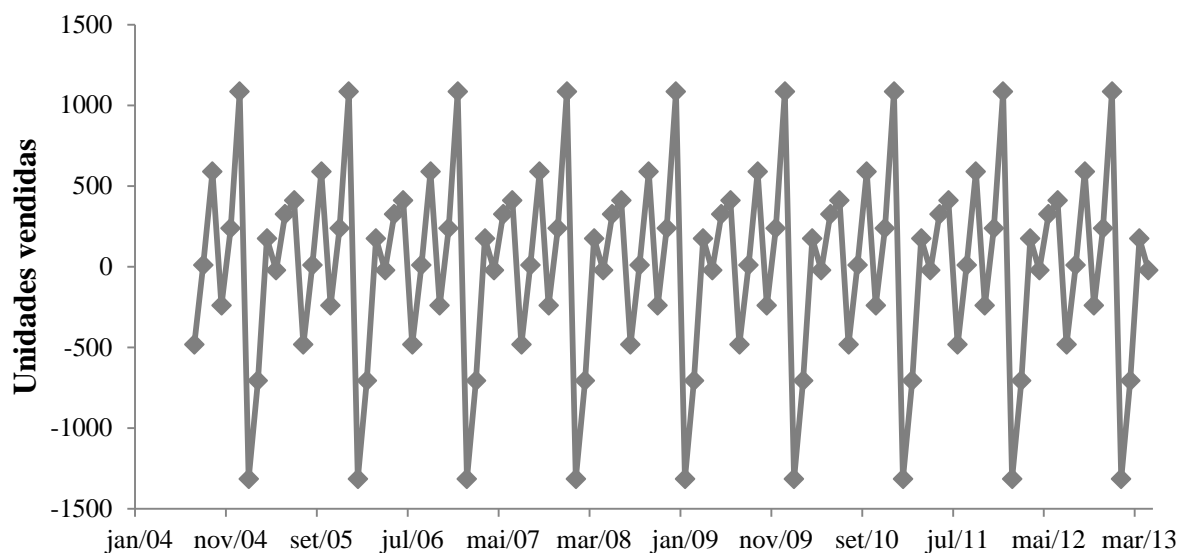
Figura 22 - Série temporal de vendas de unidades de imóveis residências na cidade de São Paulo decomposta sem a tendência



Fonte: Saída do software *NNQ Stat* adaptada pelo autor.

Tendo indícios de sazonalidade, na Figura 23 apresenta-se o gráfico da componente sazonal da série de vendas de imóveis residências em São Paulo. Salienta-se que sazonalidade são variações que se repetem todos os anos.

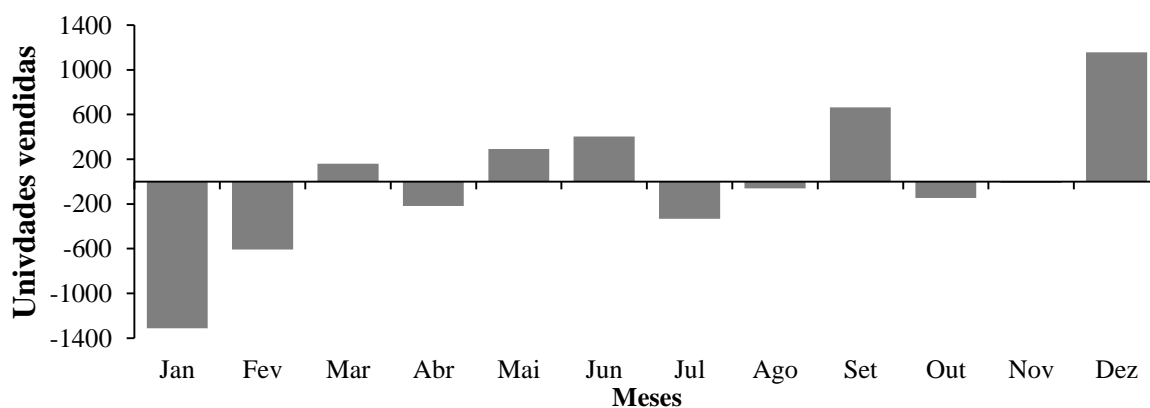
Figura 23 - Decomposição da componente sazonalidade da série temporal de vendas de unidades de imóveis residências na cidade de São Paulo



Fonte: Saída do software NNQ Stat adaptada pelo autor.

Esse fato pode ser confirmado, por meio da Figura 24, onde é apresentado o gráfico com os índices sazonais. Afirma-se assim que o mês de dezembro é o mês com maior volume de vendas de imóveis e o mês de janeiro é o mês com menor número de vendas.

Figura 24 - Índices sazonais da série temporal de vendas de unidades de imóveis residências na cidade de São Paulo



Fonte: Saída do software NNQ Stat adaptada pelo autor.

Corroborando com a análise gráfica, para confirmar a evidência de que a série apresenta componente sazonal, na Tabela 3 apresentam-se os resultados dos testes de Kruskal-Wallis e Friedman, os quais possuem como hipóteses,  $H_0$ : Não há sazonalidade e  $H_1$ : Há sazonalidade. Diante da Figura 24 (índice sazonais) resolveu-se testar tamanhos de sazonalidade de 3, 6 e 12.

Tabela 3 - Testes para detecção de presença de sazonalidade na série temporal

Testes	Sazonalidade = 3		Sazonalidade = 6		Sazonalidade = 12	
	<i>Kruskal-Wallis</i>	<i>Friedman</i>	<i>Kruskal-Wallis</i>	<i>Friedman</i>	<i>Kruskal-Wallis</i>	<i>Friedman</i>
Estatística	54,80741939	37,135135	46,58817533	45,58730	48,65432099	50,111111
p-valor	1,25522E-12	8,634E-09	6,89276E-09	1,102E-08	1,09094E-06	5,978E-07

Fonte: Saída do software *Action Stat* adaptada pelo autor.

Observando os resultados dos testes, verifica-se que para os tamanhos de sazonalidade 3, 6 e 12 rejeita-se a hipótese nula, assim afirma-se com 95% de confiança que a série apresenta componente sazonal de tamanhos 3, 6 e 12. Esse resultado servirá de base para modelar a série nas etapas de Box-Jenkins e RNA.

Findando, comenta-se que não se verificou alguma evidência que a série de vendas em São Paulo possui padrão cíclico, este fato, porém não significa que as vendas de imóveis na capital paulista não apresenta um ciclo, mas sim que pelo número de observações que a pesquisa possui não é possível impetrar um padrão cíclico.

## 4.2 MODELAGENS DA SÉRIE TEMPORAL

Esta seção refere-se a modelagens dos métodos de previsão, compreendendo as etapas 2, 3, 4 e 5 da pesquisa, apresentada na metodologia. A apresentação da etapa de modelagens dispõe da seguinte sequência, modelagem das séries temporais utilizando os modelos de: (i) SE; (ii) Box & Jenkins; (iii) Modelos Heterocedásticos; e (iv) RNAs.

### 4.2.1 Suavização Exponencial

A segunda etapa da pesquisa compreende a modelagem da série por meio dos modelos de Suavização Exponencial. Na Seção 4.1 deste capítulo a série temporal em estudo foi

decomposta e encontraram-se elementos que evidenciam a presença de tendência e sazonalidade na série. Desta maneira, seguindo a metodologia da pesquisa, os métodos de SE testados serão os apropriados para séries que possuem tendência e sazonalidade. O processo de modelagem foi realizado utilizando o software *NNQ-Stat*. As medidas de acurácia foram calculadas no software *Excel*.

Como ressaltado, a série de vendas de imóveis residências na cidade de São Paulo possui características de tendência e sazonalidade, em vista disso modelaram-se apenas os oito métodos apropriados para séries com essas duas características. Na Tabela 4 apresentam-se os parâmetros e as medidas de acurácia calculadas para cada um dos oito métodos de SE com tendência e sazonalidade.

Tabela 4 - Parâmetros e medidas de erro para os métodos de SE com tendência e sazonalidade

	AAA	MAA	AAdA	MAdA	MAM	MAdM	MMM	MMdM
<b>Parâmetros</b>	$\alpha = 0,32$	$\alpha = 0,08$	$\alpha = 0,32$	$\alpha = 0,08$	$\alpha = 0,30$	$\alpha = 0,30$	$\alpha = 0,31$	$\alpha = 0,29$
	$\beta = 0,01$	$\beta = 0,01$	$\beta = 0,01$	$\beta = 0,01$	$\beta = 0,01$	$\beta = 0,01$	$\beta = 0,01$	$\beta = 0,01$
	$\gamma = 0,01$	$\gamma = 0,25$	$\gamma = 0,01$	$\gamma = 0,23$	$\gamma = 0,01$	$\gamma = 0,01$	$\gamma = 0,01$	$\gamma = 0,01$
			$\varphi = 0,80$	$\varphi = 0,98$	$\varphi = 0,01$	$\varphi = 0,80$	$\varphi = 0,01$	$\varphi = 0,80$
<b>RMSEa</b>	552,673	614,99	552,65	<b>705,621</b>	512,693	512,836	513,086	506,75*
<b>MAPEa</b>	0,259	0,284	0,257	<b>0,254</b>	0,233	0,232	0,236	0,227*
<b>MAEa</b>	724,739	827,78	722,54	<b>554,946*</b>	694,327	691,993	694,807	688,645
<b>UTHEILa</b>	0,639	0,787	0,637	<b>0,557614*</b>	0,623	0,620	0,621	0,613
<b>RMSEp</b>	1432,685	991,858	1260,395	<b>899,513*</b>	1366,361	1217,171	1707,043	1201,246
<b>MAPEp</b>	0,991	0,629	0,868	<b>0,589*</b>	0,897	0,802	1,112	0,791
<b>MAEp</b>	1276,482	828,394	1114,434	<b>775,576*</b>	1197,481	1062,421	1506,324	1047,708
<b>UTHEILp</b>	1,221	0,845	1,074	<b>0,767*</b>	1,165	1,037	1,455	1,024

\*Melhor modelo para cada medida de acurácia.

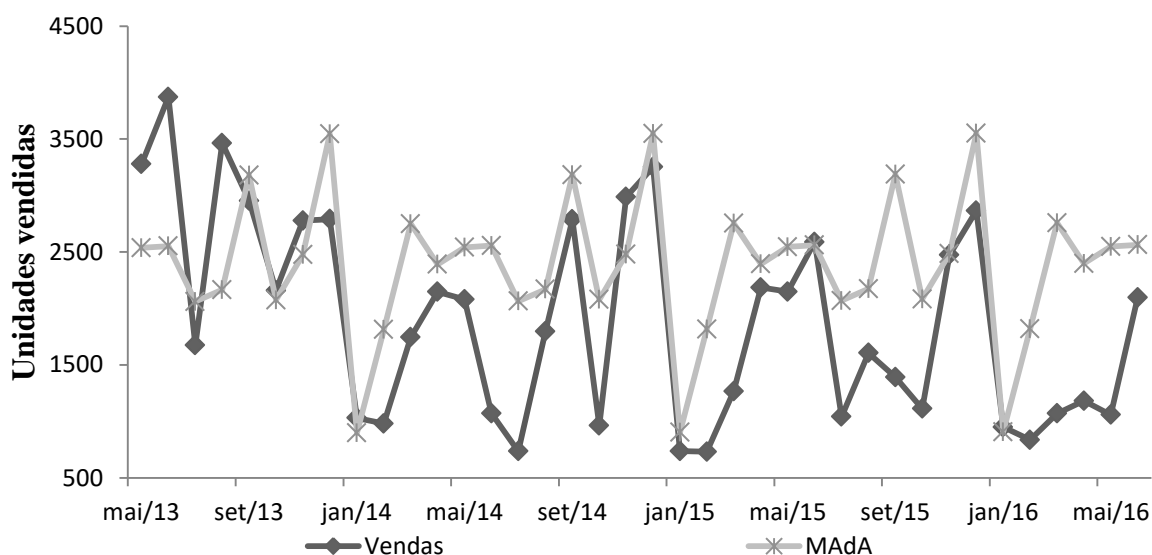
Fonte: Elaborado pelo autor.

O modelo de Tendência Multiplicativa Amortecida com Sazonalidade Multiplicativa e correção Multiplicativa do Erro (MMdM) obteve os menores valores para as medidas RMSE e MAPE de ajustamento. Entretanto para as medidas MAE e UTheil o modelo de Tendência Aditiva Amortecida com Sazonalidade Aditiva e correção do Erro Multiplicativa (MAdA) obteve as menores estatísticas, além disso, este modelo obteve os menores valores para as estatísticas de previsão, sendo assim, o modelo MAdA é o modelo que gera a previsão mais acurada dentre os métodos de SE.

Evidencia-se que mesmo o UTheil sendo menor que um, ou seja, a previsão do modelo é melhor que uma previsão ingênua, verifica-se que os valores de erros são altos para uma boa previsão, como o MAPEp de 58,9% de erro e uma média absoluta de imóveis vendidos de 775 unidades. Para uma melhor visualização dos valores de previsão, na Figura 25 observa-se

o gráfico com a previsão gerada pelo modelo MAdA comparada à série de Vendas utilizada para avaliação dos métodos.

Figura 25 - Comparação das previsões geradas pelos modelos de MAdA comparada à série de Vendas de imóveis residências na cidade de São Paulo



Fonte: Elaborado pelo autor.

Nota-se que a previsão gerada pelo método MAdA é maior que a série para praticamente todos os períodos, com exceção dos primeiros meses de avaliação. O que pode ser explicado pelo valor das constantes de alisamento  $\alpha$ . Quanto menor o valor de  $\alpha$  mais estáveis serão as previsões finais, uma vez que pesos maiores são dados às observações passadas, desta maneira, qualquer flutuação aleatória, no presente, exercerá um peso bem menor no cálculo de previsões.

O valor da constante de suavização  $\alpha$  do modelo MAdM, o qual apresentou melhor previsão, é o menor dentre todos os modelos, porém na modelagem de determinação dos valores das quatro constantes, esse modelo obteve a menor soma dos quadrados dos erros da previsão. Em tempo, considera-se que a sazonalidade e a tendência da série prevista acompanharam bem a da série de Vendas, porém o nível foi superprevisto.

Assim sendo, o uso do modelo MAdA individual para previsão implicaria em um número superior de imóveis lançados ao necessário para atender a demanda, desse modo os estoques cresceriam e os custos das incorporadoras aumentariam. Essa situação seria semelhante ao que aconteceu nos anos de 2014 e 2015 quando os estoques foram recordes devidos aos inúmeros lançamentos de anos anteriores, fato este explicado na Seção 2.1

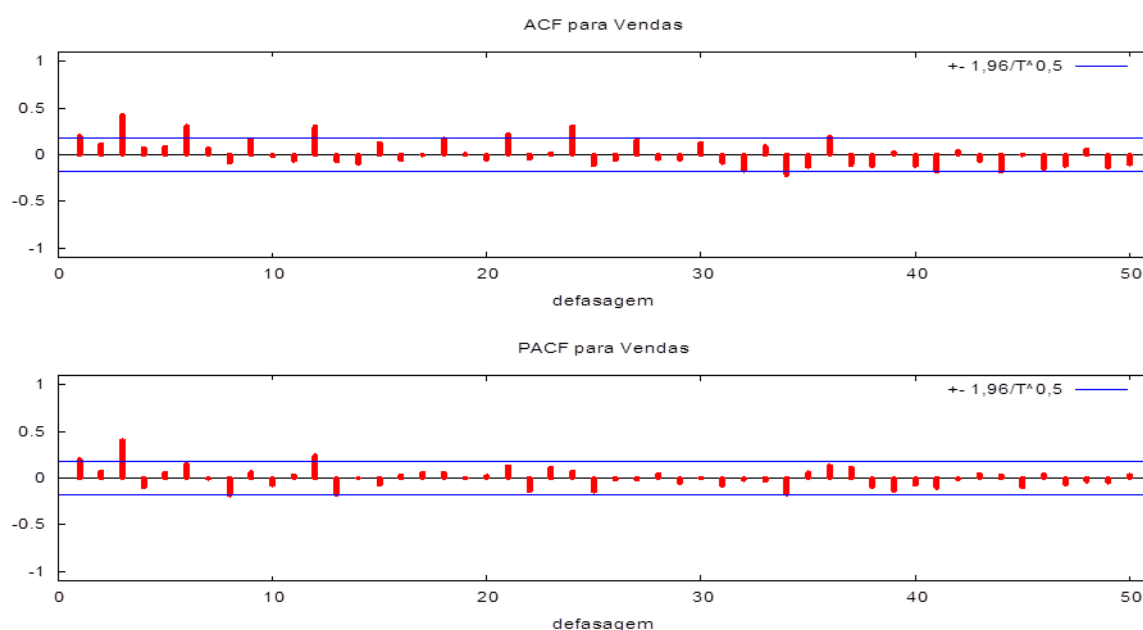
### 4.2.2 Metodologia de Box-Jenkins

Esta seção compreende a terceira etapa da pesquisa e apresenta os resultados da modelagem da série temporal das vendas dos imóveis residências na cidade de São Paulo utilizando a Metodologia Box-Jenkins. O processo de modelagem foi realizado utilizando o software *Gretl* 1.9.5 cvs. As medidas de acurácia foram calculadas no software *Excel*<sup>®</sup>.

Como pode ser observada na Figura 20, a série apresenta evidências de um comportamento não estacionário, visto que sua média varia no processo, do mesmo modo que sua variância também é diferente durante instantes díspares da série. É notório também que nos meses do final do ano há picos de vendas, fato comprovado com testes de sazonalidade, sendo possível observar indícios que a modelagem apresente parâmetros sazonais.

A primeira etapa para a utilização da metodologia de Box-Jenkins é a análise das funções de Autocorrelação (ACF) e de Autocorrelação Parcial (PACF). Essas funções, somado à visualização do gráfico da série temporal, auxiliam na verificação da condição de estacionariedade e contribui para identificar os possíveis modelos para a série temporal. Essas funções podem ser observadas na Figura 26, onde é visível que a ACF apresenta decaimento rápido, contrariando a hipótese de que a série seja não estacionária, assim evidenciando a hipótese de que a série seja estacionária.

Figura 26 - Correlograma da FAC e da FACP da série temporal de vendas de unidades residências em São Paulo



Fonte: Saída do *Gretl*.

Nota-se ainda que tanto na FAC quanto na FACP, o *lag* 12 apresenta comportamento significativo, alimentando a evidência de sazonalidade autorregressiva. Este fato foi verificado também, na Figura 24, onde apresentam-se os índices sazonais, sendo o mês de dezembro, o mês que obteve maior índice. Na FACP podem-se verificar dois *lags* significativos, demonstrando que no modelo pode ser incluso dois parâmetros autorregressivos.

Ainda em relação à estacionariedade da série, para uma adequada conclusão a respeito de tal condição, optou-se por realizar o teste de raiz unitária ADF em nível,  $I(0)$ , apresentado na Tabela 5. Nota-se que para o teste ADF a série apresentou comportamento estacionário para nível de significância de 5%, visto que o valor crítico do teste é menor que -2,88. Desse modo confirma-se a hipótese evidenciada na análise da ACF, onde se presenciou decaimento rápido, somado ao teste ADF, afirma-se que a série é estacionária.

Tabela 5 - Teste ADF aplicado à série de vendas de imóveis

		ADF	
		t-Statistic	Prob.*
I (0)	Crítico	-3,331414	0,0158
	1% level	-3,491345	
	5% level	-2,888157	
	10% level	-2,581041	

Fonte: Saída do *Gretl* adaptado pelo autor.

Após a realização do teste de verificação de raiz unitária, procedeu-se para a segunda fase da metodologia de Box-Jenkins, que é a identificação dos modelos potenciais juntamente a análise da ACF e PACF. A análise da ACF e PACF possibilitou a identificação de seis potenciais modelos para a modelagem do processo gerador da série, que são apresentados na Tabela 6 juntamente com os valores dos critérios penalizadores AIC e BIC.

Tabela 6 - Modelos concorrentes à estimação dos parâmetros e modelagem do processo

	AIC	BIC	Observações
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	Modelo selecionado para a estimação
SARIMA (2,0,0)(2,0,0) <sub>12</sub>	13,6255	13,7225	Modelo selecionado para a estimação
SARIMA (2,0,0)(3,0,0) <sub>12</sub>	13,5668	13,6882	Modelo selecionado para a estimação

Fonte: Elaborado pelo autor.

Salienta-se que vários modelos foram testados, porém não são modelos concorrentes, pois seus parâmetros foram maiores que 1 ou não foram estatisticamente significantes. Resolveu-se por escolher os três modelos com menores AIC e BIC para estimação dos parâmetros, sendo assim observa-se que o modelo SARIMA(2,0,0)(1,0,0)<sub>12</sub>, o modelo SARIMA(2,0,0)(2,0,0)<sub>12</sub> e o modelo SARIMA(2,0,0)(3,0,0)<sub>12</sub> são aqueles candidatos à estimação e posterior modelagem, pois dentre os demais, minimizaram os critérios AIC e BIC. O próximo passo foi estimar os parâmetros dos modelos selecionados e comparar os valores das medidas de acurácia para o ajustamento e previsão da série, objetivando definir qual modelo SARIMA gera a melhor previsão. Na Tabela 7 são mostrados os resultados da estimação dos parâmetros e os valores das medidas de acurácia.

Tabela 7 - Estimação dos parâmetros e acurácia de ajustamento e previsão dos modelos concorrentes da modelagem Box-Jenkins

	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>
<b>Parâmetros</b>	$\varphi_1 = 0,52288$ $\varphi_2 = 0,32194$ $\Phi_1 = 0,5479$	$\varphi_1 = 0,47478$ $\varphi_2 = 0,27504$ $\Phi_1 = 0,4091$ $\Phi_2 = 0,4051$	$\varphi_1 = 0,46748$ $\varphi_2 = 0,30427$ $\Phi_1 = 0,2891$ $\Phi_2 = 0,3219$ $\Phi_3 = 0,2961$
<b>RMSEa</b>	974,0019	<b>698,8654*</b>	864,8135
<b>MAPEa</b>	0,319	<b>0,2831*</b>	0,2832
<b>MAEa</b>	724,7394	<b>556,603*</b>	651,6471
<b>UTHEILa</b>	0,7717	<b>0,553725*</b>	0,6852
<b>RMSEp</b>	804,9413	<b>643,2946*</b>	887,4524
<b>MAPEp</b>	0,3673*	<b>0,386502</b>	0,5599
<b>MAEp</b>	636,6731	<b>541,1966*</b>	711,3658
<b>UTHEILp</b>	0,686	<b>0,548278*</b>	0,7564

\*Melhor modelo para cada medida de acurácia.

Fonte: Elaborado pelo autor.

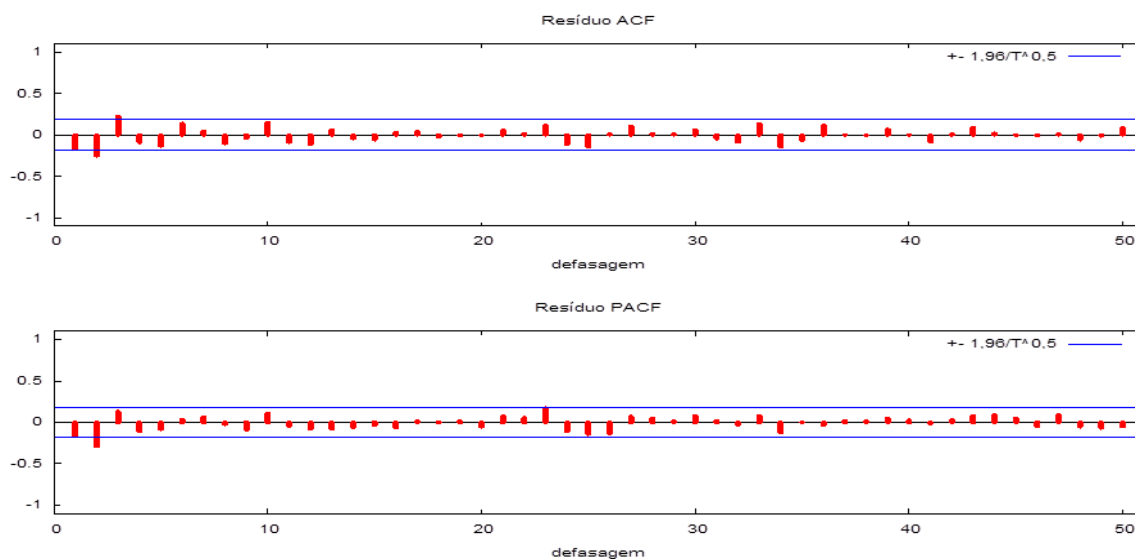
Evidencia-se que os modelos concorrentes e potenciais bons preditores encontrados são todos modelos com ruído branco, ou seja, média igual a zero e variância constante ao longo do tempo, somado ao fato de que os parâmetros são todos menores do que um.

Na Tabela 7 verifica-se que o melhor modelo de ajuste da série é o modelo SARIMA(2,0,0)(2,0,0)<sub>12</sub>, destacado em negrito, tendo apresentado os menores valores para as medidas de ajuste, do mesmo modo, o modelo obtém os menores erros de previsão para todas as medidas. Este, obteve 38,6% de erro percentual médio e um valor de UTheil de previsão de 0,54. Sendo assim, afirma-se que este modelo é o melhor predictor para a série temporal em estudo dentre os modelos de Box-Jenkins. Após estas análises, com o intuito de confirmar a parcimônia do modelo é apresentado o correlograma dos resíduos na Figura 27, o qual



demonstra graficamente que os resíduos da modelagem SARIMA(2,0,0)(2,0,0)<sub>12</sub> não são autocorrelacionados, visto que de maneira aceitável praticamente todos os *lags* estão dentro do intervalo de confiança.

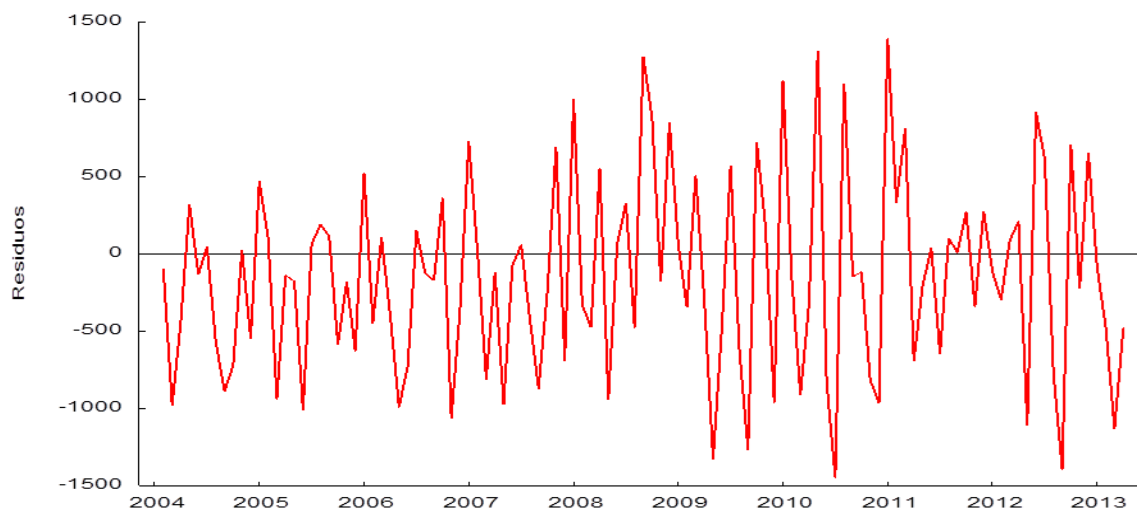
Figura 27 - Correlograma dos resíduos Modelo SARIMA(2,0,0)(2,0,0)<sub>12</sub>



Fonte: Saída do *Gretl*.

Na Figura 28 pode-se observar o gráfico dos resíduos comparado com o tempo, de maneira aceitável percebe-se que os resíduos são ruído branco, o que é confirmado no efeito ARCH na próxima seção.

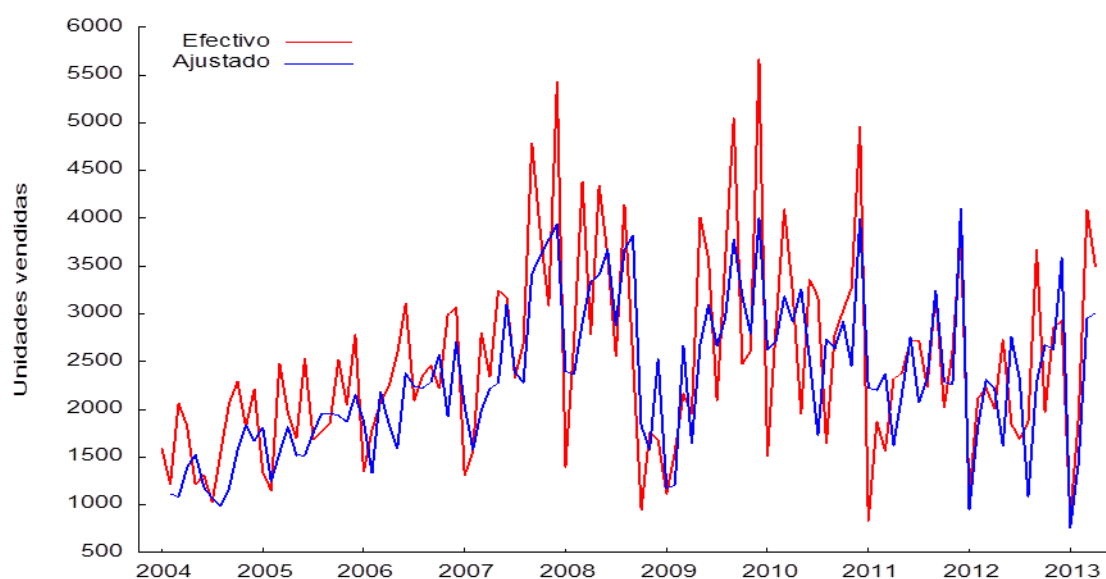
Figura 28 - Gráfico dos resíduos do modelo SARIMA(2,0,0)(2,0,0)<sub>12</sub> comparado com o tempo



Fonte: Saída do *Gretl*.

Na Figura 29 apresenta-se a série ajustada comparada a observada, nota-se que o modelo ajustado consegue representar de maneira fidedigna a série original, demonstrando um bom ajuste, salienta ainda que os valores dos critérios penalizadores AIC e BIC do modelo escolhido, mesmo não sendo os menores, foram bem próximos do melhor modelo gerador (menores critérios de AIC e BIC).

Figura 29 - Série de vendas e série ajustada pelo modelo SARIMA (2,0,0)(2,0,0)<sub>12</sub>

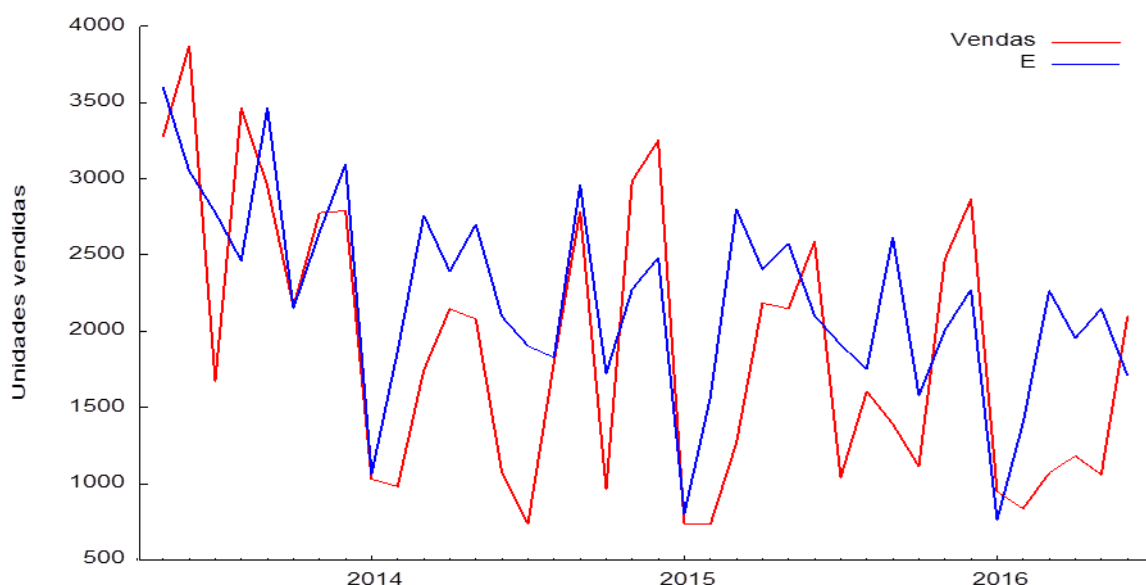


Fonte: Saída do *Gretl*.

Na Figura 30 pode-se verificar a previsão gerada pelo modelo SARIMA (2,0,0)(2,0,0)<sub>12</sub> (na figura representado pela letra E) comparada com a série de vendas utilizada para comparação da previsão. Verifica-se que a previsão gerada pelo modelo traduz bem as características da série como a sazonalidade e também a tendência, acompanhando a série com variância, até o final do ano de 2015. Essa variância pode ser explicada pelos parâmetros autorregressivos sazonais, o que difere do modelo ARIMA simples, o qual não possui parâmetro sazonal.

A partir do ano de 2016 a previsão gerada pelo modelo é bem maior que as vendas reais, desse modo constata-se que modelo gera uma boa previsão, porém em um longo prazo não traduz o comportamento ideal de vendas de imóveis, isso deve-se ao fato que os modelos de Box-Jenkins são modelos preditivos de curto prazo.

Figura 30 - Previsão gerada pelo modelo SARIMA (2,0,0)(2,0,0)<sub>12</sub> comparada com a série de vendas entre Mai/13 e Jun/16



Fonte: Saída do *Gretl*.

#### 4.2.3 Modelos com Heterocedasticidade Condicional

A quarta etapa da pesquisa compreende a modelagem dos modelos de Heterocedasticidade Condicional. Para utilizar esses modelos é necessário que os resíduos dos modelos de Box-Jenkins tenham variância evoluindo no tempo.

Para verificar se a série possui o efeito ARCH - somente se existir esse efeito pode-se modelar os métodos com Heterocedasticidade Condicional - na Tabela 8 é apresentado o resultado do teste ARCH-LM aplicado aos resíduos ao quadrado do modelo SARIMA(2,0,0)(2,0,0)<sub>12</sub>, o melhor modelo de Box-Jenkins encontrado nessa pesquisa. O teste possui as seguintes hipóteses:  $H_0$ : Hipótese de homocedasticidade e  $H_1$ : Hipótese de heterocedasticidade. Para realização do teste utilizou-se o software *Gretl*.

Verifica-se dessa maneira que para as 12 defasagens testadas o efeito ARCH não foi encontrado, somente para o intercepto, desse modo os modelos de heterocedasticidade condicional não poderão ser utilizados nesse estudo. Além deste, outro agravante seria o baixo número de observações, visto que para gerar previsões para este tipo de modelos é necessário uma gama alta de dados.

Tabela 8 - Teste de presença do efeito ARCH por meio do teste ARCH-LM

Defasagens	Estimativa	Erro Padrão	Estatística <i>t</i>	p-valor
alpha(0)	475928	23846	1,996	0,04908*
alpha(1)	-0,027658	0,107202	-0,258	0,79702
alpha(2)	-0,002252	0,117656	-0,019	0,98478
alpha(3)	0,372913	0,117714	0,468	0,64124
alpha(4)	-0,000862	0,124497	-0,007	0,99449
alpha(5)	0,137913	0,126512	1,09	0,27867
alpha(6)	-0,088139	0,127239	-0,693	0,49034
alpha(7)	0,060428	0,126	0,48	0,63273
alpha(8)	0,113738	0,126953	0,896	0,37277
alpha(9)	-0,013184	0,127183	-0,104	0,91768
alpha(10)	0,035623	0,118279	0,301	0,764
alpha(11)	-0,100653	0,118183	-0,852	0,39674
alpha(12)	-0,003423	0,117074	-0,029	0,97674

\*parâmetro significativo.

Fonte: Saída do *Gretl* adaptado pelo autor.

#### 4.2.4 Redes Neurais Artificiais

Esta seção compreende a quinta etapa da pesquisa e apresenta os resultados da modelagem da série temporal das vendas dos imóveis residências na cidade de São Paulo utilizando os modelos de RNAs. O processo de modelagem foi realizado utilizando o software *Statistica* versão 10.0. As medidas de acurácia foram calculadas no software *Excel*<sup>®</sup>.

O primeiro passo nessa etapa consistiu na definição da configuração da rede, onde foram estabelecidos, entre outros, o número de neurônios na camada de entrada. O número de neurônios na camada escondida e na camada de saída já foram previamente definidos, visto na metodologia da pesquisa. Considerando-se a existência da componente sazonal, sendo verificadas na análise gráfica, nos testes de sazonalidade e nos modelos de Box-Jenkins, as configurações testadas para a camada de entrada dos modelos MLP foram de 3, 6, 12 e 18 neurônios na camada de entrada. A configuração da rede está apresentada no Quadro 4, contendo o número de neurônios em cada camada.

Quadro 4 - Configuração das RNAs modeladas no presente estudo

Rede MLP	Número de neurônios
Camada de entrada	3, 6, 12 e 18
Camada escondida	5, 10, 15, 20, 25, 30
Camada de saída	1
Número de modelos	24

Fonte: Elaborado pelo autor.

Em seguida, realizou-se o treinamento da rede dos modelos encontrados a partir das configurações consideradas, onde foram estabelecidos os pesos das conexões entre cada camada, para cada modelo. A primeira fase do treinamento foi realizada pelo algoritmo *Backpropagation* e a segunda fase foi realizada pelo algoritmo *Conjugate Gradiente Descent*. Para todos os modelos de RNAs utilizados neste estudo, foi estipulado o número máximo de 20 mil ciclos de treinamento, tendo como objetivo evitar o superajustamento aos dados de treinamento.

Após o treinamento realizou-se a modelagem e previsão da série por meio dos modelos MLP treinados e, então, foram contabilizados os resultados. Por questão de organização, utilizou-se na descrição dos modelos a notação do modelo MLP como no exemplo – MLP (3,5,1) – sendo este o modelo MLP com 3 neurônios na camada de entrada, 5 neurônios na camada intermediária e 1 neurônio na camada de saída. Os resultados da modelagem e previsão são apresentados na Tabela 9.

Tabela 9 - Resultados dos modelos MLP/RNA utilizados para a modelagem e previsão da série temporal de Vendas de imóveis residenciais na cidade de São Paulo

Modelos MLP	Ajustamento				Previsão			
	RMSE	MAPE	MAE	UTHEIL	RMSE	MAPE	MAE	UTHEIL
MLP (3,5,1)	906,859	0,328	694,335	0,714	1091,154	0,8	934,395	0,93
MLP (3,10,1)	825,346	0,328	651,996	0,65	1768,259	1,299	1533,133	1,508
MLP (3,15,1)	797,839	0,265	597,457	0,628	812,143	0,506	692,823	0,693
MLP (3,20,1)	831,347	0,307	648,433	0,655	939,474	0,664	817,996	0,801
MLP (3,25,1)	1006,564	0,417	768,508	0,793	993,614	0,669	801,573	0,847
MLP (3,30,1)	906,265	0,321	653,475	0,714	1076,036	0,606	792,331	0,918
MLP (6,5,1)	836,343	0,298	630,733	0,65	820,603	0,463	613,658	0,7
MLP (6,10,1)	840,764	0,307	658,251	0,654	1026,609	0,734	826,597	0,875
MLP (6,15,1)	1058,669	0,394	807,78	0,823	1145,531	0,733	894,217	0,977
MLP (6,20,1)	896,374	0,307	684,943	0,697	1121,319	0,801	923,242	0,956
MLP (6,25,1)	911,18	0,36	731,318	0,709	960,876	0,649	730,305	0,819
MLP (6,30,1)	922,564	0,318	671,165	0,717	1026,054	0,598	844,808	0,875
MLP (12,5,1)	826,357	0,331	645,773	0,626	903,707	0,619	720,382	0,771
MLP (12,10,1)	789,608	0,294	635,517	0,598	759,599	0,473	624,406	0,648
MLP (12,15,1)	750,918	0,255	577,388	0,569	725,487	0,329	543,265	0,619
MLP (12,20,1)	938,611	0,344	677,885	0,711	1047,56	0,587	776,869	0,893
MLP (12,25,1)	862,404	0,235	539,625	0,653	810,658	0,536	745,618	0,691
MLP (12,30,1)	792,737	0,202	438,363	0,601	776,777	0,347	563,161	0,663
MLP (18,5,1)	833,448	0,313	662,325	0,619	763,632	0,462	648,882	0,651
MLP (18,10,1)	827,629	0,295	591,329	0,614	763,461	0,505	687,824	0,651
MLP (18,15,1)	1036,188	0,392	807,553	0,769	969,7	0,595	723,265	0,827
MLP (18,20,1)	840,577	0,16*	422,391*	0,424	551,561	0,45	496,707	0,441
MLP (18,25,1)	<b>535,369*</b>	<b>0,192</b>	<b>443,578</b>	<b>0,398*</b>	<b>402,327*</b>	<b>0,234*</b>	<b>334,986*</b>	<b>0,343*</b>
MLP (18,30,1)	950,393	0,262	514,824	0,705	846,901	0,54	786,978	0,722

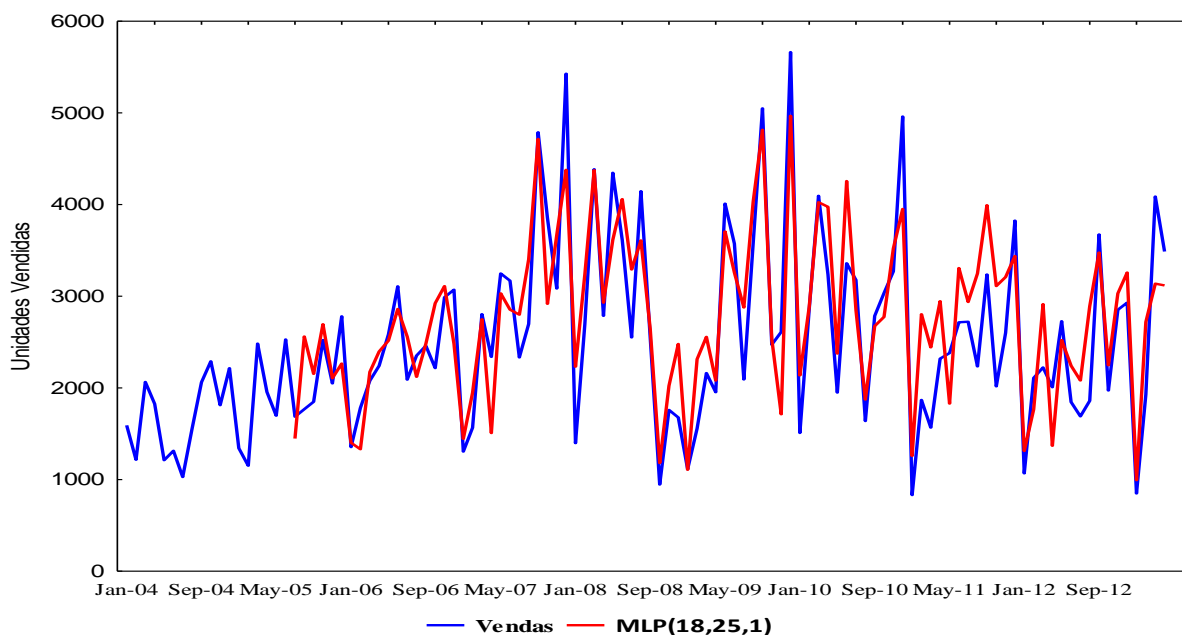
\*Melhor modelo para cada medida de acurácia.

Fonte: Elaborado pelo autor.

O modelo destacado em negrito é o que gerou o melhor ajuste e a melhor previsão. Ainda que para os valores de MAPE e MAE o modelo não apresentou melhores resultados, escolheu-se este para melhor ajuste pelo fato de que o valor de UTheil de ajuste foi muito menor que os outros modelos. O modelo RNA/MLP (18,25,1) obteve 23,4% de MAPE e UTHEIL de 0,343, mostrando que o modelo gera uma boa previsão.

Para uma análise gráfica na Figura 31 é apresentada a série ajustada gerada pelo melhor modelo MLP/RNA (18,25,1) comparada à série original de vendas de imóveis na capital paulista.

Figura 31 - Série de vendas de imóveis de São Paulo e série ajustada pelo modelo RNA/MLP (18, 25, 1)



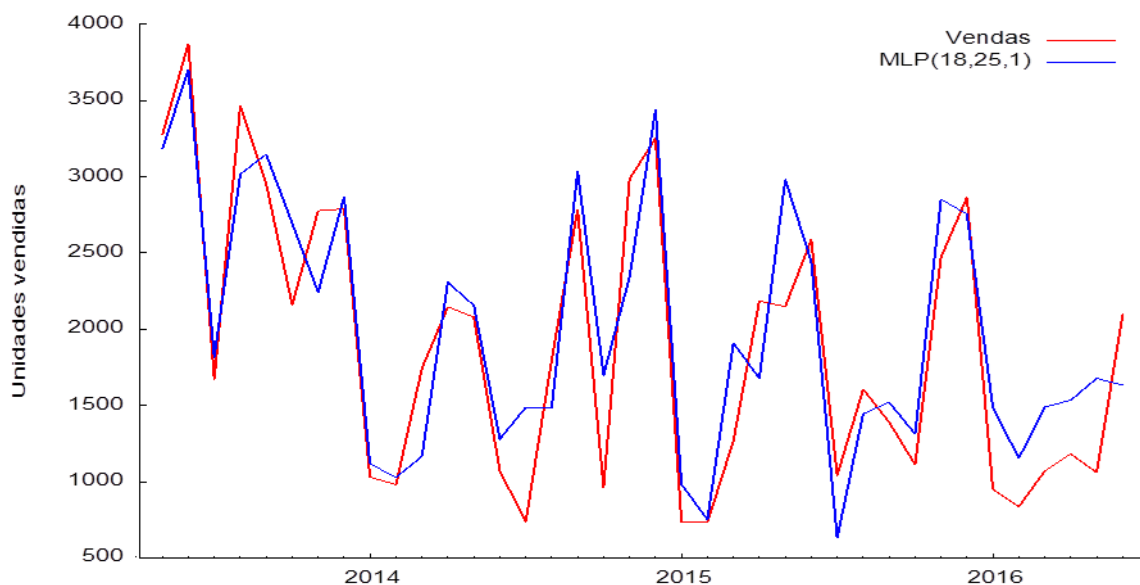
Fonte: Elaborado pelo autor.

Pela Figura 31 nota-se que o modelo ajustado consegue representar de maneira fidedigna a série original, demonstrando um bom ajuste. Na Figura 32 apresenta-se a previsão gerada pelo modelo RNA/MLP (18, 25, 1), comparada com a série de Vendas utilizada para avaliação das previsões.

Para ter a certeza que o modelo testado não apresentou super ajustamento realizou-se uma regressão linear simples, assim o grau de ajuste do modelo de regressão linear simples mediu o quanto o modelo MLP (18,25,1) foi capaz de descrever os valores alvo apresentados, sendo analisado por meio do coeficiente de determinação  $R^2$ . No treinamento, foram apresentados à RNA os valores alvo constantes no conjunto de treinamento dos modelos e, no

processo de teste foram apresentados os valores alvo não apresentados no treinamento, possibilitando verificar o potencial do modelo em prever os valores *out-of-sample* e da ocorrência de super ajustamento.

Figura 32 - Previsão gerada pelo modelo RNA/MLP (18,25,1) comparada com a série de vendas entre Mai/13 e Jun/16

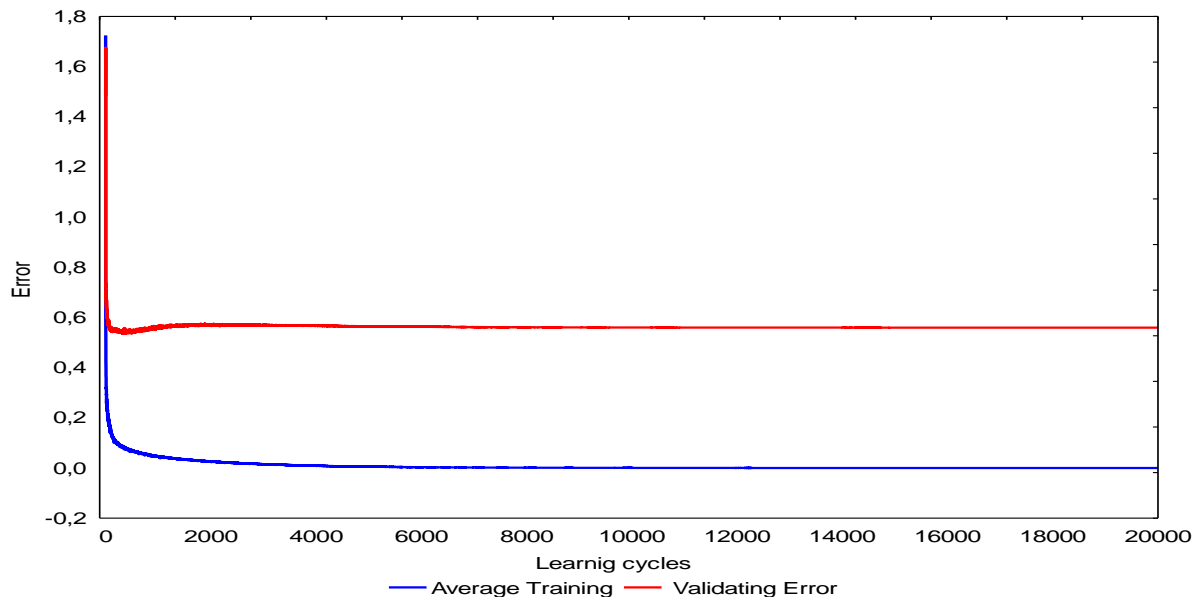


Fonte: Elaborado pelo autor.

O valor de  $R^2$  obtido no treinamento do modelo foi igual a 0,812 e, para o processo de validação, foi obtido o valor  $R^2$  igual a 0,798. Assim é possível concluir que, no treinamento do modelo MLP (18,25,1) não ocorreu o super ajustamento visto que não há grande variância entre o  $R^2$  de treinamento e  $R^2$  de validação, afirmando assim que o modelo não apresenta alta variância ao lidar com dados desconhecidos. Outro teste para confirmar que não houve super ajustamento é a análise do gráfico de treinamento do modelo MLP(18,25,1), este apresentado na Figura 33.

Observou-se que o erro médio do processo de treinamento (0,0097) convergiu ao erro meta estipulado que foi de 0,01. Já o erro do processo de validação foi de 0,566, verificando que o modelo atingiu valores aceitáveis de erro para os 20000 ciclos de treinamento.

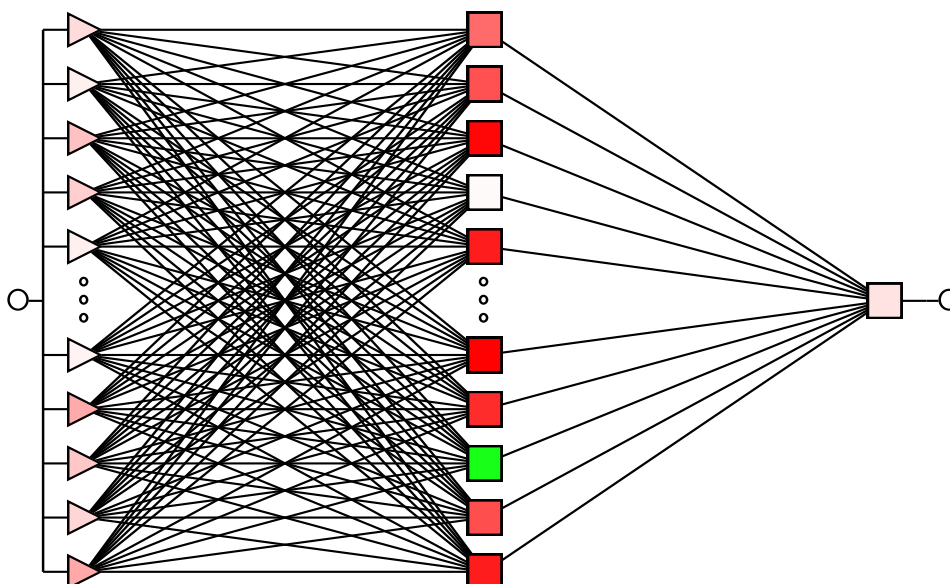
Figura 33 - Gráfico de treinamento da rede MLP(18,25,1)



Fonte: Saída do *Statistica*.

Na Figura 34 apresenta a arquitetura da rede do modelo MLP/RNA (18, 25,1), sendo a entrada que é a série temporal de vendas de unidades de imóveis residências na cidade de São Paulo, tendo 18 neurônios na camada de entrada, 25 neurônios na camada escondida, ou camada intermediária, um neurônio na camada de saída e tendo como saída a previsão da série temporal em estudo.

Figura 34 - Arquitetura da rede do modelo MLP/RNA(18,25,1)



Fonte: Saída do *Statistica*.



Encerrando esta seção salienta-se que pela série apresentar componente sazonal, podendo ser observada que na modelagem ARIMA, o tamanho da sazonalidade encontrada foi de 12. Esperava-se, assim que o método com 12 neurônios na camada de entrada fosse produzir os melhores resultados, fato este que não aconteceu, visto que o melhor modelo têm 18 neurônios na camada de entrada.

Isto pode ser explicado, pois os meses de Junho e Setembro são os dois meses que após dezembro mais registraram vendas de imóveis na cidade de São Paulo, sendo assim de 18 em 18 os meses sazonais serão junho e dezembro, ou seja, o modelo capta bem a sazonalidade da série. Esse fato é afirmado quando, novamente observado o teste de presença de sazonalidade, o qual acusou a presença de tamanhos de sazonalidade de 6 e 12 meses. Essa sazonalidade em junho e dezembro pode ser explicada pelo pagamento das parcelas do décimo terceiro salário, assim os consumidores aproveitam esse pagamento para dar entrada na compra de um novo imóvel.

#### 4.3 COMBINAÇÕES E MODELAGEM HÍBRIDA

Com a obtenção dos modelos individuais, podem-se combinar as previsões para a série utilizando os métodos de combinação descritos na seção 2.2.1.5. Esta seção apresenta a sexta etapa de pesquisa, a de combinação das previsões dos modelos MAdA, SARIMA(2,0,0)(2,0,0)<sub>12</sub> e MLP(18,25,1). Por questão de organização a notação *underline* ( ) significa a combinação entre os métodos, e a notação barra (/) significa o método de combinação utilizado.

O primeiro método de combinação realizado foi o da Média Aritmética (MD) por meio da Equação 63, gerando os respectivos modelos MAdA\_SARIMA/MD, MAdA\_RNA/MD, SARIMA\_RNA/MD e MAdA\_SARIMA\_RNA/MD.

O segundo método foi o da combinação por meio de Média Geométrica (MG) por meio da Equação 64, obtendo os seguintes modelos: MAdA\_SARIMA/MG, MAdA\_RNA/MG, SARIMA\_RNA/MG e MAdA\_SARIMA\_RNA/MG.

O método da Média Harmônica (MH), Equação 65, foi utilizado para combinação obtendo os respectivos modelos, MAdA\_SARIMA/MH, MAdA\_RNA/MH, SARIMA\_RNA/MH e MAdA\_SARIMA\_RNA/MH.

Em seguida, o método da Variância Mínima (VAM) foi utilizado para combinação, obtendo os modelos, MAdA\_SARIMA/VAM, MAdA\_RNA/VAM, SARIMA\_RNA/VAM e MAdA\_SARIMA\_RNA/VAM respectivamente. A primeira etapa para esta combinação foi a

obtenção dos pesos atribuídos a cada previsão. Na Tabela 10 são apresentados os pesos para cada modelo individual diante do modelo combinado, salienta-se que se utilizou a Equação 61 para cálculo dos pesos e Equação 62 para combinação.

Tabela 10 - Pesos  $k$  para cada modelo para definição da combinação de previsão por meio do método VAM.

	<b>MAdA_SARIMA</b>	<b>MAdA_RNA</b>	<b>SARIMA_RNA</b>	<b>MAdA_SARIMA_RNA</b>
<b>MAdA (<math>k</math>)</b>	0,505	0,605	-	0,38
<b>SARIMA (<math>k</math>)</b>	0,496	-	0,601	0,373
<b>RNA (<math>k</math>)</b>	-	0,396	0,4	0,249

Fonte: Elaborado pelo autor.

O penúltimo método de combinação utilizado foi o da Regressão Linear (RL), conseqüentemente os modelos apresentados são MAdA\_SARIMA/RL, MAdA\_RNA/RL, SARIMA\_RNA/RL e MAdA\_SARIMA\_RNA/RL. Desse modo, realizaram-se quatro regressões lineares, uma para cada método, na Tabela 11, pode-se observar os valores dos coeficientes de regressão e o  $R^2$  Ajustado para cada modelo individual diante do modelo combinado.

Tabela 11 - Coeficiente  $w$  para cada modelo para definição da combinação de previsão por meio do método RL

	<b>MAdA_SARIMA</b>	<b>MAdA_MLP</b>	<b>SARIMA_MLP</b>	<b>MAdA_SARIMA_MLP</b>
<b>Intercepto (<math>w_0</math>)</b>	-257,115	-452,052	-455,375	-530,698
<b>MAdA (<math>w</math>)</b>	0,536	0,307	-	0,123
<b>SARIMA (<math>w</math>)</b>	0,607	-	0,364	0,301
<b>RNA (<math>w</math>)</b>	-	0,831	0,793	0,762
<b><math>R^2</math> Ajustado</b>	56,24%	75,30%	76,92%	76,90%

Fonte: Elaborado pelo autor.

O último método de combinação utilizado foi o de Análise de Componentes Principais (ACP), obtendo os modelos MAdA\_SARIMA/ACP, MAdA\_RNA/ACP, SARIMA\_RNA/ACP e MAdA\_SARIMA\_RNA/ACP. Para este método procurou-se obter apenas um fator (componente principal), o qual possui a maior variância explicada, visto que estamos trabalhando com no máximo três variáveis (três modelos individuais de previsão). Cada fator gera uma combinação linear das variáveis originais com o propósito de reter o máximo de informação em termos de variação total, ou seja, de contribuição de cada variável.

Na Tabela 12 podem-se observar os valores de contribuição de cada variável e a respectiva variância explicada por cada modelo.

Tabela 12 - Contribuição de cada modelo individual no Fator 1 para combinação linear por meio do método ACP

	MAdA_SARIMA	MAdA_RNA	SARIMA_RNA	MAdA_SARIMA_RNA
<b>MAdA</b>	0,428	0,339	-	0,261
<b>SARIMA</b>	0,572	-	0,377	0,297
<b>RNA</b>	-	0,661	0,623	0,442
<b>Variância Explicada</b>	85,37%	79%	75,52%	71,26%

Fonte: Elaborado pelo autor.

Obtidas as combinações das previsões dos modelos, podem-se calcular os resultados de cada combinação para o conjunto de ajuste e para o conjunto de previsão e então selecionar a melhor para a previsão da série. Na Tabela 13 são apresentados os resultados das medidas de acurácia para o ajuste e previsão dos modelos combinados.

Tabela 13 - Resultados dos cálculos das medidas de acurácia para ajuste e previsão dos modelos combinados

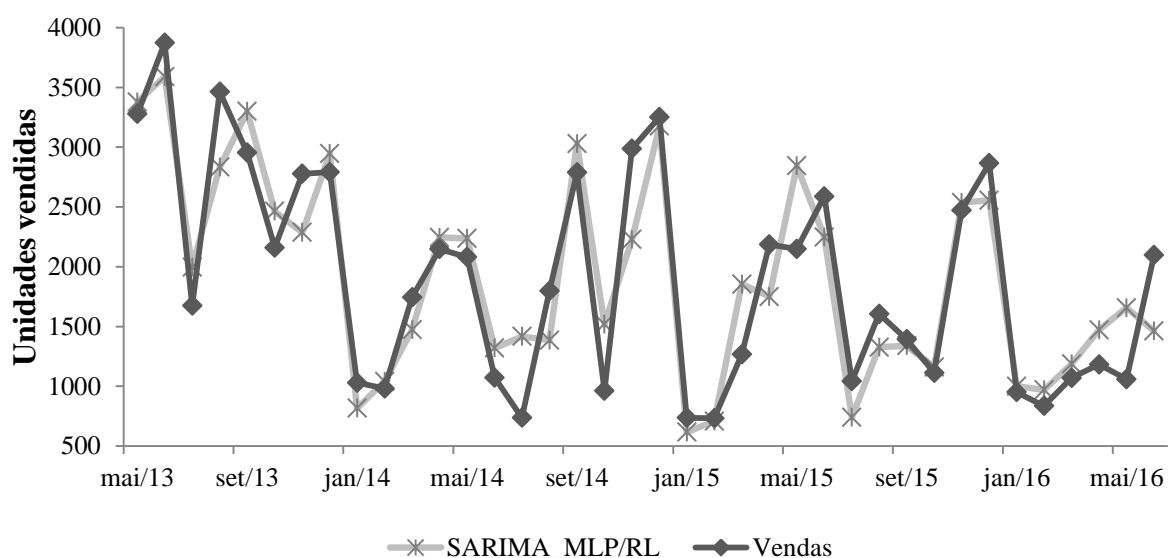
Modelos Combinados	Ajustamento				Previsão			
	RMSE	MAPE	MAE	UTHEIL	RMSE	MAPE	MAE	UTHEIL
MAdA_SARIMA/MD	641,77	0,217	498,83	0,509	725,569	0,462	594,3	0,619
MAdN_RNA/MD	544,65	0,194	446,02	0,405	570,831	0,372	486,6	0,487
SARIMA_RNA/MD	521,9	0,183	431,79	0,388	447,71	0,274	379	0,382
MAdN_SARIMA_RNA/MD	555,38	0,189	444,94	0,412	560,372	0,358	458,5	0,478
MAdA_SARIMA/MG	640,86	0,215	498,73	0,508	715,429	0,454	585,4	0,61
MAdN_RNA/MG	540,13	0,191	443,19	0,401	537,598	0,339	450,4	0,459
SARIMA_RNA/MG	526,18	0,181*	433,92	0,391	436,809	0,258	363,8	0,373
MAdN_SARIMA_RNA/MG	554,85	0,185	441,82	0,412	530,906	0,332	430,2	0,453
MAdA_SARIMA/MH	641,87	0,215	500,78	0,509	706,292	0,447	578,2	0,602
MAdN_RNA/MH	537,37	0,188	442,33	0,399	512,568	0,315	425,3	0,437
SARIMA_RNA/MH	533,33	0,182	439,97	0,396	430,691	0,252	360,4	0,368
MAdN_SARIMA_RNA/MH	556,63	0,183	445,11	0,413	506,164	0,307	406,2	0,432
MAdA_SARIMA/VAM	641,8	0,217	498,78	0,509	726,759	0,463	595,3	0,62
MAdN_RNA/VAM	571,18	0,2	458,73	0,424	632,903	0,417	545,6	0,54
SARIMA_RNA/VAM	548,65	0,187	444,1	0,407	478,837	0,293	406,5	0,409
MAdN_SARIMA_RNA/VAM	575,08	0,193	455,35	0,427	600,67	0,384	490,6	0,512
MAdA_ARIMA/RL	975,96	0,347	755,13	0,772	757,844	0,476	621,5	0,646
MAdN_RNA/RL	1093,4	0,388	840,47	0,865	390,969	0,212	321,9	0,334
SARIMA_RNA/RL	1110,8	0,39	850,34	0,878	<b>368,74*</b>	<b>0,192*</b>	<b>302,1*</b>	<b>0,315*</b>
MAdN_SARIMA_RNA/RL	1100,1	0,388	845,4	0,87	371,286	0,198	303,1	0,317
MAdA_SARIMA/ACP	677,08	0,223	534,68	0,503	706,823	0,447	576,6	0,603
MAdN_RNA/ACP	519,33	0,189	433,8	0,386	488,075	0,308	406,4	0,416
SARIMA_RNA/ACP	<b>501,71*</b>	<b>0,182</b>	<b>426,44*</b>	<b>0,373*</b>	418,083	0,253	350,1	0,357
MAdN_SARIMA_RNA/ACP	528,33	0,186	435,93	0,392	509,813	0,324	417,1	0,435

\*Melhor modelo para cada medida de acurácia.

Fonte: Elaborado pelo autor.

Por meio da Tabela 13 pode-se verificar que o modelo SARIMA\_RNA/ACP é o modelo que melhor se ajusta a série, obtendo os menores valores para as medidas de ajuste, embora segundo a MAPE tal combinação não foi aquela que gerou melhor combinação de ajuste, porém as outras três medidas indicam que esse é o modelo que gera o melhor ajuste. A combinação de previsões que impetrou o melhor resultado para a previsão foi a SARIMA\_RNA/RL, destacado em negrito, tendo os menores valores das medidas de acurácia. Salienta-se que o modelo de regressão foi validado por meio da análise dos resíduos que pode ser verificado no Apêndice A. Na Figura 35 é apresentado o gráfico dos valores previsto pelo modelo SARIMA\_RNA/RL comparado com o valor real da série.

Figura 35 - Gráfico dos valores previstos pelo modelo SARIMA\_RNA/RL comparado com a série de vendas



Fonte: Elaborado pelo autor.

Questionou-se o porquê da série que gerou o melhor ajuste não ter gerado uma boa previsão, o sentido deve-se que a ACP no primeiro fator carrega a maior variância explicada, nesse sentido sabendo que o método de regressão minimiza os quadrados dos erros, entendendo que a previsão gerada pela combinação SARIMA\_MLP/ACP há grande variância nos erros, propõe-se utilizar esta previsão como variável independente em um modelo de regressão, a fim de minimizar os erros que este modelo gerou. Utilizaram-se três modelos de regressão simples, um modelo linear, um modelo quadrático e um modelo cúbico. Na Tabela 14 podem-se verificar os valores dos coeficientes de regressão para cada modelo.

Tabela 14 – Coeficientes e medidas de acurácia para cada modelo de regressão

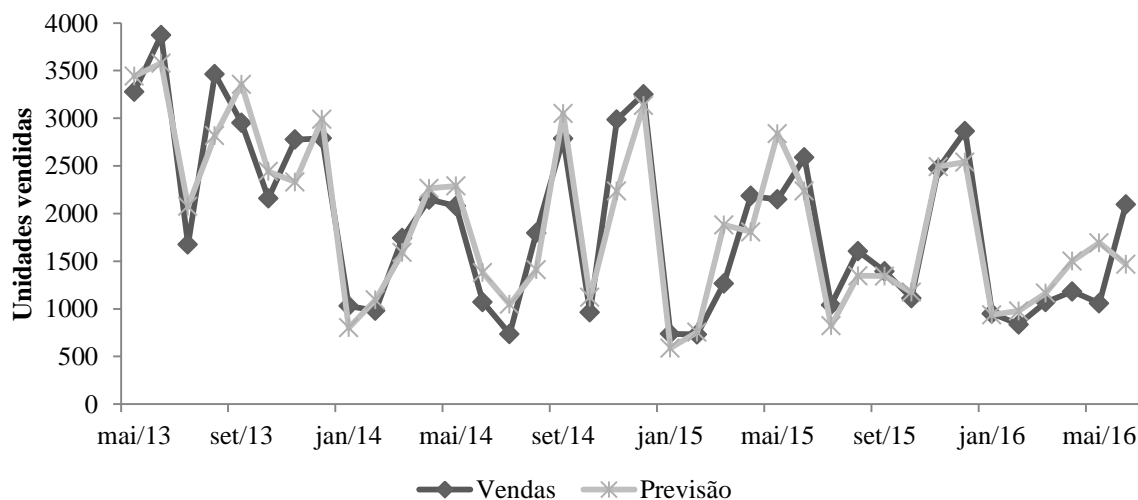
	Linear	Quadrático	Cúbico
<b>Intercepto</b>	-490,187	847,442	-256,957
<b>SARIMA-RNA/ACP</b>	1,178	0,122	1,544
<b>(SARIMA_RNA/ACP)<sup>2</sup></b>	-	0,000192317	-0,000366776
<b>(SARIMA_RNA/ACP)<sup>3</sup></b>	-	-	6,80283E-08
<b>R<sup>2</sup> Ajustado</b>	78,94%	78,86%	78,91%
<b>RMSEa</b>	483,148	460,015	<b>456,997*</b>
<b>MAPEa</b>	0,18	0,173	<b>0,168*</b>
<b>MAEa</b>	419,01	397,264	<b>390,75*</b>
<b>UTHEILa</b>	0,359	0,342	<b>0,339*</b>
<b>RMSEp</b>	<b>349,217*</b>	417,841	406,774
<b>MAPEp</b>	<b>0,171*</b>	0,26	0,237
<b>MAEp</b>	<b>287,625*</b>	359,041	336,615
<b>UTHEILp</b>	<b>0,298*</b>	0,357	0,347

\*Melhor modelo para cada medida de acurácia.

Fonte: Saída do *Action Stat* adaptado pelo autor.

Verifica-se por meio da Tabela 14, que o modelo cúbico obteve os menores valores para as medidas de acurácia de ajuste, destacado em negrito, entretanto são valores próximos dos outros dois modelos. Ao analisar os valores de acurácia para a previsão, nota-se que o modelo linear apresentou melhores resultados para a previsão, obtendo um erro percentual de 17,1% e UTheil de 0,298, assim verifica-se que a previsão gerada pelo modelo linear é considerada satisfatória e melhor que uma previsão ingênua, visto sua proximidade com o valor 0. Na Figura 36 visualiza-se o gráfico da previsão gerada pelo modelo de regressão linear SARIMA\_RNA/ACP comparada com a série de vendas.

Figura 36 - Gráfico dos valores previstos pelo modelo de RL(SARIMA\_RNA/ACP) comparado com a série de vendas



Fonte: Elaborado pelo autor.

Constata-se que a regressão conseguiu minimizar os erros dos resíduos, comparado ao modelo combinado pela ACP, visto, por exemplo, a diminuição do MAPE de previsão de até 8%, assim a solução proposta é satisfatória.

#### 4.4 COMPARAÇÃO DOS MÉTODOS

Na sétima etapa da pesquisa é realizada a comparação dos modelos encontrados na pesquisa, os modelos comparados são o melhor modelo de Suavização Exponencial: MAdA, melhor modelo de Box-Jenkins: SARIMA(2,0,0)(2,0,0)<sub>12</sub>, melhor modelo de RNA: RNA/MLP(18,25,1), o melhor modelo combinado SARIMA\_MLP/RL e o melhor modelo combinado híbrido: RL(SARIMA\_MLP/ACP). Na Tabela 15 é apresentada a comparação entre os modelos.

Tabela 15 - Medidas de acurácia para comparação entre os modelos individuais, combinado e híbrido

Modelos	Ajustamento				Previsão			
	RMSE	MAPE	MAE	UTHEIL	RMSE	MAPE	MAE	UTHEIL
MAdA	705,63	0,26	554,95	0,56	899,52	0,59	775,58	0,77
<b>SARIMA (2,0,0)(2,0,0)<sub>12</sub></b>	698,87	0,29	556,61	0,56	643,3	0,39	541,2	0,55
<b>MLP (18, 25, 1)</b>	535,37	0,2	443,58	0,4	402,33	0,24	334,99	0,35
<b>SARIMA_RNA/RL</b>	1110,8	0,39	850,34	0,88	368,74	0,2	302,1	0,32
<b>RL(SARIMA_RNA/ACP)</b>	<b>483,15*</b>	<b>0,18*</b>	<b>419,01*</b>	<b>0,36*</b>	<b>349,21*</b>	<b>0,17*</b>	<b>287,62*</b>	<b>0,29*</b>

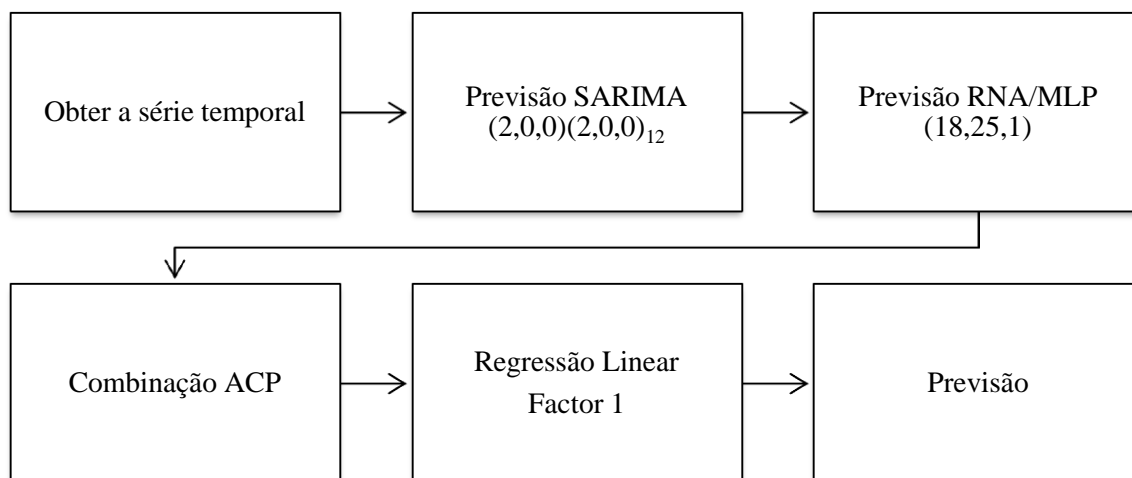
\*Melhor modelo para cada medida de acurácia.

Fonte: Elaborado pelo autor.

Ao comparar os modelos verificou-se que o modelo RL(SARIMA\_RNA/ACP), destacado em negrito, obteve para todas as medidas de acurácia, tanto para ajuste quanto para previsão, valores muito menores que para os outros modelos. O modelo teve 17% de erro relativo e uma média absoluta do erro de 287 unidades por mês. Dessa maneira, afirma-se que o melhor modelo para prever a demanda do mercado imobiliário de São Paulo é a Regressão Linear, utilizando como variável independente o modelo combinado entre SARIMA (2,0,0)(2,0,0)<sub>12</sub> e RNA/MLP (18,25,1) por ACP. Este modelo foi validado por meio da análise de resíduos, a qual pode ser observada no Apêndice B.

Assim sendo, na Figura 37 apresentam as etapas necessárias para realizar a previsão pelo modelo proposto. Essas são as etapas que os interessados em utilizar o modelo deverão seguir para obter a previsão pelo modelo proposto.

Figura 37 - Etapas pertinentes a modelagem do modelo proposto RL(SARIMA\_RNA/ACP)



Fonte: Elaborado pelo autor.

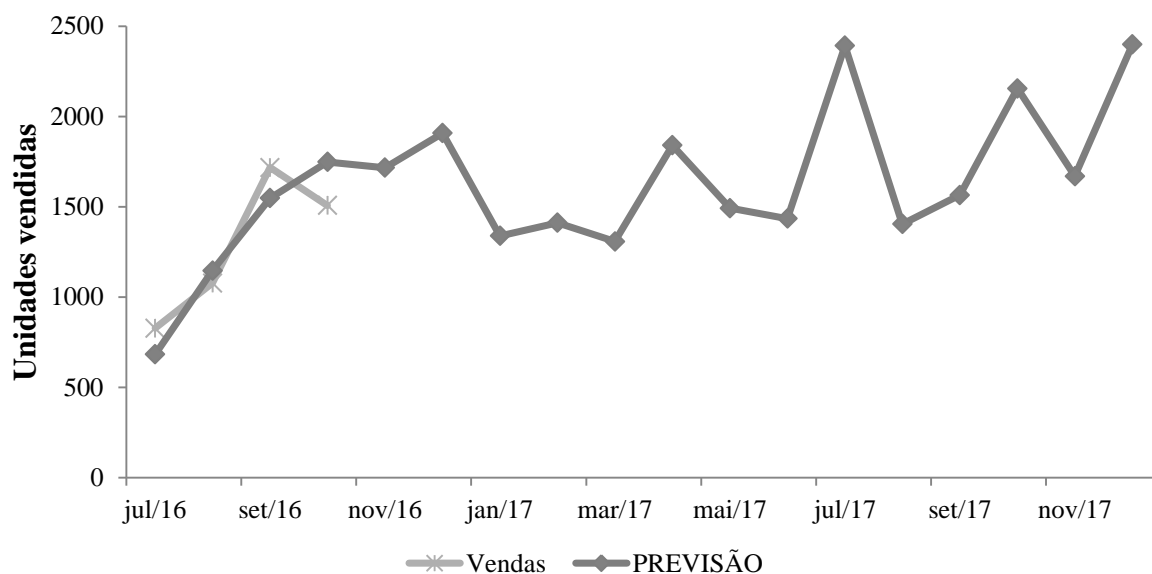
A primeira etapa é a obtenção da série temporal de vendas de imóveis na cidade de São Paulo, evidencia-se que a cada atualização do SECOVI-SP e EMBRAESP, se faz necessário inserir a nova observação na série para que a modelagem seja mais correta. As outras etapas já são descritas na metodologia e nos resultados.

#### 4.5 PREVISÃO

Esta seção corresponde à oitava etapa da pesquisa, aqui o modelo proposto é utilizado para prever a demanda de imóveis residências em São Paulo. Para esta pesquisa utilizou-se os dados compreendendo o período de janeiro de 2004 e junho de 2016, entretanto, no decorrer da pesquisa, os dados de vendas foram sendo atualizados pela EMBRAESP e SECOVI.

Desta forma, o modelo proposto é utilizado para prever a demanda de imóveis residenciais entre julho de 2016 e dezembro de 2017 sendo realizada uma comparação com os meses que não entraram na pesquisa, porém já fechadas e divulgadas as vendas. Na Figura 38 pode-se observar a previsão gerada pelo modelo atualizado comparada a série de vendas que foi atualizada pelo EMPRAESP e SECOVI. Nota-se na Figura 38 que a previsão gerada pelo modelo proposto acompanha bem a série original de vendas nos quatro primeiros meses do segundo semestre de 2016.

Figura 38 - Previsão gerada pelo modelo proposto



Fonte: Elaborado pelo autor.

Percebe-se que o modelo apresenta uma previsão com tendência positiva até o mês de dezembro, entretanto, após isso até março, ou seja, no primeiro trimestre de 2017, o modelo demonstra uma baixa de vendas. A partir do segundo trimestre de 2017 demonstra que o mercado começa a se recuperar, mas é no terceiro trimestre que se percebe uma crescente nas vendas. Além do bom ajuste aos valores de vendas atuais, o resultado do modelo vai ao encontro de vários especialistas da área.

Homer (2016) salienta que o ano de 2016 foi difícil para a concretização de negócios imobiliários. Afetado pela instabilidade política, o mercado financeiro freou financiamentos e prejudicou aqueles que iam a buscar crédito para fechar contratos. O autor ainda afirma que os corretores tiveram maiores trabalhos para convencer os clientes para fechar negócio em um mercado tão adverso. O primeiro semestre de 2016 foi particularmente ruim. Porém, a partir do segundo semestre, o mercado imobiliário começou a ganhar fôlego, o que vai ao encontro do que o modelo proposto apresenta. Esse fôlego pode ser explicado devido a decisão da Caixa Econômica de aumentar o limite de financiamento para imóveis mais caros, assim contribuindo no melhoramento dos índices do ramo imobiliário. Zylberstajn (2016) ainda salienta que além do aumento no limite do Sistema Financeiro de Habitação houve queda na taxa de juros.

Para o ano de 2017, Zylberstajn (2016) acredita que, com a redução dos juros e a retomada da confiança, o mercado pode começar a reagir, porém cita que o emprego ainda



deve demorar em voltar a ter uma dinâmica positiva, então o mercado ainda deve sofrer por algum tempo. Corroborando Araújo (2016) diz que 2017 será um ano melhor, no quesito vendas, que 2016, porém essa mudança será sutil.

O modelo proposto para previsão demonstra que essa mudança que os autores citam será apresentada apenas no terceiro trimestre do ano de 2017, isso se a economia se recuperar, ou o governo passar confiança, visto que em tempos incertos, pouca gente quer fazer dívida de longo prazo. Afinal de contas, com uma economia fortalecida e estável, há uma melhora na empregabilidade, na renda e na confiança do consumidor com o futuro, características que garantem condições propícias para o investimento em um imóvel. Corroborando Armary (2016) insiste na continuidade da redução da taxa de juros, para que o setor imobiliário retome suas atividades aos patamares normais e possa contribuir para a geração de emprego e renda e, conseqüentemente, com o aquecimento da economia.

Por fim, SECOVI-SP (2017) prevê para 2017 uma reversão das tendências negativas da economia, e esse movimento, mesmo que lento, poderá propiciar o crescimento do mercado imobiliário de 5% a 10% no volume de vendas. Neste sentido o modelo, prevê um crescimento superior que o citado pelo SECOVI, o modelo apresenta um crescimento de 20% nas vendas de imóveis na capital paulista em 2017 em relação a 2016.



## 5 CONCLUSÃO

Este capítulo está dividido em duas seções, a primeira remete-se as conclusões dos resultados obtidos e a segunda perpetra uma lista de sugestões para pesquisas futuras envolvendo o tema e resultados observados dessa pesquisa.

### 5.1 CONCLUSÕES ACERCA DOS RESULTADOS OBTIDOS

O estudo objetivou propor um modelo de previsão de demanda para o mercado imobiliário de São Paulo, o qual toasse de auxílio para o processo de tomada de decisão nas incorporadoras, construtoras e para investidores da área imobiliária. Para tanto, dispôs-se utilizar 74 métodos de previsões de séries temporais, somado aos 6 métodos de combinações de previsões, totalizando mais 24 modelos combinados.

Em um primeiro instante encontrou-se as características da série, analisando graficamente e por meio de testes, afirmou-se a presença de tendência e sazonalidade. Sob estes resultados, dos 30 métodos de SE, apenas os adequados para séries com tendência e sazonalidade foram modelados, totalizando oito. Nesta classe, o modelo de Tendência Aditiva amortecida com Sazonalidade Aditiva e correção de erro Multiplicativo, o MAdA, obteve os menores valores para as medidas de acurácia, apresentando 58,6% de MAPE e UTheil de 0,767.

Na metodologia Box-Jenkins, o modelo SARIMA (2,0,0)(2,0,0) com tamanho 12 de sazonalidade obteve os menores valores de acurácia, 30,6% de MAPE e UTheil de 0,548. O não cumprimento de suposições implicaram na não utilização dos métodos de heterocedasticidade condicional, a ausência do efeito ARCH na série impediu a modelagem dessa classe. Foram modelados 24 modelos de Redes Neurais Artificiais, sendo que o modelo RNA/MLP (18,25,1) obteve os menores valores para as estatísticas de erro de previsão, apresentando 23,4% de MAPE e um UTheil de 0,346.

Após as modelagens individuais os modelos foram combinados por meio de 6 métodos de combinação, dentre os 24 modelos combinados, o que apresentou melhores valores para acurácia foi a combinação entre o modelo SARIMA (2,0,0)(2,0,0)12 e RNA/MLP (18,25,1) combinados por meio de Regressão Linear, obtendo 19,2% de erro percentual médio (MAPE) e UTheil de 0,315. Entretanto, percebeu-se que o modelo de melhor ajuste, bem superior aos outros, o modelo que combinava por meio de ACP o SARIMA (2,0,0)(2,0,0)12 e RNA/MLP (18,25,1), não apresentava uma boa previsão, assim resolveu-se utilizar o Factor 1 da ACP

dessa combinação, utilizando-o como variável independente em um Regressão, testou-se regressões linear, quadrática e cúbica. A Regressão Linear apresentou os melhores resultados, 17,1% de erro percentual e UTheil de 0,298.

Ao comparar esses métodos, averiguou-se que o modelo híbrido gerava previsões superiores aos outros modelos, então este foi escolhido o melhor modelo para prever a demanda de imóveis em São Paulo. Por fim, utilizou-se esse modelo para prever a demanda do segundo semestre de 2016 e o ano de 2017, o resultado apresentado vai ao encontro dos profissionais da área, entretanto o modelo prevê um crescimento maior para o setor do que o estimado pelo SECOVI-SP.

Dessa forma, a metodologia mostrou-se eficiente para o objetivo proposto, encontrando um modelo adequado para a previsão de demanda de imóveis. Desse modo, a combinação e hibridização dos modelos de Box-Jenkins com os modelos MLP/RNA apresentam vantagens que a tornam interessante para realizar a previsão de demanda. Entretanto, algumas considerações quanto a sua utilização merecem ser comentadas.

Os modelos de Box-Jenkins e RNA empregados nesta pesquisa apresentam a vantagem da aplicação em uma ampla gama de séries temporais. Os modelos de Box-Jenkins aplicam-se à previsão de séries com sazonalidade, séries com tendência, séries autocorrelacionadas, além da previsão de séries com mais de uma ou todas as situações descritas. Já os modelos MLP/RNA, além destas, apresentam a capacidade de modelagem não linear. A utilização conjunta de tais modelos, por meio da combinação, possibilita a captura dos diversos possíveis padrões existente na série, os tornando interessantes para a previsão de demanda frente a outros modelos no que tange à sua acurácia.

Desse modo a realização deste estudo permitiu verificar o ganho obtido com as combinações de modelos e também com a hibridização. A dificuldade na utilização prática em organizações, destes modelos e de suas combinações para a previsão de demanda, é devido a disponibilidade de software e conhecimento/compreensão dos modelos de previsão. Sob esta ótica, torna-se necessário a presença de especialistas contratados, o que pode gerar um custo maior ao processo de planejamento. Entretanto, o resultado apresentado nessa dissertação demonstra que a utilização de um modelo de previsão de demanda de imóveis auxilia as incorporadoras e construtoras a tomada de decisão sobre as vendas e também lançamentos, visto que o modelo pode ser usado para verificar qual a tendência de vendas a um médio prazo.

Salienta que o modelo proposto é válido até que não haja nenhuma alteração ou fenômeno na economia nacional, como a criação de outro programa habitacional como o PMCMV.

## 5.2 SUGESTÕES DE PESQUISA FUTURAS

O tema dessa dissertação somado aos resultados auferidos levantaram questionamentos, os quais estimularam desenvolver uma lista de sugestões de novas pesquisas:

- a) Um modelo de regressão defasada, indicando as variáveis que influenciam nas vendas de imóveis em São Paulo;
- b) Um de modelo de Vetor Autorregressivo; indicando as variáveis que influenciam nas vendas de imóveis em São Paulo;
- c) Um modelo combinado entre séries temporais (demanda independente) e um modelo de regressão ou Vetor Autorregressivo (demanda dependente);
- d) Testar o modelo proposto em outras cidades, como Rio de Janeiro;
- e) Utilizar outros moldes de hibridização e comparar com o modelo proposto;
- f) Testar a previsão por componentes separadas da série, prever sazonalidade e tendência distintas; e
- g) Utilizar outras técnicas de combinação, como a lógica fuzzy.



## REFERÊNCIAS

- ABECIP. **Crédito imobiliário com recursos da poupança cresceu 3,4% em 2014**. Reportagem 2015. Disponível em: <<https://www.abecip.org.br/imprensa/abecip-na-midia/credito-imobiliario-com-recursos-da-poupanca-cresceu-3-4-em-2014>>. Acesso em: 15 jan. 2016.
- ABURTO, L.; WEBER, R. Improved supply chain management based on hybrid. **Applied Soft Computing**, n. 7, p. 136 -144, 2007.
- ADYA, M.; COLLOPY, F.; ARMSTRONG, J. S.; KENNEDY, M. Automatic Identification of Time Series Features for Rule-Based Forecasting. **International Journal of Forecasting**, v. 17, n. 2, p. 143-157, 2001.
- ALBANEZ, T. Impact of the cost of capital on the financing decisions of Brazilian companies. **International Journal of Managerial Finance**, v. 11, n. 3, p. 285 - 307, 2015.
- ALLEMÃO, M. A. F. **Redes Neurais Aplicadas à Previsão de Demanda de Numerário em Agências Bancárias**. 2004. 247p. Dissertação (Mestrado em Informática) - Programa de Pós-Graduação em Ciência da Computação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2004.
- ANDRAWIS, R. R.; ATIYA, A. F.; EL-SHISHINY, H. Combination of long term and short term forecasts, with application to tourism demand forecasting. **International Journal of Forecasting**, v. 27, p. 870-886, 2011.
- ANSUJ, A. P.; CAMARGO, M.E.; RADHARAMANAN, R.; PETRY, D.G. Sales forecasting using time series and neural networks. **Computers and Industrial Engineering**, v. 31, n. 1, p. 421-425, 1996.
- ARAÚJO, E. R. **Saiba quais são as expectativas para o mercado imobiliário em 2017**. Reportagem 2016. Disponível em: <<http://www.zappro.com.br/saiba-quais-sao-as-expectativas-para-o-mercado-imobiliario-em-2017/>>. Acesso em: 15 jan. 2017.
- ARAÚJO, M. A. G. **Previsão de demanda de energia elétrica por meio de redes neurais artificiais**. 2005. 112 p. Dissertação (Mestrado em Administração) - Programa de Pós-Graduação em Administração, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2005.
- ARCHER, B. Forecasting Demand: Quantitative and Intuitive Techniques. **International Journal of Tourism Management**, v. 1, n. 1, p. 5-12, 1980.
- ARCHIBALD, B. C.; KOEHLER, A. B. Normalization of seasonal factors in Winters methods. **International Journal of Forecasting**, v. 19, p. 143-148, 2003.
- ARMARY, F. **Venda de imóveis novos em São Paulo aumenta 59,3% no mês de setembro**. Reportagem 2016. Disponível em: <<https://www.jornalestacao.com.br/portal/2016/11/20/venda-de-imoveis-novos-em-sao-paulo-aumenta-593-no-mes-de-setembro/>>. Acesso em: 09 jan. 2017.

- ARMSTRONG, J. S. **Principles of forecasting: handbook for researchers and practitioners**. 1. ed. Boston: Kluwer Academic Publishers, 2001.
- ARMSTRONG, J. S.; COLLOPY, F. Error Measures for Generalizing about Forecasting Methods: Empirical Comparisons. **International Journal of Forecasting**, v. 8, n. 1, p. 69-80, 1992.
- ARMSTRONG, J. S.; COLLOPY, F.; YOKUM, J. T. Decomposition by Causal Forces: A Procedure for Forecasting Complex Time Series. **International Journal of Forecasting**, v. 21, n. 1, p. 25-36, 2005.
- ARMSTRONG, J. S.; FILDES, R. Making progress in forecasting. **International Journal of Forecasting**, v. 22, n. 3, p. 433-441, 2006.
- ARRAES, R. A.; SOUSA FILHO, E. Externalidades e formação de preços no mercado urbano brasileiro. **Economia Aplicada**, v. 12, n. 2, p. 289-319, 2008.
- ARRUDA, F. E.; FERREIRA, R. T.; CASTELAR, I. Modelos lineares e não lineares da curva de Philips para a previsão da Taxa de Inflação no Brasil. **Revista brasileira de Economia**, v. 65, n. 3, p. 237-252, 2011.
- ASLAM, M. S. Comments on Design of fractional adaptive strategy for input nonlinear Box-Jenkins systems. **Signal Processing**, v. 119, p. 169-173, 2016.
- ASSIMAKOPOULOS, E. P. V.; HATZICHRISTOS, T.; FRENCH, N. Real estate appraisal: a review of valuation methods. **Journal of Property Investment & Finance**, v. 21, n. 4, p. 383-401, 2003.
- BABU, C. N.; REDDY, B. E. A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data. **Applied Soft Computing**, v. 23, p. 27-38, 2014.
- BALL, M.; NANDA, A. Household attributes and the future demand for retirement housing, **International Journal of Housing Markets and Analysis**, v. 6, n. 1, p. 45-62, 2013.
- BALLOU, R. H. **Gerenciamento da Cadeia de Suprimentos: Planejamento, Organização e Logística Empresarial**. 4. ed. Porto Alegre: Bookman, 2001.
- BALLOU, R. H. **Gerenciamento da cadeia de suprimentos/Logística empresarial**. 5. ed. Porto Alegre: Bookman, 2006.
- BALTRUSIS, N.; MOURAD, L. N. Política habitacional e locação social em Salvador. **Cadernos CRH**, v. 27, n. 71, p. 267-284, 2014.
- BAPTISTELLA, M. **O uso de redes neurais e regressão linear múltipla na engenharia de avaliações: determinação dos valores venais de imóveis urbanos**. 2005. 123 p. Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Universidade Federal do Paraná, Curitiba, 2005.
- BATES, J. M.; GRANGER, C. W. The Combining of Forecasts. **Operational Research Quarterly**, v. 20, n. 4, p. 451-468, 1969.



- BM&FBOVESPA. **Empresas listadas**. Mar. 2016. Disponível em: <<http://www.bmfbovespa.com.br/ciaslistadas/empresaslistadas/BuscaEmpresaListada.aspx?seguimento=Constru%C3%A7%C3%A3o+Civil&idioma=pt-br>>. Acesso em: 07 mar. 2016.
- BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. **Journal of Econometrics**, v. 31, p. 307-328, 1986.
- BOPP, A. E. On Combining Forecasts: Some Extensions and Results. **Management Science**. v. 31, n. 12, p. 1492-1498, 1985.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis – forecasting and control**. 4. ed. New Jersey: Prentice Hall, 2008.
- BOX, G. E. P.; LUCENÑO, A. **Statistical control by monitoring and feedback adjustment**. 1. ed. New York: John Wiley & Sons, 1997.
- BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. **Redes Neurais Artificiais: Teoria e Aplicações**. 2. ed. Rio de Janeiro: Ed. LTC – Livros Técnicos e Científicos, 2011, 226 p.
- BRASIL. Ministério das Cidades. **Política Nacional de Habitação**. Brasília: Ministério das Cidades, 2004. Disponível em: <[http://www.sst.sc.gov.br/arquivos/id\\_submenu/230/politicanacionalhabitacao.pdf](http://www.sst.sc.gov.br/arquivos/id_submenu/230/politicanacionalhabitacao.pdf)>. Acesso em: 18 jan. 2016.
- BRAUER, K. U. **Einführung in die immobilienwirtschaft: Grundlagen der Immobilienwirtschaft**. 7. ed. Wiesbaden: Gabler, 2011.
- BRAULIO, S. N. **Proposta de uma metodologia para avaliação de imóveis urbanos baseados em métodos estatísticos multivariados**. 2005. 158 p. Dissertação (Mestrado em Métodos Numéricos em Engenharia) - Universidade Federal do Paraná, Curitiba, 2005.
- BREM, A.; MAIER, M.; WIMSCHEIDER, C. Competitive advantage through innovation: the case of Nespresso. **European Journal of Innovation Management**, v. 19, n. 1, p. 133-148, 2016.
- BROOKS, C.; TOSLACOS, S. **Real Estate Modelling and Forecasting**. 1 ed. Cambridge City: Cambridge University Press, 2014. 453 p.
- BUSINGER, M. P.; READ, R. R. Identification of Demand Patterns for Selective Processing: A Case Study. **Omega**, v. 27, n. 2, p. 189-200, 1999.
- CALDEIRA, J. F.; MOURA, G. V.; SANTOS, A. A. P.; TESSARI, C. Seleção de carteiras com modelos fatoriais heterocedásticos: aplicação para fundos de fundos multimercados. **Revista de Administração da Mackenzie**, v. 15, n. 2, p. 127-161, 2014.
- CANDIDO, B. L. **Mercado Imobiliário: uma análise sobre o comportamento dos preços dos imóveis na cidade de São Paulo**. 2012. 62 f. Monografia (Graduação em Ciências

- Econômicas). Departamento de Ciências Econômicas da Universidade Federal de Santa Catarina, Florianópolis, 2012.
- CECATTO, C.; BELFIORE, P. O uso de métodos de previsão de demanda nas indústrias alimentícias brasileiras. **Gestão e Produção**, v. 22, n. 2, p. 404-418, 2015.
- CHANG, P.; LUI, C.; FAN, C. Data clustering and fuzzy neural network for sales forecasting: A case study in printed circuit board industry. **Knowledge-Based Systems**, v. 22, p. 344-355, 2009.
- CHARNET, R.; FREIRE, C. A.; CHARNET, E. R.; BONVINO, H. **Análise de modelos de regressão linear: com aplicações**. Campinas: Editora Unicamp, 2008.
- CHEN, J.; YANG, J.; ZHAO, J.; XU, F.; SHEN, Z.; ZHANG, L. Energy demand forecasting of the greenhouses using nonlinear models based on model optimized prediction method. **Neurocomputing**, v. 174, part B, p. 1087-1100, 2016.
- CLEMENTS, M. P.; FRANSES, P. H.; SWANSON, N. R. Forecasting economic and financial time-series with non-linear models. **International Journal of Forecasting**, v. 20, p. 169-183, 2004.
- COELHO, L. S.; SANTOS, A. A. P.; COSTA JUNIOR, N. C. A. da. Podemos prever a taxa de cambio brasileira? Evidência empírica utilizando inteligência computacional e modelos econométricos. **Gestão e Produção**, v.15, n. 3, p. 635-647, 2008.
- CSESKO, M.; REED, R. Will residential aged care facilities meet long-term demand?. **Property Management**, v. 27, n. 1, p. 58-74, 2009.
- DAVIS, M.; AQUILANO, N.; CHASE, R. **Fundamentos da Administração da Produção**. Porto Alegre: Bookman, 2001.
- DE GOOIJER, J. G.; KUMAR, K. Some recent developments in non-linear time series modelling, testing, and forecasting. **International Journal of Forecasting**, v. 8, p. 135-156, 1992.
- DE LA PAZ, P. T. New housing supply and price reactions: evidence from Spanish markets. **Journal of European Real Estate Research**, v. 7, n. 1, p. 4-28, 2014.
- DENT JR., H. S. **A próxima grande depressão: como se preparar em tempos de turbulência**. 1 ed.. Rio de Janeiro: Elsevier. 2010.
- DURBIN, J.; KOOPMAN, S. J. **Time series analysis by state space methods**. Oxford: Oxford University Press, 2008.
- DUUS, H. J. Strategic forecasting: the management perspective. **Management Research Review**, v. 39, n. 9, p. 998-1015, 2016.
- ELLERO, A.; PELLEGRINI, P. Are traditional forecasting models suitable for hotels in Italian cities?. **International Journal of Contemporary Hospitality Management**, v. 26, n. 3, p. 383-400, 2014.

EL PAÍS. **Mercado imobiliário vive seu ajuste e abre espaço para barganha em 2015**. São Paulo, 1 Abr. 2015. Disponível em: <[http://brasil.elpais.com/brasil/2015/04/01/politica/1427914453\\_617511.html](http://brasil.elpais.com/brasil/2015/04/01/politica/1427914453_617511.html)>. Acesso em: 17 jan. 2016.

EMBRAESP. **Crise estrutural no setor imobiliário ou a onde que se ergueu no mar**. São Paulo, jul. 2013. Disponível em: <<https://incorporacaoimobiliaria.com/tag/embraesp/>>. Acesso em: 26 jan. 2016.

ENDERS, W. **Applied econometric time series**. 3. ed. New York: John Wiley & Sons, 1995.

ENGLE, R. F. Autoregressive conditional heteroskedasticity with estimates of the variance of UK onflation. **Econometrica**, v. 50, p. 987-1008, 1982.

ERAKER, B.; WANG, J. A non-linear dynamic model of the variance risk premium. **Journal of Econometrics**, v. 187, n. 2, p. 547-556, 2015.

FARUK, D. O. A hybrid neural network and ARIMA model for water quality time series prediction. **Engineering Applications of Artificial Intelligence**, v. 23, n. 4, p. 586-594, 2009.

FERNANDEZ, V. Linear and non-linear causality between price indices and commodity prices. **Resources Policy**, v. 41, p. 40-51, 2014.

FERREIRA, T. A. E. **Uma nova metodologia híbrida inteligente para a previsão de séries temporais**. 2006. 206 f. Tese (Doutorado Ciência da Computação)- Universidade Federal de Pernambuco, Recife, 2006.

FLORES, J. H. F. **Comparação de modelos MLP/RNA e modelos Box-Jenkins em séries temporais não lineares**. 2009. 133 p. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.

FREITAS, C. A. de; SAFADI, T. Volatilidade dos Retornos de Commodities Agropecuárias Brasileiras: um teste utilizando o modelo APARCH. **Revista de Economia e Sociologia Rural**, v. 53, n. 2, p. 211-228, 2015.

GAITHER, N.; FRAZIER, G. **Administração da produção e operações**. 8. ed. São Paulo: Thomson, 2004.

GARDNER, E. S.; MCKENZIE, E. Forecasting trends in time series. **Management Science**, v. 31, p. 1237–1246, 1985.

GARDNER, E. S. Exponential smoothing: the state of the art – Part II. **International Journal of Forecasting**, v. 22, p.637-666, 2006.

GIBLER, K. M.; CLEMENTS, J. S. Testing a forecast model to predict movement of older Americans into retirement housing. **International Journal of Housing Markets and Analysis**, v. 4, n. 1, p. 18-30, 2011.

- GIBLER, K. M.; TALTAVULL, P.; CASADO-DÍAZ, J. M.; RODRIGUEZ, V. Examining retirement housing preferences among international retiree migrants. **International Real Estate Review**, v. 12, n. 1, p. 1-22, 2009.
- GIL, A. C. **Como elaborar projeto de pesquisa**. 4. ed. São Paulo: Atlas, 2009.
- GIOVANNELLI, A.; PROIETTI, T. On the Selection of Common Factors for Macroeconomic Forecasting. **Advances in Econometrics**, v. 35, p. 593-628, 2016.
- GONÇALVES, P. S. **Administração de Materiais**. 2. ed. Rio de Janeiro: Elsevier, 2007.
- GONZÁLEZ, M. A. S.; FORMOSO, C. T. Mass appraisal with genetic fuzzy rule based systems. **The Journal of European Real**, v. 24, p. 20-3, 2003.
- GRANGER, C. W. J.; RAMANATHAN, R. Improved methods of combining forecasts. **Journal of Forecasting**, v. 3, p. 197-204, 1984.
- GUJARATI, D. N.; PORTER, D. C. **Econometria básica**. 5. ed. Porto Alegre: Bookman, 2011.
- GUO, H.; LIU, X.; SUN, Z. Multivariate time series prediction using a hybridization of VARMA models and Bayesian networks. **Journal of Applied Statistics**, v. 43, n.16, p. 2897-2909. 2006.
- HAJEJ, Z.; REZG, N.; ALI, G. An optimal production/maintenance strategy under lease contract with warranty periods. **Journal of Quality in Maintenance**, v. 22, n. 1, p. 35-50, 2016
- HANSSON, S. O.; LILIEWVIST, K; BJORNBERG, K. E.; JOHANSSON, M. V. Time horizons and discount rates in Swedish environmental policy: Who decides and on what grounds? **Futures**, v. 76, p. 55-66, 2015.
- HAYKIN, S.. **Redes Neurais: Princípios e práticas**. 2 ed. Porto Alegre: Bookman., 2001.
- HENKE, J. E.; WICHERN, D. **Business forecasting**. 9. ed. New Jersey: Prentice Hall, 2009.
- HIBON, M.; EVGENIOU, T. To combine or not combine: selecting among forecasts and their combinations. **International Journal of Forecasting**, v. 21, p. 15-24, 2005.
- HICHAM, A.; MOHAMED, B.; ABDELLAH, EL F. An improved approach based on fuzzy clustering and Back-Propagation Neural Networks with adaptive learning rate for sales forecasting: Case study of PCB industry. **International Journal of Computer Science**, v. 9 , n. 1, p. 404-414. 2012.
- HO, D. K. H.; ADDAE-DAPAH, K. Real estate market cyclical dynamics: The prime office sectors of Kuala Lumpur, Singapore and Hong Kong. **International Journal of Managerial Finance**, v. 10, n. 2, p. 241- 262, 2014.

HOFER, P.; EISL, C.; MAYR, A. Forecasting in Austrian companies: Do small and large Austrian companies differ in their forecasting processes? **Journal of Applied Accounting Research**, v. 16, n. 3, p. 359- 382, 2015.

HOLT, C. C. Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages. **International Journal of Forecasting**, v. 20, n. 1, p. 5-10, 2004.

HOMER. **Mercado imobiliário em 2017 deve ganhar fôlego. Veja balanço de 2016.** Reportagem 2016. Disponível em: <<http://homer.com.br/blog/2016/12/01/mercado-imobiliario-em-2017-deve-ganhar-folego-veja-balanco-de-2016/>>. Acesso em: 16 jan. 2017.

HUANG, Y.; WANG, Y.; GAI, S. 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**, v. 15, p. 4532-4537, 2011.

HUANG, Z.; LIU, H.; WANG, T. Modeling long memory volatility using realized measures of volatility: A realized HAR GARCH model. **Economic Modelling**, v. 52, part B, p. 812-821, 2016.

HYNDMAN, R. J., KOEHLER, A. B., SNYDER, R. D, GROSE, S. A state space framework for automatic forecasting using exponential smoothing methods. **International Journal of Forecasting**, v. 18, p. 439–454, 2002.

HYUNG, C. J.; JUNG, J. U. Simulation Model on Long-Term Housing Demand Forecasting for the Utilization of Real Estate Knowledge. **Journal of CEO and Management Studies**, v. 16, p. 119-139, 2013.

IBGE. **Pesquisa anual dos serviços.** Rio de Janeiro, 2014. Disponível em: <<http://ibge.gov.br/home/estatistica/economia/comercioeservico/pas/analisepas99.shtm>>. Acesso em 02 dez. 2015.

IBGE. **Cidades.** Rio de Janeiro, 2015. Disponível em: <<http://www.cidades.ibge.gov.br/xtras/perfil.php?lang=&codmun=355030&search=sao-paulo|sao-paulo>>. Acesso em 07 dez. 2015.

IBGE. **Série histórica IPCA.** Rio de Janeiro, jan. 2016. Disponível em: <[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)>. Acesso em: 27 jan. 2016.

IPEA. **Existe bolha no mercado imobiliário brasileiro?** Textos para discussão. São Paulo, SP. Ago., 2012

JALES, A. W. L.; SILVA, C. A. U. da. A estimação neural de tempos de viagens de ônibus sob regime de fretamento usando-se de dados de posicionamento por satélites (GPS). **Journal of Transport Literature**, v. 10, n. 1, p. 30-34, 2016.

JOZEFczyk, J.; GASIOR, D. Utility-based models and decision making problems for selected network processes. **Kybernetes**, v. 44, n. 6/7, p.1107-1121, 2015.

- KAHN, K. Benchmarking Sales Forecasting Performance Measures. **The Journal of Business Forecasting Methods and Systems**, v. 17, n. 4, p. 19-23, 1998.
- KAHN, K. An Exploratory Investigation of New Product Forecasting Practices. **The Journal of Product Innovation Management**, v. 19, n. 2, p. 133-143, 2002.
- KATRIS, C.; DASKALAKI, S. Comparing forecasting approaches for internet traffic. **Expert Systems with Applications**, v. 42, n. 21, p. 8172-8183, 2015.
- KAYA, G. O.; DEMIREL, O. F. Parameter optimization of intermittent demand forecasting by using spreadsheet. **Kybernetes**, v. 44, n. 4, p. 576-587, 2015.
- KHASHEI, M.; BIJARI, M. An artificial neural network (p,d,q) model for time series forecasting. **Expert Systems with applications**, v. 37 n. 1, p. 479-489, 2010.
- KLINDOKMAI, S.; NEECH, P.; WU, Y.; OJIAKO, U.; CHIPULU, M.; MARSHALL, A. Evaluation of forecasting models for air cargo. **The International Journal of Logistics Management**, v. 25, n. 3, p. 635- 655, 2014.
- KOEHLER, A. B.; CHATFIELD, C.; ORD, J. K.; SNYDER, R. D. A new look at models for exponential smoothing. **The Statistician**, v. 50, p. 147-159, 2001.
- KOH, Y.; LEE, H. A. The effect of financial factors on firms' financial and tax reporting decisions. **Asian Review of Accounting**. v. 23, n. 2, p.110 - 13, 2015.
- KOTLER, P.; ARMSTRONG, G. **Princípios de marketing**. 12. ed. São Paulo: Pearson, 2009.
- KOVÁCS, Z. L. **Redes Neurais Artificiais: Fundamentos e aplicações**. 4. ed. São Paulo: Ed. Livraria da Física, 2006.
- KRAJEWSKI, L. J.; RITZMAN, L. P. **Operations Management: Strategy and Analysis**. 6. ed., New Jersey: Prentice-Hall, 2002.
- KRISTJANPOLLER, W.; MINUTOLO, M. C. Gold price volatility: a forecasting approach using the artificial neural network-GARCH model. **Expert Systems with Applications**, v. 42, n. 20, p. 7245 - 7251, 2015.
- LAWRENCE, M. J.; EDMUNDSON, R. H.; O'CONNOR, M. J. The Accuracy of Combining Judgemental and Statistical Forecasts. **Management Science**. v. 32, n. 12, p. 1521-1532, 1986.
- LEE, Y.S.; TONG, L.I. Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming. **Knowledge-Based Systems**, n. 24, p. 66-72, 2011.
- LEMOS, F. O. **Metodologia de seleção de métodos de previsão de demanda**. 2006. 183 p. Dissertação (Mestrado em Engenharia de Produção) - Departamento de Engenharia de Produção e Transportes, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2000.

LINDBERG, E; ZACKRISSON, U. Deciding about the Uncertain: The Use of Forecasts as an Aid to Decision-making. **Scandinavian Journal of Management**. v. 7, n. 4, p. 271-283, 1991.

LIU, Lv.; LI, Wang. Research on the forecast of Real Estate market demand in Beijing. In: INTERNATIONAL CONFERENCE ON CONSTRUCTION & REAL ESTATE MANAGEMENT, 2009, Beijing, China. **Anais...** Beijing, ICCRSM, 2009. Disponível em: < [https://apps.webofknowledge.com/full\\_record.do?product=UA&search\\_mode=GeneralSearch&qid=2&SID=4DI2OnoqkXHK1AlrmqM&page=1&doc=1](https://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=2&SID=4DI2OnoqkXHK1AlrmqM&page=1&doc=1) >. Acesso em: 03 jan. 2016.

LUSTOSA, L.; MESQUITA, M. A.; QUELHAS, O.; OLIVEIRA, R. **Planejamento e Controle da Produção**. Rio de Janeiro: Elsevier, 2008.

LYDIA, M.; KUMAR, S. S.; KULMAR, S. I. A.; KUMAR, G. E. P. Linear and non-linear autoregressive models for short-term wind speed forecasting. **Energy Conversion and Management**, v. 112, p. 115-124, 2016.

LYNN, G.; SCHNAARS, S.; SKOV, R. Survey of New Product Forecasting Practices in Industrial High Technology and Low Technology Businesses. **Industrial Marketing Management**. v. 28, n. 6, p. 565-571, 1999.

MADDALA, G. S. **Introdução à Econometria**. 3 ed. Rio de Janeiro: Ed: LTC – Livros Técnicos e Científicos, 2003.

MAKRIDAKIS, S.; WHEELWRIGHT, S.; HYNDMAN, R. **Forecasting: Methods and Applications**. 3. ed., New York: John Wiley & Sons, 1998.

MAKRIDAKIS, S.; HIBON, M. The M3-competition: results, conclusions and implications. **International Journal of Forecasting**, v. 16, p.451-476, 2000.

MARTINS, L. P. HR leaders hold the key to effective diversity management: ... as more and more important decisions are taken at local level. **Human Resource Management International Digest**, v. 23, n. 5, p. 49-53, 2015.

MARTINS, P. G.; LAUGENI, F. P. **Administração da produção**. 2. ed. São Paulo: Saraiva, 2005.

MARTINS, V. L. M. **Combinações de Previsões: Uma proposta utilizando análise de componentes principais**. 105 f. Tese (Doutorado em Engenharia de Produção) Universidade Federal do Rio Grande do Sul, Porto Alegre, 2014.

MARTINS, V. L. M.; WERNER, L. Comparação de previsões individuais e suas combinações: um estudo com séries industriais. **Production**, v. 24, n. 3, p. 618-627, 2014.

MATOS, D.; BARTKIW, P. I. N. **Introdução ao mercado imobiliário**. 1. ed. Curitiba: IFPR-E-tec, 2013.

MEDEIROS, F. S. B. **Análise comparativa entre a aquisição dos imóveis financiados e o valor dos aluguéis**: Um estudo do Programa Minha Casa Minha Vida. 2015, 95 f.

- Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Maria, Santa Maria, 2015.
- MENDONÇA, M. J. C. O crédito imobiliário no Brasil e sua relação com a Política Monetária. **Revista brasileira de Economia**, Rio de Janeiro, v. 67, n. 4, p. 457-495, 2013.
- MENDONÇA, M. J.; MEDRANO, L. A.; SACHSIDA, A. **Avaliando o efeito de um choque de política monetária sobre o mercado imobiliário**. Brasília, IPEA, 2011.
- MENTZER, J. T.; GOMES, R. Evaluating a Decision Support Forecasting System. **Industrial Marketing Management**. v. 18, n. 4, p. 313-323, 1989.
- MEYER, J. F. P. **Demanda Residencial: adequação da análise de mercado imobiliário - o caso de São Paulo**. 2008. 212 f. Tese (Doutorado em Arquitetura e Urbanismo). Faculdade de Arquitetura e Urbanismo, Universidade de São Paulo, São Paulo, 2008.
- MILLER, N. G. Workplace trends in office space: implications for future office demand. **Journal of Corporate Real Estate**, v. 16, n. 3, p. 159-181, 2014.
- MONTGOMERY, D.; JOHNSON, L.; GARDINER, J. **Forecasting and Time Series Analysis**. New York: McGraw-Hill, 1990.
- MONTGOMERY, D. C.; RUNGER, G. C. **Applied Statistics and Propability for engineers**. New York: John Wiley & Sons, Inc., 1994.
- MONTGOMERY, D. C.; HINES, W. W.; GOLDSMAN, D. M.; BORROR, C. M. **Probabilidade e Estatística na Engenharia**. Rio de Janeiro: LTC, 2006.
- MOON, M; MENTZER, J.; SMITH, C.; GARVER, M. Seven Keys to Better Forecasting. **Business Horizons**. v. 41, n. 5, p. 44-52, 1998.
- MOORE, D.S.; McCABE, G.P.; DUCKWORTH, W.M.; SCLOVE, S.L. **A prática da estatística empresarial: como usar dados para tomar decisões**. 1. ed. Rio de Janeiro: Ed. LTC, 2006.
- MOREIRA, D. A. **Administração da Produção e Operações**. São Paulo: Cengage Learning, 2011.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2. ed. São Paulo: Bulcher, 2006.
- MORO, M. F.; COMPAGNONI, C .C. dos R.; FLORES, S. A.; WEISE, A. D.; VINCENZI, S. L. São Paulo Real Estate market: Interference of economic indicators in sales of residential Real Estates. **International Journal of Real Estate Studies**, v. 10, n. 2, 2016.
- OLIVEIRA, B. C. **Qualidade Habitacional do novo milênio**. 2011. 118 p. Dissertação (Mestrado Integrado em Engenharia Civil) - Departamento de Engenharia Civil, Faculdade de Engenharia da Universidade do Porto, Porto, Portugal, 2011.



ORD, L. K., KOEHLER, A. B., SNYDER, R. D. Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models. **Journal of the American Statistical Association**, v.92, n. 440, p. 1621 -1629, 1997.

OSADCHIY, N., GAUR, V.; SESHADRI, S. Sales forecasting with financial indicators and experts input. **Production & Operations Management**, v. 22, n. 5, p. 1056-1076, 2013.

OSMA, Y. Z.; ABDELLATIF, M. E. Improving accuracy of downscaling rainfall by combining predictions of different statistical downscale models. **Water Science**, v. 30, n. 2, p. 61-75, 2016.

PAGOURTZI, E.; MAKRIDAKIS, S.; ASSIMAKOPOULOS, V.; LITSA, A. The advanced forecasting information system PYTHIA: An application in real estate time series. **Journal of European Real Estate Research**, v. 1, n. 2, p.114-138, 2008.

PATTON, A. J.; SHEPPARD, K. Optimal combinations of realised volatility estimators. **International Journal of Forecasting**, v. 25, p. 218-238, 2009.

PAULANI, L. M. A crise do regime de acumulação com dominância da valorização financeira e a situação do Brasil. **Estudos Avançados**. v. 23, n. 66, p. 25-39, 2009.

PEGELS, C. Exponential forecasting: Some new variations. **Management Science**, v. 15, n. 3, 1969.

PELLEGRINI, F. R. **Metodologia para implementação de sistemas de previsão de demanda**. 2000. 146p. Dissertação (Mestrado em Engenharia de Produção) - Departamento de Engenharia de Produção e Transportes, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2000.

PELLEGRINI, F. R.; FOGLIATTO, F. S. Passos para implantação de sistemas de previsão de demanda – Técnicas e estudo de Caso. **Revista Produção**, v. 11, n. 1, p. 43-64, 2001.

PINDYCK, R. S.; RUBINFELD, D. L. **Microeconomia**. 7. ed. São Paulo: Pearson Education do Brasil, 2010.

POKHILCHUK, K. A.; SAVEL'EV, S. E. On the choice of GARCH parameters for efficient modelling of real stock price dynamics. **Physica A: Statistical Mechanics and its Applications**, v. 448, p. 248-253, 2016.

PRAHLADA, R.; DEKA, P. C. Forecasting of time series significant wave height using wavelet decomposed Neural Network. **Aquatic Procedia**, v. 4, p. 540-547, 2015.

RAGSDALE, C. T. **Modelagem e Análise de Decisão**. São Paulo: Cengage Learning, 2009.

RASMUSSEN, R. On Time Series Data and Optimal Parameters. **Omega**, v. 32, n. 2, p. 111-120, 2004.

RINGUEST, J. L.; TANG, K. Simple Rules for Combining Forecasts: Some Empirical Results. **Socio-Economic Planning Science**, v. 21, n. 14, p. 239-243, 1987.

RITZMAN, L. P.; KRAJEWSKI, L. J. **Administração da produção e operações**. São Paulo: Pearson Prentice Hall, 2008.

ROCHA LIMA JR, J.; MONETTI, E.; ALENCAR, C. T. **Real Estate: Fundamentos para Análise de Investimentos**. São Paulo: Elsevier, 2011.

ROSS, S. M. **Introduction to probability models**. 5. ed. San Diego: Academic Press, 1993.

SAMOHYL, R. W.; SOUZA, G.; MIRANDA, R. **Métodos Simplificados de Previsão Empresarial**. Rio de Janeiro: Editora Ciência Moderna, 2008.

SATO, R. C. Avaliação do risco no mercado de ações de companhias aéreas. **Journal of Transport Literature**, v. 7, n. 2, p. 352-372, 2013.

SECOVI-SP. **Balanco anual de 2014**: inúmeros fatores afetaram o mercado imobiliário em 2014. São Paulo, nov. 2015 C. Disponível em: <<http://balanco.secovi.com.br/2014/>>. Acesso em: 28 jan. 2015.

SECOVI-SP. **Vendas de imóveis apresentam bons resultados em novembro de 2015**. São Paulo, nov. 2015A. Disponível em: <<http://www.secovi.com.br/files/Arquivos/pmi-novembro-2015.pdf>>. Acesso em: 20 jan. 2016.

SECOVI-SP. **Na contramão do mercado imobiliário, VCG cresce 155% em 2014**. São Paulo, nov. 2015B. Disponível em: <<https://incorporacaoimobiliaria.com/tag/mercado-imobiliario/>>. Acesso em: 25 jan. 2016.

SECOVI-SP. **Associados**. São Paulo, fev. 2016A. Disponível em: <<http://www.secovi.com.br/associados/page/55/>>. Acesso em: 14 fev. 2016.

SECOVI-SP. **Outubro fraco confirma fraco desempenho de vendas de imóveis em 2015**. São Paulo, nov. 2015C. Disponível em: <<http://www.secovi.com.br/noticias/outubro-confirma-fraco-desempenho-de-vendas--de-imoveis-em-2015/10965/>>. Acesso em: 07 jan. 2016.

SECOVI-SP. **Série histórica com a evolução do IVV e também do número de unidades ofertadas, vendidas e lançadas em São Paulo**. São Paulo, 2016B. EXCEL.

SHEN, Y; GUO, J.; LIU, X.; WEI, X.; LI, W. One hybrid model combining singular spectrum analysis and LS + ARMA for polar motion prediction. **Advances in Space Research**, v. 59, n. 2, p. 513-523, 2017.

SHIMBO, L. O aquecimento imobiliário e o mercado de habitação na área central de São Paulo (2001-2010). **Revista Latinoamericana de Estudios Urbanos e Regionales**, Santiago, v. 39, n.117, p. 215-235, 2013.

SILVA, H. M. B. **Observatório do uso do solo e da gestão fundiária do centro de São Paulo**. Relatório de pesquisa. São Paulo: Laboratório de Habitação e Assentamentos Humanos (LABHAB), Universidade de São Paulo (USP), 2006. Disponível em: <[http://www.usp.br/fau/deprojeto/labhab/biblioteca/produtos/observatorio\\_pcentro.pdf](http://www.usp.br/fau/deprojeto/labhab/biblioteca/produtos/observatorio_pcentro.pdf)>. Acesso em: 04 jan. 2016.

SILVA, C. F. DA; ALVES, T. W. Dinâmica dos financiamentos habitacionais nos municípios do Rio Grande do Sul de 2006 a 2010: uma avaliação do Programa “Minha Casa, Minha Vida”. **Revista de Administração Pública**, Rio de Janeiro, v. 48, n. 1, p. 27-54, 2014.

SIQUEIRA, Luis Paulo Ribeiro. A indústria da corretagem no segmento de imóveis usados: características e Perspectivas futuras. In: CONFERÊNCIA INTERNACIONAL DA LARES. XIII, 2013, São Paulo, **Anais...** São Paulo, LARES, 2013. Disponível em: <<http://lares.org.br/Anais2013/artigos/806-1047-1-RV.pdf>>. Acesso em: 07 jan. 2016.

SIQUEIRA, M. L. **Modelos de Séries Temporais para a Previsão da Arrecadação Tributária Federal**. 2002. 220p. Dissertação (Mestrado em Economia). Universidade Federal de Pernambuco, Recife, 2002.

SLACK, N.; CHAMBERS, S.; JOHNSTON, R. **Administração da Produção**. São Paulo: Atlas, 2009.

SKITTIDES, C.; FRUH, W. Wind forecasting using Principal Component Analysis. **Renewable Energy**, v. 69, p. 365-374, 2014.

SOBREIRO, V. A.; ARAÚJO, P. H. S. L.; NAGANO, M. S.. Precificação do etanol utilizando técnicas de redes neurais artificiais. **Revista de Administração da USP**, v. 44, n. 1, p. 46-58, 2009.

SOLA, L. F. **Condições para estruturação de fundos imobiliários com renda proveniente de locações residenciais**. 2014. 216 p. Monografia (MBA em Economia Setorial e Mercados, com ênfase em Real Estate). Escola Politécnica da Universidade de São Paulo, São Paulo, 2014.

SOLA, Luis Felipe; PORTO, Paola Torneri Condições para estruturação de fundos imobiliários com renda proveniente de locações residenciais. In: CONFERÊNCIA INTERNACIONAL DA LARES , XIV, 2014, Rio de Janeiro, **Anais...** Rio de Janeiro, LARES, 2014. Disponível em: <<http://lares.org.br/Anais2014/artigos/1061-1261-1-DR.pdf>> Acesso em: 17 dez. 2015.

SOUZA, F. M. **Modelos Box & Jenkins aplicados a demanda de leitos hospitalares**. 2006. 83 p. Monografia (Especialização em Estatística e Modelagem Quantitativa) - Universidade Federal de Santa Maria, Santa Maria, RS, Brasil, 2006.

STAUDT, F.; GONCALVES, M. B.; RODRIGUEZ, C. M. T. Procedimento para implantar um modelo de previsão de demanda com incorporação de julgamento de especialistas. **Production**, Ahead to print, Epub Nov 10, 2015.

STOCK, J. H.; WATSON, M. Combination forecasts of output growth in a seven-country data set. **Journal of Forecasting**, v. 23, p.405-430, 2004.

TAYLOR, J. W. Exponential Smoothing with a Damped Multiplicative Trend. **International Journal of Forecasting**, v. 19, n. 4, p. 715 -725, 2003.

TAYLOR, J. W.; SNYDER, R. Forecasting Intraday Time Series with Multiple Seasonal Cycles Using Parsimonious Seasonal Exponential Smoothing. **Omega**, v. 40, n. 6, p. 748–757, 2012.

TERRA, L. A. A.; PASSADOR, J. L. Redes neurais artificiais na previsão da inflação: aplicação como ferramenta de apoio à análise de decisões financeiras em organizações de pequeno porte. **Revista de Administração da Mackenzie**, v.13, n.1, p. 68-86, 2012.

THEIL, H. **Applied economic forecasting**. 4. ed. Amsterdam: North-Holland Publishing Company, 1966.

THOMAS JR, G. B.; FINNEY, R. L. **Calculus and analytic geometry**. 8. ed. Addison-Wesley, Reading, MA, 1992.

TONG, H. **Non-linear time series models**. Oxford: Oxford University Press, 1990.

TSAY, R. S. **Analysis of Financial time series**. New York: Wiley, 2002.

TUBINO, D. F. **Manual de planejamento e Controle da Produção**. 2. ed. São Paulo: Atlas, 2004.

VANCLAY, J. K. Planning horizons and end conditions for sustained yield studies in continuous cover forests. **Ecological Indicators**, v. 8, p. 436-439, 2015.

VAN DONSELAAR, K. H.; PETERS, J.; DE JONG, A.; BROEKMEULEN, R. A. C. M. Analysis and forecasting of demand during promotions for perishable items. **International Journal of Production Economics**, v. 172, p. 65-75, 2015.

VARANDAS JÚNIOR, J. E. R. **Interpretação da influência das variáveis condicionantes da demanda pela produção habitacional privada: aplicação na cidade de São Paulo durante o período de 1998 a 2008**. 2010. 129 p. Dissertação (Mestrado em Engenharia). Departamento de Engenharia e Construção Civil. Escola Politécnica da Universidade de São Paulo, São Paulo, 2011.

VARANDAS JÚNIOR, J. E. R.; ALENCAR, C. T. de. Modelos de projeção de demanda aplicado ao mercado imobiliário. In: CONFERÊNCIA INTERNACIONAL DA LARES, IX, 2009, São Paulo. **Anais...** São Paulo, LARES, 2009. Disponível em: <<http://lares.org.br/2009/images/154-323-1-RV.pdf>>. Acesso em: 15 dez. 2016.

VICINI, L.; SOUZA, A. M. **Análise multivariada da teoria à prática**. Santa Maria: UFSM, CCNE, 2005.

VORTELINOS, D. I. Forecasting realized volatility: HAR against Principal Components Combining. neural networks and GARCH. **Research in International Business and Finance**, v. 39, part B, p. 824-839, 2017.

WALKER, K. B.; McCLELLAND, L. A. Management Forecasts and Statistical Prediction Model Forecasts in Corporate Budgeting. **Journal of Accounting Research**, v. 29, n. 2, p. 371-381, 1991.

WEISE, A. D. **Especulação no mercado imobiliário residencial: a avaliação do mercado por meio de índices**. 2009. 277 f. Tese (Doutorado em Engenharia Civil) - Universidade Federal de Santa Catarina, Florianópolis, 2009.

WINTERS, P. R. Forecasting sales by exponentially weighted moving averages. **Management Science**, v. 6, n. 3, p. 324-342, 1960.

WU, G.; ZHANG, H.; CHEN, P. Fitting and forecasting of nonlinear Taylor rule in China: Review on the roles of foreign exchange in monetary policies. **China Finance Review International**, v. 5 n. 4, p. 402-420, 2015.

YAO, Cy. The forecast analysis of developing trend about supply demand factors for Beijing real estate market in 2010. In: INTERNATIONAL SYMPOSIUM ON ADVANCEMENT OF CONSTRUCTION MANAGEMENT AND REAL ESTATE, 2009, Nanjing, **Anais...** Nanjing, 2009. Disponível em: <[http://apps.webofknowledge.com/full\\_record.do?product=UA&search\\_mode=GeneralSearch&qid=12&SID=4D12OnoqkXHK1AlrmqM&page=1&doc=1](http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=12&SID=4D12OnoqkXHK1AlrmqM&page=1&doc=1)>. Acesso em: 07 jan. 2016.

ZHANG, G. P. Time series forecasting using a hybrid ARIMA and neural network model. **Neurocomputing**, n. 50, p. 159-175, 2003.

ZHANG, G.; PATUWO, B. E.; HU, M.Y. Forecasting with artificial neural networks: the state of the art. **International Journal of Forecasting**, v. 14, n. 1, p.35-62, 1998.

ZOU, H. F.; XIA, G. P.; YANG, F. T.; WANG, H. Y. An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. **Neurocomputing**, n.70, p.2913-2926, 2007.

ZYLBERSTAJN, E. **Saiba quais são as expectativas para o mercado imobiliário em 2017**. Reportagem 2016. Disponível em: <<http://www.zappro.com.br/saiba-quais-sao-as-expectativas-para-o-mercado-imobiliario-em-2017/>>. Acesso em: 15 jan. 2017.

**ANEXO A - TABELA DE FORMULAÇÃO DOS MODELOS DE SUAUIZAÇÃO EXPONENCIAL POR MEIO DE MODELOS DE ESPAÇO DE ESTADO**

Tendência	Sazonalidade					
	N		A		M	
	Erro		Erro		Erro	
	A	M	A	M	A	M
	$e_t = y_t - \mu_t$	$e_t = (y_t - \mu_t)/\mu_t$	$e_t = y_t - \mu_t$	$e_t = (y_t - \mu_t)/\mu_t$	$e_t = y_t - \mu_t$	$e_t = (y_t - \mu_t)/\mu_t$
N	$\mu_t = l_{t-1}$ $l_t = \mu_t + \alpha e_t$	$\mu_t = l_{t-1}$ $l_t = \mu_t(1 + \alpha e_t)$	$\mu_t = l_{t-1} + s_{t-m}$ $l_t = (\mu_t + \alpha e_t) - s_{t-m}$ $s_t = (\mu_t + \gamma e_t) - l_{t-1}$	$\mu_t = l_{t-1} + s_{t-m}$ $l_t = \mu_t(1 + \alpha e_t) - s_{t-m}$ $s_t = \mu_t(1 + \gamma e_t) - l_{t-1}$	$\mu_t = l_{t-1}s_{t-m}$ $l_t = (\mu_t + \alpha e_t)/s_{t-m}$ $s_t = (\mu_t + \gamma e_t)/l_{t-1}$	$\mu_t = l_{t-1}s_{t-m}$ $l_t = \mu_t(1 + \alpha e_t)/s_{t-m}$ $s_t = \mu_t(1 + \gamma e_t)/l_{t-1}$
A	$\mu_t = l_{t-1} + b_{t-1}$ $l_t = \mu_t + \alpha e_t$ $b_t = (\mu_t + \beta^* e_t) - l_{t-1}$	$\mu_t = l_{t-1} + b_{t-1}$ $l_t = \mu_t(1 + \alpha e_t)$ $b_t = \mu_t(1 + \beta^* e_t) - l_{t-1}$	$\mu_t = l_{t-1} + b_{t-1} + s_{t-m}$ $l_t = (\mu_t + \alpha e_t) - s_{t-m}$ $b_t = (\mu_t + \beta^* e_t) - (l_{t-1} + s_{t-m})$ $s_t = (\mu_t + \gamma e_t) - (l_{t-1} + b_{t-1})$	$\mu_t = l_{t-1} + b_{t-1} + s_{t-m}$ $l_t = \mu_t(1 + \alpha e_t) - s_{t-m}$ $b_t = \mu_t(1 + \beta^* e_t) - (l_{t-1} + s_{t-m})$ $s_t = \mu_t(1 + \gamma e_t) - (l_{t-1} + b_{t-1})$	$\mu_t = (l_{t-1} + b_{t-1})s_{t-m}$ $l_t = (\mu_t + \alpha e_t)/s_{t-m}$ $b_t = (\mu_t + \beta^* e_t)/s_{t-m} - l_{t-1}$ $s_t = (\mu_t + \gamma e_t)/(l_{t-1} + b_{t-1})$	$\mu_t = (l_{t-1} + b_{t-1})s_{t-m}$ $l_t = \mu_t(1 + \alpha e_t)/s_{t-m}$ $b_t = \mu_t(1 + \beta^* e_t)/s_{t-m} - l_{t-1}$ $s_t = \mu_t(1 + \gamma e_t)/(l_{t-1} + b_{t-1})$
A <sub>d</sub>	$\mu_t = l_{t-1} + \phi b_{t-1}$ $l_t = \mu_t + \alpha e_t$ $b_t = (\mu_t + \beta^* e_t) - l_{t-1}$	$\mu_t = l_{t-1} + \phi b_{t-1}$ $l_t = \mu_t(1 + \alpha e_t)$ $b_t = \mu_t(1 + \beta^* e_t) - l_{t-1}$	$\mu_t = l_{t-1} + \phi b_{t-1} + s_{t-m}$ $l_t = (\mu_t + \alpha e_t) - s_{t-m}$ $b_t = (\mu_t + \beta^* e_t) - (l_{t-1} + s_{t-m})$ $s_t = (\mu_t + \gamma e_t) - (l_{t-1} + \phi b_{t-1})$	$\mu_t = l_{t-1} + \phi b_{t-1} + s_{t-m}$ $l_t = \mu_t(1 + \alpha e_t) - s_{t-m}$ $b_t = \mu_t(1 + \beta^* e_t) - (l_{t-1} + s_{t-m})$ $s_t = \mu_t(1 + \gamma e_t) - (l_{t-1} + \phi b_{t-1})$	$\mu_t = (l_{t-1} + \phi b_{t-1})s_{t-m}$ $l_t = (\mu_t + \alpha e_t)/s_{t-m}$ $b_t = (\mu_t + \beta^* e_t)/s_{t-m} - l_{t-1}$ $s_t = (\mu_t + \gamma e_t)/(l_{t-1} + \phi b_{t-1})$	$\mu_t = (l_{t-1} + \phi b_{t-1})s_{t-m}$ $l_t = \mu_t(1 + \alpha e_t)/s_{t-m}$ $b_t = \mu_t(1 + \beta^* e_t)/s_{t-m} - l_{t-1}$ $s_t = \mu_t(1 + \gamma e_t)/(l_{t-1} + \phi b_{t-1})$
M	$\mu_t = l_{t-1}b_{t-1}$ $l_t = \mu_t + \alpha e_t$ $b_t = (\mu_t + \beta^* e_t)/l_{t-1}$	$\mu_t = l_{t-1}b_{t-1}$ $l_t = \mu_t(1 + \alpha e_t)$ $b_t = \mu_t(1 + \beta^* e_t)/l_{t-1}$	$\mu_t = l_{t-1}b_{t-1} + s_{t-m}$ $l_t = (\mu_t + \alpha e_t) - s_{t-m}$ $b_t = [(\mu_t + \beta^* e_t) - s_{t-m}]/l_{t-1}$ $s_t = (\mu_t + \gamma e_t) - l_{t-1}b_{t-1}$	$\mu_t = l_{t-1}b_{t-1} + s_{t-m}$ $l_t = \mu_t(1 + \alpha e_t) - s_{t-m}$ $b_t = [\mu_t(1 + \beta^* e_t) - s_{t-m}]/l_{t-1}$ $s_t = \mu_t(1 + \gamma e_t) - l_{t-1}b_{t-1}$	$\mu_t = l_{t-1}b_{t-1}s_{t-m}$ $l_t = (\mu_t + \alpha e_t)/s_{t-m}$ $b_t = (\mu_t + \beta^* e_t)/(l_{t-1}s_{t-m})$ $s_t = (\mu_t + \gamma e_t)/(l_{t-1}b_{t-1})$	$\mu_t = l_{t-1}b_{t-1}s_{t-m}$ $l_t = \mu_t(1 + \alpha e_t)/s_{t-m}$ $b_t = \mu_t(1 + \beta^* e_t)/(l_{t-1}s_{t-m})$ $s_t = \mu_t(1 + \gamma e_t)/(l_{t-1}b_{t-1})$
M <sub>d</sub>	$\mu_t = l_{t-1}b_{t-1}^\phi$ $l_t = \mu_t + \alpha e_t$ $b_t = (\mu_t + \beta^* e_t)/l_{t-1}$	$\mu_t = l_{t-1}b_{t-1}^\phi$ $l_t = \mu_t(1 + \alpha e_t)$ $b_t = \mu_t(1 + \beta^* e_t)/l_{t-1}$	$\mu_t = l_{t-1}b_{t-1}^\phi + s_{t-m}$ $l_t = (\mu_t + \alpha e_t) - s_{t-m}$ $b_t = [(\mu_t + \beta^* e_t) - s_{t-m}]/l_{t-1}$ $s_t = (\mu_t + \gamma e_t) - l_{t-1}b_{t-1}^\phi$	$\mu_t = l_{t-1}b_{t-1}^\phi + s_{t-m}$ $l_t = \mu_t(1 + \alpha e_t) - s_{t-m}$ $b_t = [\mu_t(1 + \beta^* e_t) - s_{t-m}]/l_{t-1}$ $s_t = \mu_t(1 + \gamma e_t) - l_{t-1}b_{t-1}^\phi$	$\mu_t = l_{t-1}b_{t-1}^\phi s_{t-m}$ $l_t = (\mu_t + \alpha e_t)/s_{t-m}$ $b_t = (\mu_t + \beta^* e_t)/(l_{t-1}s_{t-m})$ $s_t = (\mu_t + \gamma e_t)/(l_{t-1}b_{t-1}^\phi)$	$\mu_t = l_{t-1}b_{t-1}^\phi s_{t-m}$ $l_t = \mu_t(1 + \alpha e_t)/s_{t-m}$ $b_t = \mu_t(1 + \beta^* e_t)/(l_{t-1}s_{t-m})$ $s_t = \mu_t(1 + \gamma e_t)/(l_{t-1}b_{t-1}^\phi)$

Fonte: Hyndman et al. (2002)

## APÊNDICE A – ANÁLISE DOS RESÍDUOS DO MODELO SARIMA\_RNA/RL

Para dar credibilidade à regressão é realizada a Análise de variância (ANOVA), para saber se a regressão é significativa, o teste F pode ser visualizado na Tabela x.

Tabela 16 - Análise da Variância do modelo SARIMA\_RNA/RL

	Soma dos Quadrados	g. l.	Média	F	p-valor
<b>Regressão</b>	74459228,8	2	37229614	155,81	<0,0001
<b>Resíduos</b>	21743696,4	91	238941,72		
<b>Total</b>	96202925,2				

Fonte: Saída do *Action Stat* adaptado pelo autor.

Com o p-valor encontrado < 00001, o modelo de regressão linear múltipla é adequado e significativo ao nível de significância de 5%. Considerando a relevância das variáveis dependentes em estudo (SARIMA e RNA), as duas são consideradas adequadas para a regressão estimada, obtendo um valor de “F” de 155,81.

Para que o modelo ajustado seja confiável, as suposições descritas no capítulo da metodologia, na seção 3.3.6 não podem ser violadas. O primeiro pressuposto analisado é que os resíduos possuem média igual a zero, tendo como resultado o atendimento do mesmo, os resultados dos outros pressupostos podem ser observados no Quadro 5.

Quadro 5 - Resultados dos pressupostos do modelo SARIMA\_RNA/RL

Pressupostos	Teste	Resultado (p-valor)
Os resíduos são homoscedásticos?	- Breusch-Pagan	0,14545985
Os erros são independentes?	- Durbin-Watson	0,041358
Os resíduos seguem a distribuição normal?	- Kolgomorov-Smirnov - Shapiro Wilk	p>0,2 p>0,2
Problemas de Multicolinearidade?	VIF	1,799
Resíduos possuem outliers?	Teste de Bonferroni	0,38005717

Fonte: Saída do *Action Stat* adaptado pelo autor.

Para testar a variabilidade dos resíduos, aplicou-se o teste de Breusch-Pagan encontrando um p-valor de 0,1454 logo não rejeita-se a hipótese de homoscedasticidade dos resíduos. Em relação ao teste de independência, verifica-se que os resíduos são independentes por meio do p-valor encontrado no teste de Durbin-Watson de 0,041358. Constata-se também

que os resíduos seguem a distribuição normal, comprovado pelos dois testes realizados, sendo os p-valores encontrados maiores que o nível de significância. A multicolinearidade foi diagnosticada por meio do VIF (*Variance Inflation Factor*), o VIF encontrado para as duas variáveis foi de 1,799, indicando assim que não há problemas de multicolinearidade, pois o valor de VIF encontrado é menor do que 5. Por fim, observou-se que na medida 30 há um possível *outlier*, então aplicou-se o teste de Bonferroni encontrando um p-valor de 0,38, sendo assim tem-se que a medida não é um *outlier* pois não rejeita-se a hipótese de que a observação não é *outlier*.



## APÊNDICE B – ANÁLISE DOS RESÍDUOS DO MODELO RL(SARIMA\_RNA/ACP)

Para dar credibilidade à regressão é realizada a Análise de variância (ANOVA), para saber se a regressão é significativa, o teste F pode ser visualizado na Tabela 17.

Tabela 17 - Análise da Variância do modelo RL(SARIMA\_RNA/ACP)

	Soma dos Quadrados	g. l.	Média	F	p-valor
<b>Regressão</b>	114065738,9	1	114065738,9	543,88	<0,0001
<b>Resíduos</b>	27264114,57	130	209723,9582		
<b>Total</b>	141329853,5				

Fonte: Saída do *Action Stat* adaptado pelo autor.

Com o p-valor encontrado  $< 0,001$ , o modelo de regressão linear simples é adequado e significativo ao nível de significância de 5%. Considerando a relevância da variável dependentes em estudo (SARIMA\_RNA/ACP), considerada adequada para a regressão estimada, obtendo um valor de “F” de 543,88.

Para que o modelo ajustado seja confiável, as suposições descritas no capítulo da metodologia, na seção 3.3.6 não podem ser violadas. O primeiro pressuposto analisado é que os resíduos possuem média igual a zero, tendo como resultado o atendimento do mesmo, os resultados dos outros pressupostos podem ser observados no Quadro 6.

Quadro 6 - Resultados dos pressupostos do modelo SARIMA\_RNA/RL

Pressupostos	Teste	Resultado (p-valor)
Os resíduos são homoscedásticos?	- Breusch-Pagan	0,13047985
Os erros são independentes?	- Durbin-Watson	0,03919988
Os resíduos seguem a distribuição normal?	- Kolgomorov-Smirnov - Shapiro Wilk	p>0,2 p>0,2
Resíduos possuem outliers?	Teste de Bonferroni	0,330326627

Fonte: Saída do *Action Stat* adaptado pelo autor.

Para testar a variabilidade dos resíduos, aplicou-se o teste de Breusch-Pagan encontrando um p-valor de 0,1304 logo não rejeita-se a hipótese de homoscedasticidade dos resíduos. Em relação ao teste de independência, verifica-se que os resíduos são independentes por meio do p-valor encontrado no teste de Durbin-Watson de 0,0391. Constata-se também que os resíduos seguem a distribuição normal, comprovado pelos dois testes realizados, sendo os p-valores encontrados maiores que o nível de significância. A multicolinearidade não foi

testada, visto que a regressão linear realizada contém apenas uma variável independente. Por fim, observou-se que na medida 30 há um possível *outlier*, repare que esta observação já havia sido apontada como possível outlier na análise de resíduos do modelo SARIMA\_RNA/RL (Apêndice A). Desse modo, aplicou-se o teste de Bonferroni encontrando um p-valor de 0,3303, sendo assim tem-se que a medida não é um *outlier* pois não rejeita-se a hipótese de que a observação não é *outlier*.

