

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS NATURAIS E EXATAS
ESPECIALIZAÇÃO EM ESTATÍSTICA E
MODELAGEM QUANTITATIVA

**O USO DO MÉTODO DE COMBINAÇÕES DE PREVISÃO
APLICADAS AOS GASTOS NO CONSUMO DA CESTA
BÁSICA DE PORTO ALEGRE - RS**

MONOGRAFIA DE ESPECIALIZAÇÃO

Fábio Roberto de Souza

Santa Maria, RS, Brasil
2019

**O USO DO MÉTODO DE COMBINAÇÕES DE PREVISÃO
APLICADAS AOS GASTOS NO CONSUMO DA CESTA
BÁSICA DE PORTO ALEGRE - RS**

Por

Fábio Roberto de Souza

Monografia apresentada ao Curso de Especialização em Estatística e Modelagem Quantitativa, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Especialista em Estatística e Modelagem Quantitativa**

Orientador: Prof. Cleber Bisognin, Dr.

Santa Maria, RS, Brasil
2019

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS NATURAIS E EXATAS
ESPECIALIZAÇÃO EM ESTATÍSTICA E
MODELAGEM QUANTITATIVA

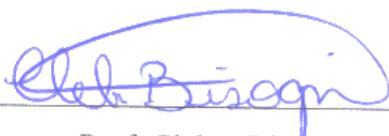
**O USO DO MÉTODO DE COMBINAÇÕES DE PREVISÃO
APLICADAS AOS GASTOS NO CONSUMO DA CESTA
BÁSICA DE PORTO ALEGRE - RS**

Elaborado por

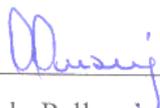
Fábio Roberto de Souza

como requisito parcial para obtenção do grau de
Especialista em Estatística e Modelagem Quantitativa

COMISSÃO EXAMINADORA:



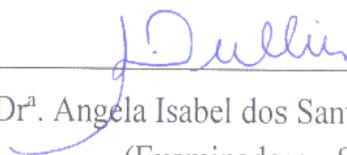
Prof. Cleber Bisognin, Dr.
(Presidente/Orientador)



Prof^a. Angela Pellegrin Ansuji, Dr^a. (UFSM)
(Examinadora)



Prof^a. Vanessa Siqueira Peres da Silva, Dr^a. (UFSM)
(Examinadora)



Prof^a. Dr^a. Angela Isabel dos Santos Dullius, Dr^a. (UFSM)
(Examinadora - Suplente)

Santa Maria, 29 de Agosto de 2019.

RESUMO

Monografia de Especialização
Especialização em Estatística e Modelagem Quantitativa
Universidade Federal de Santa Maria, RS, Brasil

O USO DO MÉTODO DE COMBINAÇÃO DE PREVISÕES APLICADAS AOS GASTOS NO CONSUMO DA CESTA BÁSICA DE PORTO ALEGRE - RS

Autor: Fábio Roberto de Souza
Orientador: Prof. Cleber Bisognin, Dr.
Local e data da defesa: Santa Maria, 29 de Agosto de 2019.

Elaborar previsões dos gastos médios no consumo da Cesta Básica de Porto Alegre, RS, permite termos uma breve aproximação dos possíveis valores a serem cobrados por esse produto. Diante desta situação, o objetivo desta pesquisa foi ajustar modelos e combinar previsões. Para isso, foram encontrados os modelos ARIMA, ARFIMA, SARIMA, SARMA e de Suavização Exponencial de Holt-Winters, os quais passaram por testes estatísticos. Para as combinações de previsões de dois, três, quatro, cinco e seis modelos foram utilizados os métodos por Média Aritmética Simples, por Mediana, por Variância Mínima, por Regressões de Mínimos Quadrados Ordinários, Robusta de Mínimos Quadrados Podados, Robusta MM e a Quantílica, além do Inverso do *Rank* e do Auto Vetor corrigido por Viés. Assim, por intermédio das medidas de Acurácias MAPE, U de Theil, RMSE e MAE, as melhores previsões para combinação de previsões, resultaram dos métodos por Regressão Quantílica e de Mínimos Quadrados Ordinários.

Palavras-chave: Gastos com Cesta Básica. Porto Alegre. Modelos de Séries Temporais. Combinações de previsão.

ABSTRACT

Specialization Monograph
Specialization in Statistics and Quantitative Modeling
Federal University of Santa Maria, RS, Brazil

O USO DO MÉTODO DE COMBINAÇÃO DE PREVISÕES APLICADAS AOS GASTOS NO CONSUMO DA CESTA BÁSICA DE PORTO ALEGRE - RS

USING THE FORECASTING COMBINATION METHOD
APPLIED TO EXPENDITURE ON BASKET CONSUMPTION
PORTO ALEGRE CITY BASICS - RS

AUTHOR: Fábio Roberto de Souza
ORIENTATOR: Prof. Cleber Bisognin, Dr.
Place and Defende Date: Santa Maria, August 29, 2019.

Making forecasts of the average consumption spending of the Basic Basket of Porto Alegre, RS, allows us to have a brief approximation of the possible values to be charged for this product. Given this situation, the objective of this research was to adjust models and combine forecasts. For this, the ARIMA, ARFIMA, SARIMA, SARMA and Holt-Winters Exponential Smoothing models were found, which underwent statistical tests. For the prediction combinations of two, three, four, five and six models, the methods used were Simple Arithmetic Average, Median, Minimum Variance, Ordinary Least Squares Regressions, Pruned Least Squares Robusta, MM Robusta and Quantile, as well as the Reverse Rank and Auto Vector corrected by Bias. Thus, through the MAPE, Theil U, RMSE and MAE Accuracy measurements, the best predictions for combining predictions resulted from the Quantum Regression and Ordinary Least Squares methods.

Key-words: Basic Basket Expenses. Porto Alegre. Time Series Models. Forecast combinations.

LISTA DE FIGURAS

Figura 1 – Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de jul/1995 à fev/2019: (a) gráfico da série temporal; (b) função de autocorrelação amostral.	68
Figura 2 – Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de jul/1995 à fev/2019 após retirada da tendência determinística: (a) gráfico da série temporal; (b) função de autocorrelação amostral. ..	71
Figura 3 – Predição e Previsão da Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de jul/1995 à fev/2020, utilizando o Modelo 4.....	74
Figura 4 – Predição e Previsão da Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de jul/1995 à fev/2020, utilizando as Combinações de Previsão, com dois (2) modelos, por Regressão Quantílica e Mínimos Quadrados.	78
Figura 5 – Predição e Previsão da Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de jul/1995 à fev/2020, utilizando as Combinações de Previsão, com três (3) modelos, por Regressão Quantílica e Mínimos Quadrados.....	78
Figura 6 – Predição e Previsão da Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de jul/1995 à fev/2020, utilizando as Combinações de Previsão, com quatro (4) modelos, por Regressão Quantílica e Mínimos Quadrados.	79
Figura 7 – Predição e Previsão da Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de jul/1995 à fev/2020, utilizando as Combinações de Previsão, com cinco (5) modelos, por Regressão Quantílica e Mínimos Quadrados.	79
Figura 8 – Predição e Previsão da Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de jul/1995 à fev/2020, utilizando as Combinações de Previsão, com seis (6) modelos, por Regressão Quantílica e Mínimos Quadrados.....	80
Figura 9 – Medidas de Acurácias dos seis modelos ajustados para os procedimentos de combinações de previsões.	83
Figura 10 – Menores RMSE por métodos de combinações de previsão e por número de previsões combinadas.	83
Figura 11 – Menores MAE por métodos de combinações de previsão e por número de previsões combinadas.	84
Figura 12 – Menores MAPE por métodos de combinações de previsão e por número de previsões combinadas.	84
Figura 13 – Menores U de Theil por métodos de combinações de previsão e por número de previsões combinadas.	85
Figura 14 – Previsões utilizando o método Combinações de Previsões por Regressão Quantílica.	85
Figura 15 – Previsões utilizando o método Combinações de Previsões por Mínimos Quadrados.....	86

LISTA DE TABELAS

Tabela 1 – Produtos alimentícios que compõem a Cesta Básica de Porto Alegre, RS, pesquisada pelo DIEESE.	25
Tabela 2 – Matriz de previsões para combinação de previsões pelos métodos da Média Aritmética Simples, da Mediana e da Variância Mínima.	55
Tabela 3 – Matriz dos Modelos Individuais para Combinação de Previsões pelos Métodos de Regressão.	57
Tabela 4 – Resultado do ajuste do modelo ARIMA(2,1,2).	70
Tabela 5 – Resultado do ajuste da tendência determinística.	70
Tabela 6 – Resultado do ajuste do modelo ARFIMA(1,d,0), após retirada da tendência determinística.	71
Tabela 7 – Resultado do ajuste do modelo SARMA(3,1)×(2,2) ₁₂ , após retirada da tendência determinística.	72
Tabela 8 – Resultado do ajuste do Modelo SARIMA(1,1,1)×(2,1,1) ₁₂	73
Tabela 9 – Resultado do ajuste do Modelo de Suavização Exponencial Holt-Winters com efeitos sazonais aditivos.	73
Tabela 10 – Resultado do ajuste do Modelo de Suavização Exponencial Holt-Winters com efeitos sazonais multiplicativos.	73
Tabela 11 – Medidas de Acurácias para predição dos Modelos 1 a 6.	74
Tabela 12 – Medidas de Acurácia para predição para as Combinações de Previsão utilizando dois (2) modelos de previsões individuais para cada método de combinação.	75
Tabela 13 – Medidas de Acurácia para predição para as Combinações de Previsão utilizando três (3) modelos de previsões individuais para cada método de combinação.	76
Tabela 14 – Medidas de Acurácia para predição para as Combinações de Previsão utilizando quatro (4) modelos de previsões individuais para cada método de combinação.	76
Tabela 15 – Medidas de Acurácia para predição para as Combinações de Previsão utilizando cinco (5) modelos de previsões individuais para cada método de combinação.	77
Tabela 16 – Medidas de Acurácia para predição para as Combinações de Previsão utilizando os seis (6) modelos de previsões individuais para cada método de combinação.	77
Tabela 17 – Medidas de Acurácia de Previsão para o Modelo 4.	80
Tabela 18 – Medidas de Acurácia de Previsão para a Combinação de Previsão utilizando Regressão Quantílica - RQ.	81
Tabela 19 – Medidas de Acurácia de Previsão para a Combinação de Previsão utilizando Regressão por Mínimos Quadrados - MQO.	81
Tabela 20 – Previsão da Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de mar/2019 à fev/2020 utilizando Combinação de Previsões, por meio da Regressão Quantílica - RQ.	82
Tabela 21 – Previsão da Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de mar/2019 à fev/2020 utilizando Combinação de Previsões, por meio da Regressão por Mínimos Quadrados - MQO. ..	82

LISTA DE SIGLAS

EIG2 – Combinação de previsão de Auto Vetor corrigida por Viés

FAC – Função de Autocorrelação

FACP – Função de Autocorrelação Parcial

InvR – Combinação de previsão Inverso do *Rank*

MA – Média Aritmética Simples

MAE – Erro Médio Absoluto

MAPE – Erro Percentual Médio Absoluto

MED – Mediana

MM – Regressão Robusta MM

MQO – Regressão Mínimos Quadrados

MQP – Regressão Robusta Mínimos Quadrados Podados

QR – Regressão Quantílica

RMSE – Raiz do Erro Quadrático Médio

U de Theil – Coeficiente

VM – Variância Mínima

d – Ordem do Processo ARIMA, ARFIMA, SARIMA e SARMA

p – Ordem do Processo ARIMA, ARFIMA, SARIMA e SARMA

q – Ordem do Processo ARIMA, ARFIMA, SARIMA e SARMA

D – Ordem do Processo SARIMA e SARMA

P – Ordem do Processo SARIMA e SARMA

Q – Ordem do Processo SARIMA e SARMA

SUMÁRIO

1.	INTRODUÇÃO.....	11
1.1.	OBJETIVOS.....	13
1.2.	JUSTIFICATIVA.....	14
2.	REVISÃO DE LITERATURA.....	16
2.1.	ALGUNS CONCEITOS INICIAIS DE SÉRIES TEMPORAIS.....	16
2.2.	ALGUNS ESTUDOS COM MÉTODOS DE SÉRIES TEMPORAIS.....	16
2.3.	APLICAÇÕES COM MÉTODOS DE SÉRIES TEMPORAIS.....	18
2.4.	COMBINAÇÕES DE PREVISÃO.....	21
3.	METODOLOGIA.....	23
3.1.	COLETA DE DADOS PARA PESQUISA.....	23
3.2.	FERRAMENTAS UTILIZADAS.....	25
3.3.	PROCESSO ESTACIONÁRIO.....	26
3.3.1.	Análise gráfica.....	27
3.3.2.	Teste de Normalidade de Shapiro-Wilk.....	27
3.3.3.	Análise de Resíduos.....	28
3.3.3.1.	Função de Autocorrelação Amostral.....	29
3.3.3.2.	Função de Autocorrelação Parcial Amostral.....	31
3.3.3.3.	Teste de Ljung-Box.....	32
3.4.	PROCESSO NÃO ESTACIONÁRIO.....	34
3.4.1.	Teste para verificação de tendência de Wald-Wolfowitz.....	34
3.4.2.	Testes de Raízes Unitárias.....	36
3.4.2.1.	Teste de Dickey-Fuller Aumentado (ADF).....	36
3.4.2.2.	Teste de Phillips-Perron.....	38
3.4.2.3.	Teste de Kwiatkowski-Phillips-Schmidt-Shin (KPSS).....	39
3.5.	MODELOS PARA PREVISÕES DE SÉRIES TEMPORAIS.....	40
3.5.1.	Modelo 1: Modelos Auto-Regressivos Integrados e de Médias Móveis – ARIMA(p,d,q)......	40
3.5.2.	Modelo 2: Modelos de Memória Longa ARFIMA(p,d,q)......	45
3.5.3.	Modelo 3: Modelo Sazonal Auto-regressivo e de Médias Móveis – SARMA(p,q) \times (P,Q) $_s$	47
3.5.4.	Modelo 4: Modelos Auto-Regressivos Integrados e de Médias Móveis Sazonais – SARIMA(p,d,q) \times (P,D,Q) $_s$	47
3.5.4.1.	Métodos de Previsão utilizando a representação Auto regressiva infinita.....	48
3.5.5.	Modelo 5: Modelo de Suavização Exponencial – Aditivo.....	52
3.5.6.	Modelo 6: Modelo de Suavização Exponencial Multiplicativo.....	53
3.6.	MÉTODOS DE COMBINAÇÕES DE PREVISÕES.....	53
3.6.1.	Combinação de Previsão pelo Método da Média Aritmética Simples (MS).....	54
3.6.2.	Combinação de Previsão pelo Método da Mediana (MED).....	56
3.6.3.	Combinação de Previsão pelo Método da Variância Mínima (VM).....	56
3.6.4.	Combinação de Previsão pelo Método de Regressão de Mínimos Quadrados (MQO).....	58

3.6.5.	Combinação de Previsão pelo Método de Regressão Robusta de Mínimos Quadrados Podados (MQP).....	59
3.6.6.	Combinação de Previsão pelo Método de Regressão Robusta (MM).....	60
3.6.7.	Combinação de Previsão pelo Método de Regressão Quantílica (RQ).....	61
3.6.8.	Combinação de Previsão pelo Método do Inverso do Rank.....	62
3.6.9.	Combinação de Previsão pelo Método de Auto Vetor Corrigida pelo Viés (EIG2).....	63
3.7.	MEDIDAS DE ACURÁCIA.....	65
3.7.1.	Erro Absoluto Médio – MAE.....	65
3.7.2.	Raiz do Erro Quadrático Médio – RMSE.....	65
3.7.3.	Erro Percentual Médio Absoluto – MAPE.....	66
3.7.4.	Coefficiente U de Theil.....	67
4.	RESULTADOS E DISCUSSÃO	68
5.	CONSIDERAÇÕES FINAIS	87
	REFERENCIAS	89
	ANEXO A	93
	APÊNDICE A	95
	APÊNDICE B.....	95
	APÊNDICE C.....	97

1. INTRODUÇÃO

A palavra Cesta Básica é um termo designado para fazer referência a um conjunto de mercadorias, as quais fazem parte do consumo básico mensal das famílias. No entanto, com base nas características próprias que cada estado brasileiro apresenta, sejam elas culturas, de produção ou de consumo, é normal que a composição da Cesta Básica varie de acordo com a localidade, pois alguns produtos que a compõem em um determinado estado ou região não são os mesmos que compõe a de outros.

No Brasil, uma quantidade de produtos de gêneros alimentícios de primeiras necessidades à nutrição básica do trabalhador foi definida por lei como ração-tipo essencial mínima (LAVINAS, 1998), e o conjunto de mercadorias popularmente conhecida nos dias atuais como Cesta Básica de Alimentos permanece com o mesmo propósito: fornecer provisões para uma alimentação balanceada e saudável, a base de proteínas, calorias, ferro, cálcio e fósforo, não podendo conforme previsto em lei seu valor total ultrapassar o valor do salário mínimo estabelecido (BRASIL, 1938).

Na segunda metade da década de 1930, indícios de subnutrição começaram a ser percebidas e consideradas como um problema sério de cunho social a saúde pública. Admitindo-se descontrole no sistema de saúde e práticas alimentares inapropriadas, o governo do presidente Getúlio Vargas percebeu que não era possível reparar aquela situação de forma imediata, o que exigiu com urgência severas medidas para conter a generalização das condições apresentadas (SILVA, 1995).

A par das referidas necessidades, ainda na segunda metade da década de 30, o Governo propôs duas medidas que possivelmente causaria impacto na alimentação da classe trabalhadora. Sendo uma das medidas efetuada sob o salário mínimo, que desde então deveria ser um valor suficiente para satisfazer as necessidades básicas de alimentação do trabalhador e de sua família e a outra, foi relacionada ao Serviço de Alimentação da Previdência Social (SAPS), criado em 1940, com a intenção de introduzir refeitórios nas grandes empresas e fornecer alimentação nas menores, realizar a venda de alimentação a baixo custo ao trabalhador que continha um amplo grupo familiar, oportunizar educação alimentar, além de incentivar a formação técnica especializada e também apoiar pesquisas de alimentos e da situação alimentar da nação (SILVA, 1995).

Em 30 de abril de 1938, por intermédio do Decreto-Lei nº. 399 regularizou-se a Lei nº 185 de 14 de Janeiro de 1936, determinando o salário mínimo como a

remuneração base pela força de trabalho de todo cidadão brasileiro em idade adulta, que estivesse exercendo atividade produtiva remunerada, ficando sob a responsabilidade das Comissões de Salário Mínimo descrito no Art. 29 a incumbência da fixação do salário mínimo da região, ou zona, de sua competência (BRASIL, 1938). Desde então, por intermédio das medidas anteriores os sucessivos governos intensificaram melhorias e criação de programas sociais e alimentares em prol do combate a fome e desnutrição da população nacional.

Estudos censitários realizados e consentidos pela Comissão criada pela Lei nº 185, serviram como base para estipular a remuneração mínima a ser paga pela jornada de trabalho em cada região, juntamente com um conjunto de itens alimentícios, que obtinham quantidades suficientes de produtos que nutrissem um trabalhador no período de um mês (PAIVA *et al.*, 2017).

Assim, os estados foram distribuídos em três Regiões conforme Decreto Lei 399 de 1938, da seguinte forma: na Região 1, ficaram os estados de São Paulo, Minas Gerais, Espírito Santo, Rio de Janeiro, Goiás e o Distrito Federal; a Região 2, formada pelo maior número de estados foi composta por Pernambuco, Bahia, Ceará, Rio Grande do Norte, Alagoas, Sergipe, Amazonas, Pará, Piauí, Tocantins, Acre, Paraíba, Rondônia, Amapá, Roraima e Maranhão e a Região 3, foi composta pelo estado do Paraná, Santa Catarina, Rio Grande do Sul, Mato Grosso e Mato Grosso do Sul. Deste modo, foram instituídas três regiões brasileiras, separadas pela similaridade dos produtos que compunha a Cesta Básica (PAIVA *et al.*, 2017).

A oposição, a fim de conter a fome e desnutrição no território brasileiro, foi motivada por práticas vivenciadas que ganharam força entre os períodos de 1940 a 1960, por meio de pesquisas realizadas por instituições universitárias, centros de pesquisas, organizações e intelectuais dedicados à produção de materiais sobre segurança alimentar, nutrição e planos de combate à fome e pobreza no país (AMORIM, 2017).

Com alterações feitas no teto salarial mínimo, atualizada por meio da Constituição Federal (ou Constituição Cidadã) de 1988, adicionaram-se direitos, nos quais, o salário mínimo deveria amparar, pois com o acréscimo do promovimento a educação, saúde, lazer e previdência social, além das correções obrigatórias na renda para que não ocorressem perdas do poder de compras do trabalhador, a proposta era de garantir pelo menos de forma constante o acesso a bens de primeiras necessidades a população brasileira (PAIVA *et al.*, 2017).

Com base no Decreto-Lei nº 399, o Departamento Intersindical de Estatística e Estudos Socioeconômicos, (DIEESE, 1993) refere-se a Cesta Básica de Alimentos como Ração Essencial Mínima (REM). E, seguindo as exigências impostas em lei, as quais determinam que os produtos integrantes da Cesta Básica sejam ricos em: proteínas, calorias, ferro, cálcio e fósforo, o DIEESE desde 1959, realiza em São Paulo o monitoramento mensal da variação dos preços¹ de 13 mercadorias (carne, leite, arroz, feijão, banana, batata, tomate, açúcar, banha/óleo, café, farinha, manteiga e o pão francês) essenciais à nutrição dos brasileiros, por intermédio da Pesquisa Nacional da Cesta Básica de Alimentos (PNCBA).

No entanto, em Porto Alegre este monitoramento só se iniciou em janeiro de 1977, sendo a segunda capital brasileira a ter os preços das mercadorias de primeiras necessidades nutricionais coletadas e analisadas, para então verificar a situação dos residentes daquela região. E, ao decorrer dos anos, a aplicação da PNCBA nas demais capitais e no Distrito Federal se tornou muito importante para o país, não apenas como uma ferramenta para informar as sucessivas variações de preços sofridas pelos produtos, mas também serviu como instrumento de revisão e ajuste dos salários mínimos (nacional e regionais), e possibilitou a verificação da qualidade de vida da população.

Diante das situações mencionadas anteriormente, a proposta deste estudo foi a de analisar um conjunto de observações referentes aos Gastos² médios mensais do consumo da Cesta Básica de Alimentos de Porto Alegre, RS, com a finalidade de realizar previsões por intermédio da Combinação de Modelos de Séries Temporais, a fim de, se chegar o mais próximo possível dos preços que serão cobrados futuramente pelo conjunto de mercadorias que compõem a Cesta Básica.

1.1. OBJETIVOS

O objetivo desta pesquisa foi identificar modelos que justificassem o comportamento da série, possibilitando a realização de previsões para formular e descrever as constantes variações “sofridas” pelos gastos mensais médios, praticados pelos consumidores de Porto Alegre - RS, em produtos de alimentação básica, com base nos objetivos específicos descritos a seguir:

¹ A palavra preços faz referência ao valor cobrado pela Cesta Básica de Alimentos pelos estabelecimentos comerciais, como as redes de supermercados ou mercados familiares.

² Nesta pesquisa, quando usado a palavra Gastos estamos fazendo referência ao valor pago pelos consumidores.

- I. Verificar se o comportamento da série está atrelado às componentes, tendência e sazonalidade;
- II. Ajustar modelos apropriados às observações para análise do consumo da Cesta Básica;
- III. Realizar combinações de previsões entre os modelos, e identificar o mais adequado para inferir sobre os dados.

1.2. JUSTIFICATIVA

Mediante os sucessivos aumentos do preço médio cobrado pela Cesta Básica de Alimentos de Porto Alegre, que se manteve desde o mês de novembro de 2015 acima dos R\$ 400,00, verificou-se pelas pesquisas mensais realizadas pelo Departamento Intersindical de Estatística e Estudos Socioeconômicos (DIEESE) que de setembro de 2016 a janeiro de 2018 o valor médio da Cesta Básica cobrada em Porto Alegre foi consecutivamente a mais cara quando comparada com a Cesta Básica das demais capitais brasileiras e também do Distrito Federal. E, a mesma manteve-se após esse período até fevereiro de 2019 entre as capitais com os valores médios mais elevados cobrados pela Cesta Básica no Brasil.

Com o aumento contínuo dos produtos de gêneros alimentícios da Cesta Básica, um percentual considerável do orçamento familiar dos consumidores (assalariados) de Porto Alegre mantinha-se destinados às necessidades alimentares, reduzindo a possibilidade de investir, poupar ou de até mesmo consumir outros bens ou serviços. Diante destes fatos, se verificou a importância de estar sempre alerta às alterações de preços “sofridas” por essas mercadorias, já que, o aumento de preços resulta em redução do poder de compra do trabalhador e conseqüentemente a restrição da renda gera endividamento e inadimplência por parte dos mesmos, ocasionados pela necessidade de gastar mais do que é ganho.

A análise de preços médios de mercadorias consumidas em determinado período é uma maneira de também verificar o custo de vida daquela região (MINE *et al.*, 2016), de tal forma que, quanto maior for o custo de vida em uma determinada localidade, menor também será o poder de compra de um assalariado e maior se torna a necessidade de receber uma remuneração que possa manter pelo menos constante o padrão de compras de produtos necessários para seu bem estar.

Diante das constantes variações de preços cobrados mensalmente pelas 13 mercadorias (carne, leite, arroz, feijão, banana, batata, tomate, açúcar, banha/óleo, café, farinha, manteiga e o pão francês) que fazem parte da composição da Cesta Básica de Porto Alegre, para o presente estudo foi crucial o monitoramento dos produtos conjuntamente, pois tendo em vista que, os preços cobrados aos consumidores dependem principalmente da produção, que normalmente depende do clima, dos gastos em insumos, do preço dos fretes, do valor do combustível, das políticas adotadas por instituições empresariais, pelo aumento de impostos aprovados pelos governos, etc., a proposta desta pesquisa foi limitada em analisar apenas a variação média dos gastos totais com a Cesta Básica de Porto Alegre, desconsiderando a influência exercida pelas variáveis mencionadas.

Como descrito anteriormente, a proposta desta pesquisa foi observar e prever as variações dos Gastos da Cesta Básica de Porto Alegre, RS, pelo método de combinação de previsões, onde foram estudadas conjuntamente duas variáveis com 284 observações mensais cada, nos períodos de julho de 1995 a fevereiro de 2019, referentes às variáveis Gastos (ou preços) médios no consumo da Cesta Básica e os Números Índices do Índice Nacional de Preços ao Consumidor Amplo (IPCA) da localidade de Porto Alegre.

Além das 284 observações das variáveis mencionadas, também foram coletadas e deflacionadas observações dos meses de março a julho de 2019, que serviram para comparar com as previsões encontradas passos á frente. No entanto, as observações dos meses de junho e julho de 2019 foram inseridas neste estudo, conforme foram sendo atualizadas as bases de dados do Departamento Intersindical de Estatística e Estudos Socioeconômicos (DIEESE) e também o banco de dados do Instituto Brasileiro de Geografia e Estatística (IBGE).

Como principal justificativa do uso do método de combinação de previsões no consumo médio mensal da Cesta Básica de Porto Alegre, foi considerada a mesma proposta feita em 1969 por Bates e Granger, quando identificaram que previsões realizadas pela combinação de modelos individuais com parâmetros significativos possivelmente resultavam em melhores previsões, tendo em vista que, ao descartar outros modelos individuais também significativos para estudar apenas o que apresentou resultados mais favoráveis estão sendo desprezadas as informações individuais únicas que cada modelo possui.

2. REVISÃO DE LITERATURA

Esta seção é composta por quatro subseções, onde a primeira aborda alguns conceitos iniciais sobre séries temporais, a segunda explora a aplicação e áreas de estudos, a terceira demonstra os estudos realizados e a quarta faz uma breve introdução sobre combinações de previsão.

2.1. ALGUNS CONCEITOS INICIAIS DE SÉRIES TEMPORAIS

Entre inúmeras definições de um conjunto de dados expresso em função do tempo, a mais corriqueira denominação de séries temporais é a que descreve um conjunto de observações realizadas de forma ordenada ao longo do tempo. Para Morettin e Toloí (2004) série temporal é definida como sendo uma trajetória de dados coletados e alinhados em um determinado espaço do tempo, na qual, uma parte ou fração do total das observações da variável é escolhida e isolada para fins de análise. Na visão de Wooldridge (2007) série temporal é a coleção de um fluxo de observações de uma ou mais variáveis obtidas ao decorrer de períodos passados. Gujarati e Porter (2011) denominam as observações de uma variável coletada em ciclos de intervalos distintos como sendo uma série temporal. E, Rossi e Neves (2014) definem uma série temporal ou histórica como uma sucessão de informações quantitativas coletadas ordenadamente.

Com base nas definições de séries temporais, o principal objetivo de estudar um conjunto de dados da série (função do tempo), é gerar previsões que possibilitem prever resultados que ajudem a resolver questões de diferentes áreas do conhecimento, a fim de otimizar e reduzir perdas de quaisquer naturezas para a sociedade como um todo.

2.2. ALGUNS ESTUDOS COM MÉTODOS DE SÉRIES TEMPORAIS

Simulações efetuadas por metodologias distintas são realizadas em intervalos de dados históricos, a fim de, projetar modelos que deem suporte a elaboração de um conjunto de medidas, as quais possam ser utilizadas como recursos em projeções futuras e adotadas por especialistas, seja na área econômica, meteorológica, biológica, da saúde ou em quaisquer outras, tendo como base elementos que possam ser explorados e modelados mediante técnicas complexas, com intuito de pressupor previsões que

possibilitem a formulação estratégica a ser traçada pelas informações verificadas nos dados (BRESSAN, 2004).

Estudos relatados ao longo dos anos demonstraram que variáveis econômicas que têm como principais mercadorias as de alimentação, “sofrem fortes” influências provocadas pelas variações do fator tempo, pois caracterizado pela presença de efeitos sazonais, de forma direta e indireta impossibilitam uma avaliação clara e precisa das observações analisadas (CARVALHO, FERRAZ e SAFADI, 2006). Fato esse, que comprova que mesmo com a utilização de modelos de previsões sofisticados, ainda assim, as estimativas de previsões encontradas por intermédio de técnicas de modelagem de dados não são totalmente confiáveis, pois tendo em vista que a produção agrícola é diretamente afetada pelas oscilações do tempo, não é possível obter exatidão das constantes variações do clima, para que se possam modelar os dados com melhor precisão.

As aplicações com uso de séries em função do tempo têm se mostrado cada vez mais exigidas. E nitidamente se percebe que a análise de série temporal tornou-se um instrumento valioso para interpretações de dados nos diferentes campos de estudos de cursos universitários (CARVALHO, FERRAZ e SÁFADI, 2006), pois, por intermédio de aplicações, estudantes e pesquisadores desenvolvem estudos científicos feitos por simulações de dados direcionados a suas respectivas áreas de conhecimentos, desenvolvendo novos modelos e aperfeiçoando técnicas que apresente melhoria no ajuste dos dados para que se alcance resultados mais favoráveis nas estimações.

Para Bueno (2008), a introdução de técnicas temporais como instrumento de análise na área econômica tem implicações positivas, pois, além de facilitar interpretações de resultados estimados de um conjunto de observações, ainda permitem que se tenha uma modelagem equacionada para diferentes situações, as quais possam resultar em melhorias posteriores do contexto de uma sociedade. Conforme, ainda Bueno (2008), as técnicas econométricas juntamente com modelos temporais ajudam na condução de esclarecimentos de um determinado acontecimento. Mesmo que tais justificativas sejam de teor preliminar, ainda sim, permitem que se justifiquem teorias e também se faça previsões, sejam essas previsões para alguma política a ser adotada ou para alguma já adotada e malconduzida, ou quaisquer outras questões que possam atingir desde a credibilidade de uma empresa até mesmo o bem-estar da população.

Do mesmo modo, estudos realizados utilizaram técnicas de séries temporais em pesquisas referentes ao setor de energia elétrica, visando encontrar novas soluções,

como por exemplo, a diminuição da poluição para reduzir alterações climáticas (GONÇALVES e SAFADI, 2008), pois, sendo este um setor possuidor de fontes renováveis e não renováveis e, de uso essencial para o crescimento e desenvolvimento de um Estado, é preciso explorar as informações que dizem respeito a esse bem. Além disso, este setor também é responsável pelos avanços e criações de novas tecnologias que suprem as necessidades pessoais e empresariais da sociedade.

Esta análise está também presente na área de meteorologia, cuja necessidade de investigar e “buscar” uma melhor compreensão a respeito das causas e efeitos de precipitações pluviométricas e variações nas temperaturas (variações climáticas) é de relevância indiscutível para que se possa encontrar padrões cíclicos e tendências em determinadas estações do ano (COSTA, FERREIRA e CORDEIRO, 2015). As interpretações destas variações podem contribuir para produção e manutenção de produtos agrícolas em períodos de safra e, na prevenção, em casos de falta de água potável em regiões mais secas. Além disso, esses estudos também podem vir a contribuir na precaução de catástrofes, como tempestades, granizos, enxurradas, etc., que são os fenômenos mais recorrentes no território brasileiro.

Em epidemiologia, as análises de observações de uma variável temporal baseadas em técnicas de previsões, influenciam em fatos que causam risco a saúde pública (ANTUNES e CARDOSO, 2015). Isso demonstra que a antecipação dos resultados vistos por meio do aprofundamento da análise quantitativa, em muitos casos podem ser o caminho ou indicação para uma possível realização de um diagnóstico, em outras palavras, essa seria uma possível solução para realizar intervenções juntamente com um planejamento bem elaborado, a fim de, evitar que algum tipo de contaminação se espalhe.

Essas são algumas das áreas que a exploração de observações em função do tempo é constantemente analisada, no entanto, técnicas de análise envolvendo séries temporais podem ser utilizadas em várias outras áreas de conhecimento, como na física, biologia, saúde, etc.

2.3. APLICAÇÕES COM MÉTODOS DE SÉRIES TEMPORAIS

Na verificação das variáveis temporais climáticas, temperatura do ar, umidade relativa e precipitação pluviométrica da cidade de Ariquemes, em Rondônia, Carvalho *et al.* (2016) nortearam uma pesquisa objetivando explorar informações sazonais e

tendenciosas das variáveis climáticas em questão, baseados em dados de dezembro de 2010 a fevereiro de 2014, coletados pelo Instituto Nacional de Meteorologia (INMET). Por intermédio da técnica de ajustamento de modelos individuais sazonais auto-regressivos integrados de médias móveis (SARIMA(p,d,q) \times (P,D,Q) $_s$) geraram e identificaram padrões nas variáveis temporais que possibilitaram melhorias na distinção do clima da região. Além disso, concluiu-se que o modelo SARIMA(p,d,q) \times (P,D,Q) $_s$ é recomendável para estudos com séries de tempo de variáveis climáticas, como temperatura e umidade relativa do ar, no entanto, não se chegou a conclusões favoráveis com esse modelo ao se tratar da variável precipitação pluviométrica.

Buscando a melhor compreensão sobre as políticas de preços mínimos praticados, que é garantia de uma remuneração justa e incentivo ao produtor, na continuidade da produção e abastecimento de alimentos no mercado interno, Saraiva *et al.* (2014) projetaram um estudo visando mostrar o efeito das políticas de intervenção adotadas pelo governo, na remuneração de produtores de arroz do estado do Rio grande do Sul. Utilizando-se, testes estatísticos para melhor interpretação dos resultados, realizou-se primeiramente o teste de raiz unitária de Dickey-Fuller Aumentado (ADF) e o Kwiatkowski-Phillips-Schmidt-Shin (KPSS), procedimento feito para averiguar a presença raiz unitária na série. Além disso, foi necessário o uso de variáveis *dummies* para ajustar a existência de sazonalidade. Após identificar o modelo ARIMA, usou-se o a Análise de Intervenção para interpretação do preço do arroz e, concluiu-se que os contratos de opções de venda juntamente com o prêmio de escoamento de venda (este da última safra) foram às políticas que influenciaram nos preços de comercialização da safra do produto, entre os períodos de 2001-2002 e 2006-2007.

Na visão de Saraiva *et al.* (2014), é notável que a trajetória da componente sazonal é encontrada em distintas posições de equilíbrio geradas ao decorrer do tempo, entre a safra e a entre-safra, pois durante a temporada das safras, via de regra, a curva de oferta é deslocada para direita, tendo em vista que exista uma grande quantidade de produtos a serem comercializados no mercado. Por outro lado, no período da entre-safra, ocorre efeito oposto, pois os volumes de produtos para venda no comércio tendem a se reduzir deslocando a curva de oferta para a esquerda.

Em pesquisa realizada sobre o custo da cesta básica na cidade mineira de Lavras do Sul, que teve como enfoque o comportamento sazonal das observações, do período de janeiro de 1995 até dezembro de 2004, Carvalho, Ferraz e Sáfadi (2006) constataram a presença de sazonalidade nos dados referentes ao custo da Cesta Básica do município.

No mesmo estudo, por intermédio da verificação gráfica dos efeitos sazonais, percebeu-se elevação no custo da cesta de alimentos em épocas chuvosas, sendo o caso, dos meses de outubro a março. Por outro lado, os meses de abril a agosto foram os que registraram menores índices de chuva e, os que apresentaram menor custo na cesta de mercadorias no município. Diante das evidências de sazonalidade nos dados foi aplicado o método de dessazonalização X-12 $ARIMA(p,d,q)$ e o melhor modelo encontrado foi o $SARIMA(2,1,0) \times (0,1,1)_{12}$, comprovando que o método de dessazonalização apresentou confiança e também qualidade, sendo recomendado a utilização da análise sazonal em quaisquer séries de caráter econômica em estudo.

Estudos referentes aos preços da cesta básica, investigado por Silva, Pires e Ferraz (2015) em capitais dos estados nordestinos, “contaram” com a utilização do Teste de Raiz Unitária na intenção de encontrar a ordem de integração e o Teste de Causalidade de Granger para distinguir a direção de influência. Esta pesquisa teve o intuito de verificar a relação de causalidade entre os preços das capitais, na qual, segundo a interpretação dos testes, foi encontrada integração de ordem 1 e, relação causal dos preços entre todas as capitais em análise, além da dependência aparecer em maior parte de maneira bidirecional. Na mesma análise percebeu-se que a capital do estado do Rio Grande do Norte e a de Pernambuco apresentaram grande influência sobre as demais, pois à medida que ocorriam choques na economia destas cidades havia a possibilidade de também ocorrer nas outras capitais do Nordeste.

Em pesquisa feita por Barreto e Holanda (2007), na cidade de Fortaleza, estado do Ceará, modelou-se variáveis que influenciaram na variação do custo da cesta de alimentos, por intermédio de modelo de Regressão e de Co-integração referente ao custo mensal do cesto e, constatou-se que o valor da cesta de alimentação sofreu oscilações de todas as variáveis explicativas analisadas na equação de regressão múltipla, de modo que, a curto prazo, as variáveis que determinaram a cesta positivamente foram a do salário, do ICMS sobre os produtos alimentícios dividido pelo salário, da cesta básica defasada, da variação do preço do óleo diesel e da taxa de câmbio. Por outro lado, as que exerceram influência negativamente foram: o índice de desemprego e a pluviometria, de forma que a longo prazo o câmbio demonstrou resultados nada conclusivos. Quanto à análise de Co-integração, a mesma foi empregada para projetar um modelo de equilíbrio a longo prazo e um dinâmico a curto prazo, para que se pudesse concluir sobre o custo da cesta de alimentação.

Na investigação feita por Amorim, Sousa e Coronel (2013), sobre a Cesta Básica nas capitais da região Sul do Brasil, realizou-se técnicas de identificação espacial para analisar gastos com a cesta de alimentação básica. Por meio dos testes de raiz unitária de Dickey-Fuller Aumentado (ADF), o Dickey-Fuller GLS e o de Kwiatkowski-Phillips-Schmidt-Shin (KPSS) averiguou-se que os gastos mensais em alimentos foram integrados de ordem 1, e o teste de Co-integração, por outro lado, justificaram que os gastos mensais com a cesta de mercadorias nas capitais apresentaram integração espacial, em outras palavras, as variações tanto na oferta quanto na demanda dos mercados das três capitais da região Sul podiam afetar os preços umas das outras.

2.4. COMBINAÇÕES DE PREVISÃO

O fato de existir inumeráveis métodos para estimar a previsão, tem mudado e possibilitado a realização de pesquisas com maior ganho de desempenho na precisão. E, diversas são as situações, na qual, para análise de uma determinada série finita de dados são encontradas mais de uma possível previsão gerada por intermédio do ajuste individual dos modelos. Desta forma, o método de combinar previsão é uma notável possibilidade conjunta de ajustamento, para que se consiga obter um menor erro de previsão (CHATFIELD, 2000).

O interesse de Bates e Granger (1969) na análise de previsões combinadas deu-se por indagações a respeito de por qual razão realizar uma seleção entre modelos que apresentavam previsões próximas entre si, isto é, por que aceitar apenas a melhor previsão para inferir sobre decisões futuras, desprezando as demais? Mediante esses questionamentos, Bates e Granger (1969) constataram que se o objetivo principal era realizar melhoria na previsão, seria interessante aceitar outros modelos ao invés de um único, pois ao desconsiderar os demais modelos, informações individuais, únicas e independentes também estariam sendo desprezadas, como por exemplo: 1) uma previsão ajustada é embasada em variáveis ou informações que a outra desconsiderou e 2) a previsão ajustada pode realizar suposição diferente sobre a relação entre as variáveis. No entanto, pelo segundo caso não é regra que o método de combinação de previsão seja sempre melhor que a previsão individual, pois em diferentes pesquisas existe a possibilidade de um método ser mais adequado que o outro.

Na suposição teórica de Timmermann (2006), ao menos que seja identificado um modelo específico para previsão, que origine menores erros de previsão que os

concorrentes, o método de combinação será um ganho de diversificação por permitir combinar modelos individuais com objetivo de verificar ocorrência de menores erros de previsão. Ainda assim, caso o modelo individual produza melhores resultados em cada ponto do tempo que o método de combinação, o segundo método ainda pode ocasionar ganhos de diversificação, apesar de seu sucesso depender do quão bem os pesos da combinação possam ser determinados.

O método de combinar previsão proposto por Bates e Granger (1969), contribuiu para melhorar a qualidade do ajuste e diversificar a modelagem da série de interesse. Essa técnica, além de gerar novas possibilidades para predição, também redirecionou a visão tradicional que era de ajustar apenas um modelo dentre muitos outros.

Mediante as considerações de Bates e Granger, o recurso de combinar modelos para realizar previsões é um procedimento que ajuda na melhoria das medidas de Acurácias, pois esta técnica reúne os fatores que afetam as previsões dos modelos individuais, os quais são compactados para então realizar o procedimento de combinação de previsões em busca de resultados próximos ao valor real (COSTANTINI e PAPPALARDO, 2010).

A seguir é abordada a metodologia deste estudo.

3. METODOLOGIA

Nesta seção são descritas sete subseções utilizadas no desenvolvimento do presente estudo, a começar pela coleta de dados para a pesquisa, a ferramenta utilizada para processar os dados, o processo estacionário, o processo não estacionário, os modelos para previsões individuais, às combinações de previsões e por fim as medidas de Acurácias.

3.1. COLETA DE DADOS PARA PESQUISA

A necessidade da obtenção de dados, referentes aos preços cobrados na comercialização de produtos alimentícios essenciais para nutrição da população, direcionaram o Departamento Intersindical de Estatística e Estudos Socioeconômicos (DIEESE) a realizar regularmente a Pesquisa Nacional da Cesta Básica de Alimentos (PNCBA) e implantar a coleta de preços dos produtos de gêneros alimentícios em São Paulo. E, a partir 1959, o monitoramento dos preços tornaram-se objetos de análise, efetuadas por intermédio de operações estatísticas com finalidade de avaliar (medir) a variação do Índice de Custo de Vida (ICV) para verificar a situação da população paulista (DIEESE, 2019).

Atualmente, o Departamento Intersindical de Estatística e Estudos Socioeconômicos (DIEESE) expandiu-se e realiza a coleta de preços dos produtos da Cesta Básica em todas as capitais brasileiras, e inclusive no Distrito Federal. E, a coleta conduzida pelo Departamento, segue as normas metodológicas da Pesquisa de Orçamento Familiar (POF) do Instituto Brasileiro de Geografia e Estatística (IBGE).

Fundamentado na POF, a instituição analisa os Gastos médios mensais das famílias em produtos das Cestas Básicas, adquiridas nas 26 capitais e no Distrito Federal, sendo as coletas realizadas por membros dos escritórios do DIEESE, com levantamento dos dados realizados em estabelecimentos comerciais e até mesmo com pessoas físicas (caso dos vendedores ambulantes), mediante a um cadastro aplicado e aprovado pela instituição (DIEESE, 2016).

Para execução do presente estudo foram utilizadas observações de natureza secundária, disponível e coletadas do banco de dados do Departamento Intersindical de Estatística e Estudos Socioeconômicos (<https://www.dieese.org.br/cesta/>). A sequência de dados obtidos para realização desta pesquisa teve como característica a exploração

das variações do Gasto médio mensal no consumo da Cesta Básica de Alimentos de Porto Alegre. Na qual, foram verificados um conjunto de dados referentes aos períodos de julho de 1995 a fevereiro de 2019, totalizando 284 observações³ para análise, os quais em sua forma ordenada possibilitou uma breve visualização do comportamento dos Gastos médios presentes na série ao longo de sua trajetória.

Como a variável em estudo dispõe de dados monetários, ou seja, as observações são referentes à moeda brasileira, foi necessário deflacionar os Gastos médios em períodos mensais. Sendo que, para realizar a deflação das observações foi preciso trabalhar com os Números Índices pertencentes ao Índice Nacional de Preços ao Consumidor Amplo (IPCA) da região metropolitana de Porto Alegre, dos períodos de julho de 1995 a fevereiro de 2019, disponível na base de dados do IBGE (<https://www.ibge.gov.br>, mais especificamente <http://bit.do/fdRnY>). Tendo, a partir de julho de 1995 os valores nominais deflacionados mensalmente e o período base a ser considerado foi o mês de fevereiro de 2019.

Com base nos números índices (I_i e I_j) referentes aos períodos já mencionados e coletados⁴ do site do Instituto Brasileiro de Geografia e Estatística (IBGE), foram deflacionadas as observações coletadas da base de dados do Departamento Intersindical de Estatística e Estudos Socioeconômicos (DIEESE) e, calculada por meio da fórmula,

$$Pr_{i,j} = \left(\frac{P_i}{I_i} \right) * I_j. \quad (1)$$

em que:

i – representa os períodos mensais da série;

j – representa o período base mensal da série;

$Pr_{i,j}$ – é o preço real da Cesta Básica de Alimentos do período i em relação ao período j ;

P_i – é o preço nominal da Cesta Básica no período i ;

I_i – são os números índices dos preços médios nos períodos i , que representa os meses de julho de 1995 a janeiro de 2019;

I_j – representa o número índice de preço médio no período j , ou seja, é o índice de preço do mês base, que neste estudo é denotado pelo mês de fevereiro de 2019.

³ As observações em estudo foram escolhidas com a finalidade de analisar as variações dos gastos médios após o Plano Real, porém os doze meses anteriores a julho de 1995 não foram incluídos nesta pesquisa por apresentarem fortes variações causadas pelo início do Plano.

⁴ Ver Anexo A do presente estudo.

Os produtos alimentícios que compõem a Cesta Básica de Porto Alegre têm as quantidades equivalentes de cada produto composta na Região 3 do levantamento de dados da pesquisa praticada pelo DIEESE (2016). E essas mercadorias propostas para o levantamento das observações conduzidas pelo DIEESE seguem o Decreto-Lei nº 399, de 30 de abril de 1938, e o conjunto formado pelos 13 itens essenciais para alimentação foram divididos e distribuídos em quatro categorias que melhor se enquadram cada produto, mostradas na Tabela 1.

Tabela 1 – Produtos alimentícios que compõem a Cesta Básica de Porto Alegre, RS, pesquisada pelo DIEESE.

Bovinos		Grãos		Hortifrúti		Industrializados	
Carne	6,6 kg	Arroz	3,0 kg	Banana	90 un.	Açúcar	3,0 kg
Leite	7,5 lit.	Feijão	4,5 kg	Batata	6,0 kg	Banha/Óleo	900 gr.
				Tomate	9,0 kg	Café em pó	600 gr.
						Farinha	1,5 kg
						Manteiga	750 gr.
						Pão Francês	6,0 kg

Fonte: Elaborado com base na metodologia do DIEESE (1993).

A distribuição dos produtos na Tabela 1 foi apenas para identificar qual categoria cada um dos itens pertence, e o que se verificou foi que das 13 mercadorias que compõem a Cesta Básica de Alimentos, seis passam por um processo industrial antes de serem comercializadas.

3.2. FERRAMENTAS UTILIZADAS

Para os procedimentos de análise utilizou-se o *software* livre *R Core Team* (2018), o qual gerou modelos do conjunto de observações da variável gastos médio no consumo da Cesta Básica de Porto Alegre, RS. Este *software* é uma ferramenta programável que possibilita a realização de cálculos estatísticos por meio de pacotes criados e disponibilizados para gerar resultados gráficos e numéricos de modelos e testes simples e complexos.

A seguir, foi realizada uma breve descrição sobre o processo estacionário.

3.3. PROCESSO ESTACIONÁRIO

Para realizar o procedimento de estimação de uma série temporal é necessário recordar o conceito de estacionaridade, o qual identifica um processo estacionário como sendo aquele em que a série se desenvolve no tempo de forma aleatória e em torno de uma média constante, reproduzindo de certo modo a estabilidade dos dados da amostra (MORETTIN e TOLOI, 2004). Deste modo, é possível inferir sobre os parâmetros estimados fundamentando-se em um processo estocástico (BUENO, 2008), afim de, se chegar a resultados conclusivos da variável em questão.

Formalmente, um processo estocástico denotado por $\{Z_t\}_{t \in \mathbb{Z}}$ é dito estacionário, caso o mesmo se desenvolva no tempo e a origem das observações seja irrelevante. Em outros termos, significa que sendo $\{Z_t\}_{t \in \mathbb{Z}}$ um processo estocástico estacionário, as propriedades de $Z_{(t+k)}$, para todo k , é idêntico aos de $\{Z_t\}_{t \in \mathbb{Z}}$ (MORETTIN e TOLOI, 2004).

Na análise de séries temporais, existem dois casos de processos estocásticos estacionários, sendo: um deles conhecido como processo estacionário fraco ou de segunda ordem e o outro como processo forte ou estritamente estacionário.

Pela definição do primeiro caso, um processo estocástico $\{Z_t\}_{t \in \mathbb{Z}}$ será estacionário fraco ou de segunda ordem quando as três condições abaixo forem satisfeitas.

- $E\{Z_t\} = \mu_t = \mu$, constante, para todo $t \in \mathbb{Z}$;
- $E\{Z_t^2\} = \sigma^2 < \infty$, para todo $t \in \mathbb{Z}$;
- $\gamma_{(t,t+k)} = Cov\{Z_t, Z_{(t+k)}\}$, é uma função de k , para todo $t \in \mathbb{Z}$.

E, pela definição do segundo caso, um processo estocástico $\{Z_t\}_{t \in \mathbb{Z}}$ é dito forte ou estritamente estacionário, somente se não houver mudança em todas as distribuições finito-dimensionais, isto é, caso as mesmas se mantenham inalterada em relação ao tempo, como:

$$F(z_1 \dots z_n; t_1 + k \dots t_n + k) = F(z_1 \dots z_n; t_1 \dots t_n), \quad (2)$$

para cada $t_1 \dots t_n, k$ pertencente a \mathbb{Z} .

Isso justifica que todas as distribuições unidimensionais não variam ao longo do tempo e, portanto, a média μ_t e a variância σ_t^2 , permanecem constantes, como:

$$\mu_t = \mu \quad \text{e} \quad \text{Var}(Z_t) = \sigma^2, \quad (3)$$

em todo $t \in \mathbb{Z}$.

3.3.1. Análise gráfica

Entre inúmeros recursos utilizados para verificação e avaliação de um conjunto de dados, uma ferramenta muito útil nas distintas áreas do conhecimento científico é o método gráfico, por ser constituído com saída de aparência visual e normalmente ser o primeiro procedimento de demonstração do comportamento da série, o mesmo possibilita a inspeção do que acontece com as observações em questão, de modo claro e dinâmico.

A utilização da ferramenta gráfica é essencial para uma boa análise, já que, por intermédio da elaboração gráfica inúmeras informações particulares da série podem ser descritas, como por exemplo, *outliers*, variabilidade, tendência, sazonalidade, normalidade, etc. (MORETTIN e TOLOI, 2004). Uma alternativa de visualização gráfica frequentemente utilizada em análises de séries temporais é o histograma, o qual permite verificar se os dados seguem uma distribuição normal ou qualquer outra distribuição.

Para testar a normalidade da amostra da série gastos médios mensais no consumo da Cesta Básica de Porto Alegre realizou-se o seguinte teste.

3.3.2. Teste de Normalidade de Shapiro-Wilk

O teste de Shapiro-Wilk tem como propósito fornecer um procedimento estatístico para classificar se a amostra em estudo segue uma distribuição normal, ou seja, se $N(0, \sigma^2)$. Este teste é arranjado dividindo-se o quadrado de uma associação linear de uma ordem estatística amostral pela estimativa simétrica da variância.

$$W = \frac{(\sum_{t=1}^n a_t z_t)^2}{\sum_{t=1}^n (z_t - \bar{z})^2}. \quad (4)$$

em que: W denota o quanto os dados são normalizados, a_t representa um coeficiente linear, z_t é o valor da variável em estudo e \bar{z} representa a média do valor da variável. Sendo o teste de Shapiro-Wilk uma técnica para verificação de normalidade da, este teste condicionado à análise de amostras de até 2000 observações (ROYSTON, 1982).

As hipóteses a serem testadas são as seguintes:

H_0 – A amostra segue distribuição normal;

H_1 – A amostra não segue distribuição normal.

A verificação Shapiro-Wilk para análise de normalidade dos resíduos é calculado pelo *software R Core Team* (2018), por meio da seguinte rotina, *shapiro.test*, e pacote *stats*, demonstrando se a saída do valor-p do conjunto de dados apresenta significância quanto a normalidade.

Na sequência são descritos os procedimentos técnicos para verificação de um ruído branco da série, gastos médios mensais no consumo da Cesta Básica.

3.3.3. Análise de Resíduos

Para identificar se um modelo está corretamente especificado, é essencial avaliar os resíduos amostrais $\hat{\varepsilon}_t$ do modelo estimado, os quais são estimativas do ruído branco. Deste modo, as autocorrelações das observações não devem apresentar significância, para que assim o modelo esteja adequado.

Para identificar se o conjunto de observações amostrais resulta em um ruído branco, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é necessário verificar se:

- $E(\varepsilon_t) = 0; \forall t;$
- $E(\varepsilon_t^2) = \sigma^2; \forall t;$
- $E(\varepsilon_t \varepsilon_{t-j}) = 0, \text{ para todo } j \neq 0.$

isto é, uma sequência $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é identificado como sendo um ruído branco, se cada valor obter média zero, variância constante e não existir correlação entre os valores da série, ou seja, a autocorrelação deve ser igual a zero (BUENO, 2008).

Portanto, para verificar se a série deriva das propriedades de um processo ruído branco é importante analisá-la por meio das funções de autocorrelações amostrais e pelo teste de Ljung-Box, os quais são descritos a seguir.

3.3.3.1. Função de Autocorrelação Amostral

Considerando um processo $\{Z_t\}_{t \in \mathbb{Z}}$ qualquer, a vantagem de realizar análise da função de autocorrelação amostral em inúmeras defasagens por intermédio do Correlograma, justifica-se pela possibilidade da extração de informações da série, como: aleatoriedade nos dados ou se possuem tendência ou sazonalidade, a ponto de reproduzir um ruído branco, gerando então, um modelo apropriado para o conjunto de observações (BUENO, 2008).

Como normalmente só é possível investigar a amostra de determinada população, na prática é a função de autocorrelação da amostra, $\hat{\rho}_k$ que é analisada. Deste modo, dado um conjunto de observações $\{Z_t\}_{t=1}^n$, é preciso encontrar o valor da Covariância da amostra na defasagem h , e da Variância amostral (GUJARATI e PORTER, 2011). Para o cálculo da Covariância amostral o procedimento é dado por:

$$\hat{\gamma}_k = \frac{\sum_{t=1}^{n-k} (z_t - \bar{z})(z_{t+k} - \bar{z})}{n} \quad (5)$$

E para encontrar a Variância amostral o procedimento é:

$$\hat{\gamma}_0 = \frac{\sum_{t=1}^n (z_t - \bar{z})^2}{n} \quad (6)$$

sendo que, $k = 0, 1, 2, \dots$ é o número de defasagens na amostra, e em ambas expressões n representa o tamanho da amostra e $\bar{z} = \sum_{t=1}^n \frac{z_t}{n}$ é a média amostral da série.

Então, a função de autocorrelação amostral com h defasagens resulta da divisão de (5) por (6) dada pela expressão,

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad (7)$$

E os resultados amostrais de $\hat{\rho}_k$ contra k , é mais bem distribuídos e visualizados pelo Correlograma. Em um processo estacionário Gaussiano, Bartlett em 1946 demonstrou que sendo $k > 0$ e $k + j > 0$ (WEI, 2006), a expressão apropriada é:

$$\begin{aligned} Cov(\hat{\rho}_k, \hat{\rho}_{k+j}) \cong \frac{1}{n} \sum_{i=-\infty}^{\infty} (\rho_i \rho_{i+j} + \rho_{i+k+j} \rho_{i-k} - 2\rho_k \rho_i \rho_{i-k-j} \\ - 2\rho_{k+j} \rho_i \rho_{i-k} + 2\rho_k \rho_{k+j} \rho_i^2), \end{aligned} \quad (8)$$

Neste processo, para uma quantidade grande da amostra representada por n , a função de autocorrelação amostral, $\hat{\rho}_k$ é aproximadamente normal com média ρ_k e variância expressa por,

$$Var(\hat{\rho}_k) \cong \frac{1}{n} \sum_{i=-\infty}^{\infty} (\rho_i^2 + \rho_{i+k} \rho_{i-k} - 4\rho_k \rho_i \rho_{i-k} + 2\rho_k^2 \rho_i^2), \quad (9)$$

Rearranjando (9), têm-se à aproximação do processo de *Bartlett's*, com $\rho_k = 0$, para $k > m$, como segue.

$$Var(\hat{\rho}_k) \cong \frac{1}{n} (1 + 2\rho_1^2 + 2\rho_2^2 + \dots + 2\rho_m^2). \quad (10)$$

Na realidade, $\rho_i (i = 1, 2, \dots, m)$ são desconhecidos e substituídos por suas estimativas amostrais $\hat{\rho}_i$ e apresentam erro padrão para grande defasagem representado na autocorrelação amostral $\hat{\rho}_k$ (WEI, 2006).

$$S_{\hat{\rho}_k} = \sqrt{\frac{1}{n} (1 + 2\hat{\rho}_1^2 + \dots + 2\hat{\rho}_m^2)}, \quad (11)$$

Portanto, o erro padrão amostral avalia a significância estatística de qualquer $\hat{\rho}_k$ em h defasagens. A justificativa de Bartlett caso uma série temporal seja puramente aleatória (origina um ruído branco), é que os coeficientes de correlação amostral serão aproximadamente $\hat{\rho}_k \sim N(0, 1/n)$. Sendo que, em situações que as observações amostrais são grandes, esses coeficientes seguem distribuição normal com média zero e

variância $\sigma_{\hat{\rho}_k}^2$ igual a um dividido pelo tamanho da amostra n , (GUJARATI e PORTER, 2011).

E, seguindo as propriedades de uma distribuição normal padrão, para um intervalo de confiança de 95%, para dada população, ρ_k o erro padrão para esse intervalo (GUJARATI e PORTER, 2011; WEI, 2006), é calculado por:

$$S_{\hat{\rho}_k} = \sqrt{\frac{1}{n}} \quad (12)$$

então,

$$Prob[\hat{\rho}_k - t_\alpha * (S_{\hat{\rho}_k}) \leq \rho_k \leq \hat{\rho}_k + t_\alpha * (S_{\hat{\rho}_k})] = 0,95 \quad (13)$$

em que, t_α representa o resultado da estatística t de Student, com $n - 1$ graus de liberdade. As hipóteses a serem testadas são:

H_0 : Se autocorrelação populacional for igual à zero, $\rho_k = 0$;

H_1 : Se autocorrelação populacional for diferente de zero, $\rho_k \neq 0$.

O seguinte teste é o da autocorrelação amostral.

3.3.3.2. Função de Autocorrelação Parcial Amostral

Dado um conjunto de dados, pela regra de Cramer para $k = 1, 2, \dots$, a função de autocorrelação parcial, Φ_{kk} (WEI, 2006), é expressar por:

$$\Phi_{kk} = \frac{\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & \cdot & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{bmatrix}}{\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & \cdot & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{bmatrix}} \quad (14)$$

A função de autocorrelação parcial da amostra é encontrada por intermédio da substituição de ρ_i pela estimação de $\hat{\rho}_i$ visto na expressão (14), e denotado pela simbologia $\hat{\Phi}_{kk}$. A FACP amostral estimada foi representada por Durbin em 1960, por meio da equação (15), pois para evitar o desenvolvimento de cálculos longos e complexos envolvendo os determinantes de um elevado número de k (14), o artifício encontrado foi iniciar igualando $\hat{\Phi}_{11} = \hat{\rho}_1$ para então calcular e obter $\hat{\Phi}_{kk}$ (WEI, 2006), como segue:

$$\hat{\Phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\Phi}_{kj} \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^k \hat{\Phi}_{kj} \hat{\rho}_j}, \quad (15)$$

portanto, temos:

$$\hat{\Phi}_{k+1,j} = \hat{\Phi}_{kj} - \hat{\Phi}_{k+1,k+1} \hat{\Phi}_{k,k+1-j}, \quad \text{com } j = 1, \dots, k \quad (16)$$

O procedimento metodológico, também é valido para calcular a função de autocorrelação parcial teórica $\hat{\Phi}_{kk}$.

Em 1949, Quenouille constatou que na hipótese de um processo implícito constituir uma sequência de ruído branco, então a variância de $\hat{\Phi}_{kk}$ pode ser aproximada por:

$$\text{Var}(\hat{\Phi}_{kk}) \cong \frac{1}{n}. \quad (17)$$

portanto, é possível utilizar $\pm 2/\sqrt{n}$ como limites críticos na função de autocorrelação parcial, $\hat{\Phi}_{kk}$ para testar as hipóteses de um processo ser ruído branco.

O próximo teste a ser descrito é o de Ljung-Box, como segue.

3.3.3.3. Teste de Ljung-Box

Uma alternativa que não exige testar a significância estatística dos coeficientes de correlação pode ser realizada por meio da Estatística Q , a qual foi desenvolvida por Box e Pierce para elaborar o teste de hipótese de forma conjunta, da função de autocorrelação, ρ_k (GUJARATI e PORTER, 2011) calculada pela seguinte expressão:

$$Q = n \sum_{k=1}^m \hat{\rho}_k^2 \quad (18)$$

em que, n é o tamanho da amostra e h representa o tamanho da defasagem.

A Estatística Q de Box-Pierce é muito aplicada para avaliar se determinada série temporal é um ruído branco, e em grandes amostras, tem distribuição próxima da distribuição de Qui-quadrado, χ^2 , com m graus de liberdade e as hipóteses a serem testadas são:

H_0 : O somatório dos ρ_k é igual à zero, $\sum_{k=1}^m \rho_k = 0$;

H_1 : Pelo menos um dos ρ_k é diferente de zero, $\sum_{k=1}^m \rho_k \neq 0$.

O teste de Ljung-Box é uma generalização da Estatística Q de Box-Pierce, embora desenvolvida com o mesmo propósito da Estatística Q e, ambas seguirem a distribuição de Qui-quadrado, χ^2 , com m graus de liberdade, para uma grande quantidade de observações, a estatística de Ljung-Box é mais apropriada do que a Q quando o número de amostra é pequena (MORETTIN e TOLOI, 2004; BUENO, 2008; GUJARATI e PORTER, 2011). O teste de Ljung-Box é calculado pela equação:

$$LB = n(n+2) \sum_{k=1}^m \left(\frac{\hat{\rho}_j^2}{n-k} \right) \sim \chi^2 m. \quad (19)$$

As seguintes hipóteses testadas são:

H_0 : O somatório dos ρ_k é igual à zero, $\sum_{k=1}^m \rho_k = 0$;

H_1 : Pelo menos um dos ρ_k é diferente de zero, $\sum_{k=1}^m \rho_k \neq 0$.

O teste de Ljung-Box para verificação do ruído branco no *software R Core Team* (2018) é executado pela rotina, *Box.test* e por intermédio do pacote *stats*.

A seguir, veremos uma breve descrição de um processo não estacionário.

3.4. PROCESSO NÃO ESTACIONÁRIO

Na análise de séries temporais, muitas das séries estudadas apresentam características não estacionárias, como é o caso de variáveis econômicas e financeiras que na prática exibem ciclos, tendências e padrões sazonais, além da média ser diferente de zero e a variância não ser constante ao longo do tempo.

Um procedimento prático bastante usual para tornar uma série estacionária, é tomar sucessivas diferenças da mesma, até que as observações sejam normalizadas, isto é, a série se estabilize com média zero e variância constante (MORETTIN e TOLOI, 2004; BUENO, 2008). Assim, é possível realizar previsões e inferir sobre os parâmetros estimados.

O procedimento a seguir é utilizado para verificar se há ocorrência de tendência na série, sendo as observações testadas pela metodologia de Wald-Wolfowitz.

3.4.1. Teste para verificação de tendência de Wald-Wolfowitz

Antes de aplicar qualquer procedimento de estimação referente às observações, é viável verificar se há resquício da componente tendência na série, pois havendo a condição que a série constata uma componente sazonal, além da tendência, será imprescindível a eliminação da sazonalidade para posteriormente testar a presença de tendência nos dados. Caso queiramos testar a sazonalidade, o inverso também vale, isto é, elimina-se a tendência para realizar o teste (MORETTIN e TOLOI, 2004).

O teste de Wald-Wolfowitz possibilita a verificação de tendência em uma série independente e identicamente distribuída (*i.i.d*). Esse teste é calculado baseando-se na mediana amostral, em outras palavras, supondo N observações $\{Z_t\}_{t=1}^N$, de um conjunto de dados, onde m representa a mediana dos valores, e cada valor de Z_t é denotado por A , se for maior ou igual a m , e B se for menor que m .

Então, $N = (n_1 \text{ pontos } A) + (n_2 \text{ pontos } B)$. Sendo a estatística utilizada:

$$T_1 = \text{número total de sequências (grupos de símbolos iguais)}.$$

As hipóteses do teste são as seguintes:

H_0 – Não há existência de tendência;

H_1 – Há existência de tendência.

Este teste analisa a hipótese de independência da série, examinando se o conjunto composto de sucessivas observações ocorre de maneira aleatória. Sendo uma amostra $\{x_1, \dots, x_N\}$ de tamanho N maiúsculo, com diferenças $\{x_1, \dots, x_N\}'$ referentes às observações x_1 e a média da amostra \bar{x} , para esse teste a estatística é expressa por:

$$R = \sum_{i=1}^{N-1} x_1'x_{i+1}' + x_1'x_N' \quad (20)$$

Considerando as observações independentes, é possível presumir que a expressão R seja normal com média e variância R , representada da seguinte forma:

$$E(R) = \frac{2n_1n_2}{N} + 1. \quad (21)$$

$$\sigma^2(R) = \sigma_R^2 = \frac{2n_1n_2(2n_1n_2 - N)}{(N)^2(N_1 - 1)}. \quad (22)$$

A rotina executada para o teste de verificação de tendência de Wald-Wolfowitz no *software R Core Team* (2018) é, *runs.test*, gerada por intermédio do pacote *randtests version 1.0*. No teste de Wald-Wolfowitz, os dados são transformados em um vetor dicotômico (a variável resposta assume somente dois valores) e o limite de referência usado na execução dos procedimentos de verificação é a mediana da amostra, que resulta do caso especial do teste, com $n_1 = n_2$, testado acima e abaixo da mediana. E os valores alternativos que definem a hipótese alternativa são *two.sided*, *left.sided* e *right.sided*. Empregando a hipótese alternativa *right.sided*, a hipótese contrária de aleatoriedade é testada contra uma tendência, e caso se utilize a hipótese alternativa *left.sided*, a hipótese nula de aleatoriedade é testada contra uma correlação serial negativa de primeira ordem.

A tomada de diferença é uma técnica (estacionária) que permite remover a tendência, seja ela estocástica ou determinista encontrada em um modelo. Porém, se Z_t for um processo integrado de ordem d , é preciso tomar a d -ésima diferença para que ocorra estacionariedade. No entanto, é de suma relevância compreender como se estimar uma série com tendência determinística, procedendo da seguinte forma.

Sendo $\{Z_t\}_{t \in \mathbb{Z}}$ um conjunto de observações variando em função do tempo, e composto por tendência determinística, é necessário estimar-se $\{Z_t\}_{t \in \mathbb{Z}}$ contra o tempo e guardar os resíduos, estes por sua vez, depois de armazenados compõem uma nova série que precisará ser modelado de forma separada.

O passo seguinte testar se a série possui raiz unitária em seu conjunto de dados.

3.4.2. Testes de Raízes Unitárias

Existem vários testes que possibilitam a detecção de raiz unitária na série, porém neste estudo foram introduzidos os mais conhecidos e descritos em trabalhos científicos e na literatura. A começar pelo teste de Dickey-Fuller Aumentado.

3.4.2.1. Teste de Dickey-Fuller Aumentado (ADF)

O teste de Dickey-Fuller Aumentado (ADF) foi desenvolvido para reparar uma falha existente no teste de Dickey-Fuller (criado em 1979), pois este considerava o erro como sendo um ruído branco, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ (é um processo ruído branco), com média (μ) zero e variância σ_ε^2 , que regularmente é descrito (o erro) como um processo estacionário, que se não corrigido poderia acarretar em distorções no poder do teste unitário (BUENO, 2008).

O propósito do teste ADF para correção do problema mencionado, era o de estimar o modelo por meio de variáveis auto-regressivas, possibilitando a correção do desvio do valor correto da estatística, em outras palavras, significa achar os desvios de z_t em relação às suas médias, para que se possa deslocar a distribuição de α na direção de zero, se acaso H_0 for verdadeiro, dadas a seguintes hipóteses, a serem testadas:

H_0 – Existe raiz unitária (a série não é estacionária);

H_1 – Não existe raiz unitária (a série é estacionária).

O modelo pode ser descrito da seguinte forma:

$$Z_t = \theta_0 \sum_{i=1}^p \phi_i Z_{t-i} + \varepsilon_t, \quad (23)$$

ou, então

$$\nabla Z_t = \theta_0 + \phi_1^* Z_{t-1} + \phi_2^* Z_{t-1} + \dots + \phi_p^* \nabla Z_{t-p+1} + \varepsilon_t, \quad (24)$$

em que, $\phi_1^* = \sum_{i=1}^p \phi_i - 1$, $\phi_j^* = -\sum_{i=j+1}^p \phi_i$, $j = 2, \dots, p$.

Caso o polinômio auto-regressivo $\phi(B)$ possuir uma raiz unitária, logo $\phi(1) = 0$, isto é, $1 - \phi_1 - \phi_2 - \dots - \phi_p = 0$, ou $\sum_{i=1}^p \phi_i = 1$ e por conseguinte $\phi_1^* = 0$. O que demonstra que ao testar a hipótese de que o polinômio auto-regressivo tem uma raiz unitária, equivale a testar a hipótese que $\phi_1^* = 0$.

Sabendo-se que por ϕ_1^* é possível estimar como sendo um coeficiente de Z_{t-1} na regressão de mínimos quadrados de ∇Z_t sobre $1, Z_{t-1}, \nabla Z_{t-1}, \dots, \nabla X_{t-p+1}$, para N maiúsculo, as estatísticas,

$$N(\hat{\phi}_1^* - 1) \quad \text{e} \quad \hat{t}_\mu = \hat{\phi}_1^* / \widehat{e \cdot p}(\hat{\phi}_1^*) \cdot \quad (25)$$

e o teste composto por \hat{t}_μ é conhecido como de Dickey-Fuller Aumentado (ADF).

O propósito do teste ADF é ajustar o teste original DF, para que assim se possa eliminar a possibilidade de correlação serial nos termos de erros, quando aplicado diferença na série (GUJARATI e POTER, 2011).

A rotina executada para este teste no *software R Core Team* (2018) é o, *adf.test*, com auxílio de recurso como o pacote, *tseries*. Pelo teste ADF, a equação geral de regressão AD, possui uma tendência constante e linear utilizando a estatística t para o coeficiente auto-regressivo de primeira ordem, AR(1), sendo que, o número de *lags* envolvido na expressão é denotado por k . E o valor referente ao truncamento é $(n - 1)^{1/3}$, onde n é o tamanho da série temporal e representa o limite superior sugerido, o qual, o número de defasagens k , tem que proceder de forma crescente, de acordo com o crescimento do tamanho da amostra um processo ARMA(p, q), Banerjee *et al.* (1993).

Para entender os dois próximos testes que definem a estacionariedade da série, é preciso compreender o estimador de Newey e West.

Newey e West⁵ desenvolveram em 1987, um método viável capaz de estimar a matriz de variância e covariância, de forma consistente em casos com presença de

⁵ Maiores informações, consultar Whitney K. Newey and Kenneth D. West. A. *Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix*. *Econometrica*, Vol. 55, Nº. 3 (May, 1987), pp. 703-708.

heteroscedasticidade e autocorrelação. Este método denominado de estimador de Newey-West, foi criado e aplicado na correção dos erros do método de Mínimos Quadrados Ordinários (MQO), sendo bastante útil em procedimentos com grandes amostras. Dado o estimador,

$$S_{N\ell}^2 = N^{-1} \sum_{t=1}^N \hat{\varepsilon}_t^2 + 2N^{-1} \sum_{j=1}^{\ell} \sum_{t=j+1}^N \hat{\varepsilon}_t \hat{\varepsilon}_{t-j}, \quad (26)$$

e portanto, sendo $S_{N\ell}^2$ definido como um estimador consistente para variância σ^2 dos resíduos de um modelo adequado, que não garante estimativa positiva para um conjunto de observações finitas (BANERJEE *et al.*; 1993). Rearranjando a expressão anterior, o estimador de Newey-West torna-se consistente e não negativo, sendo representado por:

$$\hat{S}_{N\ell}^2 = N^{-1} \sum_{t=1}^N \hat{\varepsilon}_t^2 + 2N^{-1} \sum_{j=1}^{\ell} w_{\ell}(j) \sum_{t=j+1}^N \hat{\varepsilon}_t \hat{\varepsilon}_{t-j}. \quad (27)$$

em que, $w_{\ell}(j) = 1 - j(\ell + 1)^{-1}$. Sendo, $\hat{S}_{N\ell}^2$ o estimador da variância σ^2 , N o tamanho da amostra ou última observação da série, ℓ o lag parâmetro de truncagem que serve para indicar que diferentes opções de ℓ acarretam em diferentes resultados e $\hat{\varepsilon}_t$ representa os resíduos.

O próximo teste é o de Phillips-Peron.

3.4.2.2. Teste de Phillips-Perron

O teste de Phillips-Perron é a implementação de uma correção não paramétrica ao teste de Dickey-Fuller, possibilitando a consistência do teste mesmo que se tenham variáveis defasadas dependentes e correlação serial nos erros. Desta forma, o teste de Phillips-Perron não necessita da especificação de um modelo com ordem autoregressivas considerável para corrigir a correlação serial dos resíduos (BUENO, 2008).

A estatística de Phillips-Perron é calculada pela expressão a seguir:

$$z_{t,\mu} = \hat{t}_{\mu} \left(\frac{\hat{\sigma}}{\hat{\nu}} \right) - \frac{1}{2} \left(\frac{\hat{\nu}^2 - \hat{\sigma}^2}{\hat{\nu} \sqrt{N^{-2} \sum_{t=1}^N z_{t-1}^2}} \right). \quad (28)$$

em que, os termos \hat{v}^2 (de longo prazo) e $\hat{\sigma}^2$ estimadores consistentes dos parâmetros da variância, $\hat{\tau}_\mu$ é a estatística de Dickey-Fuller e N é o número de observações.

Este teste verifica as seguintes hipóteses:

H_0 – Existe raiz unitária (a série não é estacionária);

H_1 – Não existe raiz unitária (a série é estacionária).

Na aplicação de Phillips-Perron para verificação da hipótese de raiz unitária nos dados, o mesmo não necessita de diferenciação das observações como diagnóstico no tratamento de autocorrelação serial (GUJARATI e POTER, 2011).

A rotina executada para este teste de raiz unitária no *software R Core Team* (2018) é, *pp.test*, com ajuda do pacote, *tseries*.

Este teste fundamenta-se na equação geral de Dickey-Fuller com tendência e constante linear, e a estatística de uso é $Z_{t/\alpha}$ para o coeficiente de primeira ordem, AR(1). E tendo como seu estimador da matriz de variância, σ^2 , o Newey-West, caso *lshort* for verdadeiro, o *lag* do parâmetro truncado será $(4 * (n/100)^{1/4})$, do contrário, o truncamento será $(12 * (n/100)^{1/4})$ sendo n o tamanho da série temporal (BANERJEE *et al.*, 1993).

O último teste aplicado neste estudo é conhecido como KPSS.

3.4.2.3. Teste de Kwiatkowski-Phillips-Schmidt-Shin (KPSS)

O teste de Kwiatkowski-Phillips-Schmidt-Shin (KPSS) foi elaborado para aprimorar o teste (de raiz unitária) criado por *Dickey-Fuller* (1979, 1981), visto que um problema encontrado na estatística de *Dickey-Fuller* era o baixo poder do teste, e para conseguir rejeitar a hipótese nula.

O teste de KPSS é denotado por:

$$KPSS = \sum_{t=1}^N \frac{S_t^2}{N^2 \hat{v}^2} \quad (29)$$

sendo assim, como no teste de Phillips-Perron o termo \hat{v}^2 denota um estimador da variância de longo prazo, S_t^2 é a soma parcial dos resíduos e N são as observações.

No teste *KPSS*, as hipóteses a serem testadas são:

H_0 – Não existe raiz unitária (a série é estacionária);

H_1 – Existe raiz unitária (a série não é estacionária).

Segundo Bueno (2008), na aplicação da estatística *KPSS* se tem como intenção associar o *KPSS* aos testes de raiz unitárias, para que se possa identificar a raiz de um conjunto de observações, na qual, os dados não são totalmente conclusivos. A rotina executada para este teste no *software R Core Team* (2018) é, *kpss.test*, por meio do pacote *tseries version 0.10-46*. O *KPSS* usa como estimador de σ^2 , o estimador de *Newey-West*, caso *lshort* for verdadeiro, o *lag* do parâmetro truncado será $(4 * (n/100)^{1/4})$, do contrario o truncamento será dado por $(12 * (n/100)^{1/4})$, onde n é o tamanho da série temporal.

A seguir são apresentados os modelos ajustados por meio da série temporal gastos médios como o consumo da Cesta Básica.

3.5. MODELOS PARA PREVISÕES DE SÉRIES TEMPORAIS

Nesta seção, foram descritos apenas os seis modelos que apresentaram parâmetros significativos e passaram pelos testes de normalidade dos resíduos e, ao nível de 5% de significância os modelos ajustados foram: $ARIMA(p,d,q)$, $ARFIMA(p,d,q)$, $SARIMA(p,d,q) \times (P,D,Q)_s$, $SARMA(p,q) \times (P,Q)_s$ e o de Suavização Exponencial (Aditivo e Multiplicativo) de *Holt-Winters*.

O primeiro modelo a ser descrito, se desenvolve de uma generalização da classe do processo estacionário $ARMA(p,q)$, sendo denominado na literatura como processo $ARIMA(p,d,q)$.

3.5.1. Modelo 1: Modelos Auto-Regressivos Integrados e de Médias Móveis – $ARIMA(p,d,q)$.

Esse modelo surgiu para solucionar casos de observações que manifestem ausência de estabilidade, ou seja, para procedimentos que os dados não apresentavam

estacionariedade, e pela necessidade de diferenciação do processo, $\{Z_t\}_{t \in \mathbb{N}}$ foi denominado integrado de ordem d e denotado por $Z_t \sim I(d)$, (MORETTIN, 2011).

Segundo Box, Jenkins e Reinsel (1994) o processo $ARIMA(p,d,q)$ para séries não estacionárias é uma classe de modelos de fundamental importância para realização de previsão e controle da série. Pois, o processo $ARIMA(p,d,q)$ é adequado para modelagem de séries de dados com e também sem tendência (BROCKWELL e DAVIS, 2002).

Definição 1: Dado um conjunto de observações, se d for um inteiro não negativo, $\{Z_t\}_{t \in \mathbb{N}}$ será considerado um processo $ARIMA(p,d,q)$, se e somente se, $W_t = \nabla^d Z_t$ seguir um modelo $ARMA(p,q)$ causal (BROCKWELL e DAVIS, 2002).

Pela definição 1, $\{Z_t\}_{t \in \mathbb{N}}$ satisfaz a equação de diferença expressa por:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d Z_t = \theta_0 + (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t. \quad (30)$$

A expressão geral (30) pode ser reduzida para (31):

$$\phi(B) \nabla^d Z_t = \theta(B) \varepsilon_t. \quad (31)$$

em que, $\phi(\cdot)$ e $\theta(\cdot)$ são os polinômios de ordens p e q , respectivamente, B é o operador de retardo, $\nabla^d = (1 - B)^d$ é o operador diferença não sazonal de ordem d , $\{Z_t\}_{t \in \mathbb{N}}$ é a série temporal de interesse e $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é o processo ruído branco.

Caso, $d \geq 1$ é possível adicionar uma tendência polinomial arbitrária de grau $(d - 1)$ na série temporal de interesse $\{Z_t\}_{t \in \mathbb{Z}}$, sem receio de violar a expressão (31) de diferença, e o modelo $ARIMA(p,d,q)$ será útil para representar observações com tendência. No entanto, o processo $ARIMA(p,d,q)$ também é apropriado para determinar séries com ausência de tendência, com exceção daqueles que apresentarem $d = 0$, pois é o caso em que a média de $\{Z_t\}_{t \in \mathbb{Z}}$ não é determinada pela expressão (31) e pode ser zero em particular. Uma vez que para $d \geq 1$, a expressão (31) determina as propriedades de segunda ordem de $(1 - B)^d Z_t$, porém não as de $\{Z_t\}_{t \in \mathbb{N}}$, a estimação de ϕ , θ , σ^2 e serão baseadas nas diferenças observadas de $(1 - B)^d Z_t$ (BROCKWELL e DAVIS, 2002).

No caso do processo apresentar uma constante, a expressão (31) será reescrita na forma (32), com adição da constante μ denominada de parâmetro da tendência.

$$\phi(B)\nabla^d Z_t = \mu + \theta(B)\varepsilon_t. \quad (32)$$

A finalidade de identificar um modelo integrado é o de estabelecer valores para as ordens p , d e q de um processo $ARIMA(p,d,q)$, e estabelecer as estimativas antecedentes dos parâmetros que são utilizadas no momento da estimação (MORETTIN e TOLOI, 2004). Para identificar as ordens (p,d,q) de um modelo $ARIMA(p,d,q)$ a ser ajustado, as funções estimadas de autocorrelação e autocorrelação parcial amostrais são dois métodos de avaliação que demonstram a representatividade das observações, de modo que essas funções sejam adequadas às quantidades teóricas, as quais não são conhecidas (MORETTIN e TOLOI, 2004).

O procedimento padrão executado para encontrar um modelo $ARIMA(p,d,q)$ no *software R Core Team* (2018) é gerado pela rotina *arima()*, e necessita do pacote *stats*. E essa rotina utiliza três métodos de ajuste para estimar os parâmetros do modelo. O primeiro estimador é o de soma dos quadrados condicionais (CSS), o seguinte é o de máxima verossimilhança (ML) e por fim, o de máxima verossimilhança (CSS-ML), isto é, os resultados iniciais encontrados pela soma dos quadrados condicionais (CSS) são utilizados para otimização dos valores finais. Para essa rotina, ao menos que a série tenha observações omissas, por padrão primeiro é executado a soma dos quadrados condicionais (CSS) para encontrar os resultados iniciais e após esse procedimento o modelo final é calculado pela estimação de máxima verossimilhança (CSS-ML). Além disso, a rotina *arima()* utiliza o critério de informação AIC como valor para selecionar o melhor modelo da estimação de máxima verossimilhança (ML).

Juntamente com a identificação do modelo é possível a obtenção de estimativas preliminares para os parâmetros do processo, pois os valores iniciais encontrados servem de base para a solução iterativa da estimação por máxima verossimilhança dos parâmetros, sendo tais estimativas originadas pelas autocorrelações amostrais da série $W_t = \nabla^d Z_t$ (MORETTIN e TOLOI, 2004).

Supondo um modelo $ARIMA(p,d,q)$ com seus parâmetros $p + q + 1$ no vetor $\xi = (\phi, \theta, \sigma_\varepsilon^2)$, com $\phi = (\phi_1, \dots, \phi_p)$ e $\theta = (\theta_1, \dots, \theta_q)$. Neste caso, quando $d > 0$, supõe-se que $\mu_\omega = 0$, do contrário, μ é mais um parâmetro que deve ser estimado com $p + q + 2$ parâmetros, tendo $\eta = (\phi, \theta)$.

Para estimação de ξ um método utilizado é o de máxima verossimilhança, tendo N observações Z_1, \dots, Z_N . Considerando a função de verossimilhança (EMV)

$L(\xi|Z_1, \dots, Z_N)$ como função de ξ . Os estimadores de máxima verossimilhança de ξ são valores que maximizam L ou $l = \log l$ (MORETTIN e TOLOI, 2004).

Ao tomar-se d diferenças para tornar a série estacionária o número de observações é escrita de forma que $n = N - d$ para W_1, \dots, W_n , em que, $W_t = \nabla^d Z_t$, e sendo o modelo ARMA(p, d) estacionário e invertível, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é escrito como:

$$\varepsilon_t = \tilde{W}_t - \phi_1 \tilde{W}_{t-1} - \dots - \phi_p \tilde{W}_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (33)$$

em que, $\tilde{W}_t = W_t - \mu_\omega$.

Pelo procedimento condicional, supõe-se que os $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ são normais, tendo a função densidade conjunta de $\varepsilon_1, \dots, \varepsilon_n$, na forma de:

$$f(\varepsilon_1, \dots, \varepsilon_n) = (2\pi)^{\frac{-n}{2}} (\sigma_\varepsilon)^{-n} \exp \left\{ - \sum_{t=1}^n \frac{\varepsilon_t^2}{2\sigma_\varepsilon^2} \right\} \quad (34)$$

Calculando $\varepsilon_1, \dots, \varepsilon_n$ com base na equação (33), é necessário partir da suposição que os dados p valores W_t e q valores $\{\varepsilon_t\}_{t \in \mathbb{Z}}$, os quais são escritos como W^* e ε^* , nesta ordem. Então, a partir das equações (33) e (34), a função de Verossimilhança condicional, a escolha dos W_t e $\{\varepsilon_t\}_{t \in \mathbb{Z}}$, pode ser encontrada por:

$$L(\xi|W, W^*, \varepsilon^*) = (2\pi)^{\frac{-n}{2}} (\sigma_\varepsilon)^{-n} \exp \left\{ - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^n (\tilde{W}_t - \phi_1 \tilde{W}_{t-1} - \dots - \phi_p \tilde{W}_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q})^2 \right\} \quad (35)$$

e ao aplicar o logaritmo em L , obtém-se,

$$\ell(\eta|W, W^*, \varepsilon^*) \propto -n \log \sigma_\varepsilon - \frac{S(\eta|W, W^*, \varepsilon^*)}{2\sigma_\varepsilon^2}, \quad (36)$$

em que,

$$S(\eta|W, W^*, \varepsilon^*) = \sum_{t=1}^n \varepsilon_t^2(\eta|W, W^*, \varepsilon^*), \quad (37)$$

que é a Soma de Quadrados condicional. Onde, ℓ e S são ditos condicionais ao serem escritos com asterisco, e isso demonstra que são condicionais a $W = (W_1, \dots, W_N)$, $W^* = (W_1^*, \dots, W_1^*)$, $\varepsilon^* = (\varepsilon_1^*, \dots, \varepsilon_q^*)$, como:

$$\ell_*(\xi) = -n \log \sigma_\varepsilon - \frac{S_*(\eta)}{2\sigma_\varepsilon^2}, \quad (38)$$

$$S_*(\eta) = \sum_{t=1}^n \varepsilon_t^2(\eta|W, W^*, \varepsilon^*). \quad (39)$$

É importante aqui destacar que maximizar $\ell_*(\xi)$ equivale a minimizar $S_*(\eta)$ e assim, os estimadores de máxima verossimilhança se tornam de mínimos quadrados e o estudo de $\ell_*(\xi)$ equivale ao de $S_*(\eta)$, (MRETTIN e TOLOI, 2004).

Quando $p = 1$ e $d = 0 = q$, temos um caso particular dos modelos ARIMA(p, d, q), denotado por AR(1). O estimador de máxima verossimilhança para calcular a amostra de forma condicional de um modelo AR(1), (HAMILTON, 1994), vem da expressão,

$$\begin{aligned} & f_{w_T, w_{T-1}, \dots, w_1}(w_T, w_{T-1}, \dots, w_1; \theta) \\ &= f_{w_1}(w_1, \theta) \prod_{t=2}^T f_{w_t|w_{t-1}}(w_t|w_{t-1}; \theta), \end{aligned} \quad (40)$$

A função de log verossimilhança, representada por $\mathcal{L}(\theta)$, pode ser identificada por meio da equação (33) e reescrita como:

$$\mathcal{L}(\theta) = \log f_{z_1}(z_1, \theta) + \sum_{t=2}^T \log f_{z_t|z_{t-1}}(z_t|z_{t-1}; \theta). \quad (41)$$

Podemos notar nitidamente que o valor expresso no parâmetro populacional θ , que maximiza a equação (40) é similar ao valor que maximiza a expressão (41).

Na equação seguinte,

$$\frac{1}{2} \log |\sigma^{-2} L' L| = \frac{1}{2} \log \{ \sigma^{-2T} |L' L| \}$$

$$\begin{aligned}
&= \frac{1}{2} \log \sigma^{2T} + \frac{1}{2} \log |L'L| \\
&= \left(\frac{-T}{2} \right) \log \sigma^2 + \log |L|,
\end{aligned} \tag{42}$$

L é uma matriz triangular inferior, e seu determinante é conhecido pelo produto dos termos ao longo da diagonal principal, $|L| = \sqrt{1 - \phi^2}$ de uma matriz (HAMILTON, 1994). Assim temos que;

$$\frac{1}{2} \log |\sigma^{-2} |L'L| = \left(\frac{-T}{2} \right) \log \sigma^2 + \log(1 - \phi^2). \tag{43}$$

O modelo a seguir é bem conhecido na literatura de séries temporais por sua característica de longa dependência.

3.5.2. Modelo 2: Modelos de Memória Longa ARFIMA(p, d, q).

Este modelo vem a ser a generalização do processo ARIMA(p, d, q), ou seja, é um modelo ARIMA(p, d, q) fracionário também conhecido como ARFIMA(p, d, q), e seu parâmetro de diferença, d assume valores não inteiros. Este processo era normalmente usado em séries hidrológicas, mas segundo Morettin e Tolo (2004), a mesma já é utilizada em dados de variáveis financeiras. Para um modelo ARFIMA(p, d, q), sua maior característica é a longa dependência quando $d \in (0, 0,5)$ e pequena dependência, quando $d \in (-0,5; 0,0)$.

Para todo número real $d > -1$ determina-se o operador de diferença fracionária da forma a seguir:

$$\begin{aligned}
(1 - B)^d &= \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k \\
&= 1 - dB + \frac{1}{2!} d(d-1)B^2 - \frac{1}{3!} d(d-1)(d-2)B^3 + \dots
\end{aligned} \tag{44}$$

Considera-se, $\{Z_t\}_{t \in \mathbb{Z}}$ um processo auto-regressivo fracionário integrado de médias móveis, ou ARFIMA(p, d, q), com $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$.

Definição 2: Seja $\{Z_t\}_{t \in \mathbb{Z}}$ um processo estocástico que satisfaz a seguinte expressão:

$$\phi(B)(1 - B)^d(Z_t - \mu) = \theta(B)\varepsilon_t. \quad (45)$$

em que: B é operador de retardo ou defasagem, $\phi(\cdot)$ e $\theta(\cdot)$ são polinômios de ordens p e q , Z_t é a série de interesse, μ é a média do processo e $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ representa o ruído branco.

O propósito de adotar esse processo para modelar um conjunto de observações, com procedimento de memória longa, vem a ser pelo efeito causado pelo parâmetro d em observações distantes, pois este decai hiperbolicamente (de forma suave) à medida que a distância aumenta. Por isso, d tende a ser apontado com intuito de especificar a estrutura de autocorrelação de ordens altas do conjunto de dados, enquanto $\phi(\cdot)$ e $\theta(\cdot)$ descrevem a estrutura de ordens baixas da série (MORETTIN e TOLOI, 2004).

Em relação à estacionariedade e invertibilidade, Hosking (1981), constatou que o processo ARFIMA(p, d, q), do modelo (44) é:

- (I) Estacionário caso $d < \frac{1}{2}$ e todas as raízes de $\phi(B) = 0$ estejam fora do círculo unitário;
- (II) Invertível caso $d > -\frac{1}{2}$ e todas as raízes $\theta(B) = 0$ estejam fora do círculo unitário.

Dados os Processo ARFIMA(p, d, q) (ver Definição 2), Hosking (1981), apresenta as condições para estacionariedade e invertibilidade destes processos além das suas funções densidade espectral e de autocorrelação, incluído a estrutura de decaimento destas funções que está diretamente relacionada com a relação entre as variáveis aleatórias de $\{Z_t\}_{t \in \mathbb{Z}}$, o que demonstra memória longa do processo.

Quanto à função de autocorrelação e densidade espectral, Hosking em 1981, determina que um conjunto de observações $\{Z_t\}_{t \in \mathbb{Z}}$, também dado por (44), apresenta estacionariedade e invertibilidade, e sendo $f(\lambda)$ uma função densidade espectral de $\{Z_t\}_{t \in \mathbb{Z}}$, então;

- (I) $\lim_{\lambda \rightarrow 0} \lambda^{2d} f_Z(\lambda)$ existe e é finito,
- (II) $\lim_{k \rightarrow \infty} k^{1-2d} \rho_Z(k)$ existe e é finito.

Em procedimentos semi-paramétricos, a procedência para estimar os parâmetros são desenvolvidas em: (I) estimar o parâmetro de memória longa, d , por meio de um

modelo de regressão do logaritmo da função do periodograma, e logo após estimar os parâmetros auto-regressivos de médias móveis.

A estimação utilizada no procedimento $ARFIMA(p,d,q)$ é o estimador de máxima verossimilhança aproximada, que necessita do cálculo de n integrais para cada valor do teste de *Whittle's* (BERAN, 1994).

A seguir apresentamos o modelo $SARMA(p,q) \times (P,Q)_s$ que é um caso particular do modelo $SARIMA(p,d,q) \times (P,D,Q)_s$.

3.5.3. Modelo 3: Modelo Sazonal Auto-regressivo e de Médias Móveis – $SARMA(p,q) \times (P,Q)_s$.

Este modelo é um caso particular do modelo $SARIMA(p,d,q) \times (P,D,Q)_s$, porém quando as observações forem estacionárias não é necessário aplicar procedimento de diferença sazonal e não sazonal, logo, $d = 0 = D$. A rotina executada no *software R Core Team* (2018) para o modelo $SARIMA(p,d,q) \times (P,D,Q)_s$, quando apresentam $d = 0 = D$ gera o caso particular $SARMA(p,q) \times (P,Q)_s$.

O próximo modelo é uma generalização do modelo $ARIMA(p,d,q)$, proposto por Box-Jenkins e, é utilizado para modelagem de séries que apresentam sazonalidade.

3.5.4. Modelo 4: Modelos Auto-Regressivos Integrados e de Médias Móveis Sazonais – $SARIMA(p,d,q) \times (P,D,Q)_s$

Este modelo nada mais é do que uma generalização do processo $ARIMA(p,d,q)$ sazonal multiplicativo, que possui uma parte livre de sazonalidade, com parâmetros (p,d,q) , e uma outra parte sazonal, com parâmetros $(P,D,Q)_s$.

Definição 3: Seja $\{Z_t\}_{t \in \mathbb{Z}}$ um processo estocástico satisfazendo a equação a seguir:

$$\emptyset(B)\Phi(B^S)(1-B)^d(1-B^S)^D(Z_t - \mu) = \theta(B)\Theta(B^S)\varepsilon_t. \quad (46)$$

em que: $\nabla^d = (1-B)^d$ é operador diferença não sazonal de ordem d , $\nabla_s^D = (1-B^S)^D$ é operador diferença sazonal de ordem D , μ representa a média do processo, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ denota o processo ruído branco, $s \in \mathbb{N}$ é a sazonalidade, B é o operador de retardo ou de defasagem, em outras palavras, $B^j(Z_t) = Z_{t-j}$ e $B^S(Z_t) = Z_{t-sj}$, para $j \in \mathbb{N}$, $\emptyset(\cdot)$,

$\theta(\cdot)$, $\Phi(\cdot)$, $\Theta(\cdot)$ os quais são polinômios de ordem p , q , P e Q , representados pelas expressões (47) e (48):

$$\phi(z) = \sum_{\ell=1}^p (-\phi_{\ell})z^{\ell}, \quad \theta(z) = \sum_{m=1}^q (-\theta_m)z^m, \quad (47)$$

$$\Phi(z) = \sum_{r=1}^P (-\phi_r)z^r, \quad \Theta(z) = \sum_{l=1}^Q (-\theta_l)z^l, \quad (48)$$

em que: ϕ_{ℓ} , $1 \leq \ell \leq p$, θ_m , $1 \leq m \leq q$, ϕ_r , $1 \leq r \leq P$, θ_l , $1 \leq l \leq Q$, são constantes reais e $\phi_0 = \Phi_0 = -1 = \theta_0 = \theta_0$. Portanto, $\{Z_t\}_{t \in \mathbb{Z}}$ vem a ser um processo sazonal de ordem $(p,d,q) \times (P,D,Q)_s$, com sazonalidade s , denotado por $\text{SARIMA}(p,d,q) \times (P,D,Q)_s$, sendo d o grau de diferenciação e D o grau de diferenciação sazonal.

A função padrão $\text{arima}(\cdot)$ executada no *software R Core Team* (2018), é generalizada para rotina $\text{auto.arima}(\cdot)$, executada por intermédio do pacote *forecast*, essa rotina tem por finalidade encontrar um modelo $\text{SARIMA}(p,d,q) \times (P,D,Q)_s$. A rotina $\text{auto.arima}(\cdot)$, utiliza como método de ajuste o estimador de máxima verossimilhança (ML) ou a soma dos quadrados condicionais (CSS). Porém, ao menos que a série tenha observações omissas, essa rotina calcula por padrão primeiro a soma dos quadrados condicionais para encontrar os resultados iniciais e após esse procedimento o modelo final é calcula pela estimação de máxima verossimilhança (CSS-ML). Além disso, a rotina $\text{auto.arima}(\cdot)$, também utiliza os critérios de informações como aproximação para escolha do modelo AIC, AICc e BIC.

Para as combinações dos quatro modelos temos:

3.5.4.1. Métodos de Previsão utilizando a representação Auto regressiva infinita.

Nesta seção apresentamos o método de previsão para os modelos $\text{ARIMA}(p,d,q)$, $\text{ARFIMA}(p,d,q)$, $\text{SARIMA}(p,d,q) \times (P,D,Q)_s$ e $\text{SARMA}(p,q) \times (P,Q)_s$, que utilizam à representação $\text{AR}(\infty)$.

Na análise de uma série temporal, a verificação de um conjunto de dados ordenados no tempo tem por objetivo possibilitar que se encontre um modelo matemático, capaz de ajustar observações passadas para estimar resultados futuros denominados como previsão.

Segundo Wei (2006), o objetivo central dos estudos envolvendo modelos para análise de séries temporais de um conjunto de observações, são os resultados estimados dos métodos de previsões. Em outras palavras, o propósito geral a ser alcançado são os resultados previstos de determinada série.

Para o modelo $ARIMA(p,d,q)$ a expressão geral considerada quando $d \neq 0$ é:

$$\phi(B)(1-B)^d Z_t = \theta(B)\varepsilon_t, \quad (49)$$

em que, $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ e $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$, são polinômios de graus p e q , respectivamente (BERAN, 1994). Pela forma invertível, a equação de previsão do erro quadrático médio para o modelo $ARIMA(p,d,q)$ e expressa da seguinte forma:

$$\varepsilon_t = \frac{\phi(B)}{\theta(B)} (1-B)^d Z_t, \quad (50)$$

Sendo o modelo $SARIMA(p,d,q) \times (P,D,Q)_S$ a generalização propriamente dita de um processo $ARIMA(p,d,q)$ em sua forma invertível, um processo $SARIMA(p,d,q) \times (P,D,Q)_S$ expressa sua previsão de erro quadrático médio por:

$$\varepsilon_t = \frac{\Phi(B^S)\phi(B)}{\Theta(B^S)\theta(B)} (1-B^S)^D (1-B)^d Z_t, \quad (51)$$

O modelo $ARFIMA(p,d,q)$ ou de memória longa, é denotado pela sua equação inversa de previsão para o erro quadrático médio como;

$$\varepsilon_t = \sum_{j \geq 0}^{\infty} \pi_j B^j Z_t = \sum_{j \geq 0}^{\infty} \pi_j Z_{t-j}. \quad (52)$$

Expresso por $\{\pi_j\}_{j \geq 0}$ os coeficientes de representação do modelo auto-regressivo infinito $AR(\infty)$, representam os pesos do modelo $ARIMA(p,d,q)$ e também dos modelos $ARFIMA(p,d,q)$, $SARIMA(p,d,q) \times (P,D,Q)_S$ e $SARMA(p,q) \times (P,Q)_S$ (MORETTIN e TOLOI, 20004).

A identificação do padrão não aleatório de um conjunto de observações passadas é um determinante para análise da variável de interesse, pois por meio do padrão

comportamental da série é possível realizar previsões com base no passado, para então direcionar a tomada de decisões.

Presumindo que $\{Z_t\}_{t \in \mathbb{Z}}$ seja um processo linear, a representação dada para esse processo é denotada pela expressão (53) de previsão do erro quadrático médio (PRIESTLEY, 1981).

$$\hat{Z}_t(h) = \mathbb{E}(Z_{t+h} | Z_\ell, \ell \leq t), \quad (53)$$

presumindo que as observações em $\{Z_\ell, \ell \leq t\}$ são conhecidas, somente até o instante t , e sendo t denominado origem da previsão, a expressão $\hat{X}_t(h)$ especifica a previsão de origem t , com horizonte $h \geq 1$. E, portanto, esse resultado reduz o erro quadrático médio da previsão representado por $\mathbb{E}(Z_{t+h} - \hat{Z}_t(h))$.

$$e_t(h) = Z_{t+h} - \hat{Z}_t(h). \quad (54)$$

E para o cálculo da previsão são consideradas as restrições a seguir:

- I. $\mathbb{E}(Z_{t+h} | Z_\ell, \ell \leq t) = \begin{cases} Z_{t+h}, & \text{se } h \leq 0, \\ \hat{Z}_t(h), & \text{se } h > 0, \end{cases}$
- II. $\mathbb{E}(\varepsilon_{t+h} | Z_\ell, \ell \leq t) = \begin{cases} \varepsilon_{t+h}, & \text{se } h \leq 0, \\ 0, & \text{se } h > 0. \end{cases}$

Sendo assim, o cálculo das previsões pode ser realizado pela,

- I. Substituição das esperanças passadas (sendo $h \leq 0$) por valores percebidos Z_{t+h} e ε_{t+h} , isto é, valores que são conhecidos, e
- II. Substituição das esperanças futuras (sendo $h > 0$) pelas previsões $\hat{Z}_t(h)$ e 0.

Considerando $\{Z_t\}_{t \in \mathbb{Z}}$ um processo SARIMA(p, d, q) \times (P, D, Q) $_s$ na sua forma inversível, determinada pela expressão (54) e tendo média zero e sazonalidade $s \in \mathbb{N}$. O interesse primordial é obter a previsão do erro quadrático médio mínimo, estabelecido pela estimação $\hat{X}_t(h)$, para qualquer $h \geq 1$.

Portanto, o processo $\{Z_t\}_{t \in \mathbb{Z}}$ dispõe da ilustração do processo de média móvel e auto-regressivo infinitos, nesta mesma ordem, por intermédio das ligações de,

$$\varepsilon_t = \sum_{k \in \mathbb{Z}_{\geq}} \pi_k Z_{t-k}, \quad (55)$$

Rearranjando a equação (55) no tempo $t + h$ é possível encontrar

$$\varepsilon_{t+h} = \sum_{k \in \mathbb{Z}_{\geq}} \pi_k Z_{t+h-k}, \quad (56)$$

ou então, em de forma equivalente,

$$\varepsilon_{t+h} = Z_{t+h} + \sum_{k \in \mathbb{N}} \pi_k Z_{t+h-k}, \quad (57)$$

supondo, $t = n$ como origem, então seguindo as definições de previsões de erro quadrático médio mínimo demonstradas em (53) e (57) chegamos na equação (58).

$$\begin{aligned} \hat{Z}_n(h) &= \mathbb{E}(Z_{n+h} | Z_\ell, \ell \leq n) = \mathbb{E}\left(\varepsilon_{n+h} - \sum_{k \in \mathbb{N}} \pi_k Z_{n+h-k} | Z_\ell, \ell \leq n\right) \\ &= \mathbb{E}(\varepsilon_{n+h} | Z_\ell, \ell \leq n) - \sum_{k \in \mathbb{N}} \pi_k \mathbb{E}(Z_{n+h-k} | Z_\ell, \ell \leq n) \\ &= - \sum_{k \in \mathbb{N}} \pi_k \hat{Z}_n(h-k), \quad \text{para } h \geq 1 \end{aligned} \quad (58)$$

em que, $\mathbb{E}(\varepsilon_{n+h} | X_\ell, \ell \leq n) = 0$ e $\hat{X}_n(j) = X_{n+j}$, para qualquer $j \in \mathbb{Z}_{\geq}$. O que demonstra a validade da equação de previsão de erro quadrático médio mínimo, dado por:

$$\hat{Z}_n(h) = - \sum_{k \in \mathbb{N}} \pi_k \hat{Z}_n(h-k). \quad (59)$$

A metodologia de *Holt-Winters* é a generalização do modelo de suavização exponencial de *Holt*, método aplicado em séries com presença de tendência. Porém no presente trabalho, só serão tratados os dois procedimentos de *Holt-Winters*, aditivo e multiplicativo, os quais são definidos (e/ou diferenciados) somente pelo fator sazonal presente na série.

O método de suavização para modelos sazonais é um procedimento aplicado em séries que possuem um nível comportamental de maior complexidade, caso da

suavização exponencial sazonal de *Holt-Winters*, onde as técnicas utilizadas nas séries sazonais aditivas e multiplicativas, são baseadas em três equações cada, que dispõem de constantes de suavização distintas, relacionadas às componentes da série, como o nível, tendência e sazonalidade (MORETTIN e TOLOI, 2004).

A seguir veremos o método de Suavização Exponencial Aditivo e Multiplicativo de *Holt-Winters*.

3.5.5. Modelo 5: Modelo de Suavização Exponencial – Aditivo.

Definição 5: Em modelos com implicações sazonais aditivas, supõe-se que os dados são gerados por meio da soma do nível, da tendência, de um fator sazonal e também de um erro aleatório (WERNER *et al.*, 2017), os quais estão presentes nos modelos sazonais aditivo de *Holt-Winters*, descrito pela equação (60) a seguir,

$$Z_t = L_t + T_t + F_t + \varepsilon_t, \quad (60)$$

em que, $t = 1, \dots, n$ representa a quantidade de observações da série. E as três equações seguintes denotam as estimativas do conjunto de dados no tempo t , sendo o nível indicado por L_t , a tendência indicada por T_t , e o fator sazonal indicado por S_t .

$$L_t = \alpha(Z_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (61)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}, \quad (62)$$

$$S_t = \gamma(Z_t - L_t) + (1 - \gamma)S_{t-s}. \quad (63)$$

em que: $(0 \leq \alpha \leq 1)$ é a restrição do coeficiente de ponderação exponencial de nível α , $(0 \leq \beta \leq 1)$ é a restrição do coeficiente de ponderação exponencial de tendência β e $(0 \leq \gamma \leq 1)$ é a restrição do coeficiente de ponderação exponencial do fator sazonal γ .

Para realizar previsões h passos à frente, é dada pela expressão (64):

$$\hat{Z}_t(h) = L_t + hT_t + F_{t-s+h} \quad (64)$$

em que, $\hat{Z}_t(h)$ representa a previsão no tempo t e horizonte $h \geq 1$.

Para modelos multiplicativos, tem-se:

3.5.6. Modelo 6: Modelo de Suavização Exponencial Multiplicativo.

Definição 6: Para modelos com efeitos sazonais multiplicativos, é suposto que a série vem da tendência aditiva e do fator sazonal da multiplicativo representado por F_t e de um erro aleatório (WERNER *et al.*, 2017), como vemos na equação.

$$Z_t = L_t F_t + T_t + \varepsilon_t \quad (65)$$

em que, $t = 1, \dots, n$ representa a quantidade de observações da série. E as três equações seguintes denotam as estimativas do conjunto de dados no tempo t , sendo o nível indicado por L_t , a tendência indicada por T_t , e o fator sazonal indicado por S_t , porém neste caso os modelos são multiplicativos.

$$L_t = \alpha \left(\frac{Z_t}{S_{t-s}} \right) + (1 - \alpha)(L_{t-1} + T_{t-1}), \quad (66)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)S_{t-s}, \quad (67)$$

$$S_t = \gamma \left(\frac{Z_t}{L_t} \right) + (1 - \gamma)S_{t-s}, \quad (68)$$

em que: $(0 \leq \alpha \leq 1)$ é a restrição do coeficiente de ponderação exponencial de nível α , $(0 \leq \beta \leq 1)$ é a restrição do coeficiente de ponderação exponencial de tendência β e $(0 \leq \gamma \leq 1)$ é a restrição do coeficiente de ponderação exponencial do fator sazonal γ .

Para realizar previsões h passos à frente, é dada pela expressão (69):

$$\hat{Z}_t(h) = (L_t + hT_t) \times S_{t-s+h} \quad (69)$$

em que, $\hat{Z}_t(h)$ representa a previsão no tempo t e horizonte $h \geq 1$.

Na próxima seção, foram introduzidos os métodos de combinações de previsão para os modelos.

3.6. MÉTODOS DE COMBINAÇÕES DE PREVISÕES

Nesta seção, são apresentados os nove métodos de combinações de previsões calculados pelo *software R Core Team* (2018) pelo comando *Forecast_comb_all()*, o

qual, utiliza o pacote *ForecasCombinations* que possibilita a combinação de diferentes previsões com seus respectivos pesos.

Pela proposta metodológica de Bates e Granger (1969) de combinar modelos para realizar previsão, o método para esse procedimento deveria ser formulado pela combinação linear de previsões não viciadas. Com o desenvolvimento de novos modelos de previsão o número de combinação propostos por Bates e Granger em 1969 para dois modelos, também aumentou e a notação para as previsões dos modelos⁶ individuais, então é descrita como:

$$\hat{\mathbf{Z}}^1 = (\hat{Z}_1^1, \dots, \hat{Z}_n^1), \text{ previsão da série temporal calculada utilizando o } \mathbf{Modelo 1:}$$

Modelo ARIMA(p, d, q);

$$\hat{\mathbf{Z}}^2 = (\hat{Z}_1^2, \dots, \hat{Z}_n^2), \text{ previsão da série temporal calculada utilizando o } \mathbf{Modelo 2:}$$

Modelo ARFIMA(p, d, q);

$$\hat{\mathbf{Z}}^3 = (\hat{Z}_1^3, \dots, \hat{Z}_n^3), \text{ previsão da série temporal calculada utilizando o } \mathbf{Modelo 3:}$$

Modelo SARMA(p, q) \times (P, Q)_s;

$$\hat{\mathbf{Z}}^4 = (\hat{Z}_1^4, \dots, \hat{Z}_n^4), \text{ previsão da série temporal calculada utilizando o } \mathbf{Modelo 4:}$$

Modelo SARIMA(p, d, q) \times (P, D, Q)_s;

$$\hat{\mathbf{Z}}^5 = (\hat{Z}_1^5, \dots, \hat{Z}_n^5), \text{ previsão da série temporal calculada utilizando o } \mathbf{Modelo 5:}$$

Modelo de Suavização Exponencial Aditiva;

$$\hat{\mathbf{Z}}^6 = (\hat{Z}_1^6, \dots, \hat{Z}_n^6), \text{ previsão da série temporal calculada utilizando o } \mathbf{Modelo 6:}$$

Modelo de Suavização Exponencial Multiplicativo.

O primeiro método para combinação de previsões é o mais simples, porém o mais usual entre os métodos é o de média aritmética simples.

3.6.1. Combinação de Previsão pelo Método da Média Aritmética Simples (MS).

Este método é o que possui a abordagem mais elementar entre todos os métodos, porém é a mais utilizada dada sua simplicidade de combinar previsões de modelos para então gerar a média dessas previsões (Weiss, 2018).

A matriz (Tabela 2) é a representação dos Modelos 1, ..., 6 gerados pelas observações dos gastos médios mensais com o consumo da Cesta Básica de Alimentos

⁶ Ver seção 3, subseção 3.4.

de Porto Alegre. Esta representação (Tabela 2) é utilizada nos métodos de combinações de previsões pela média aritmética simples, mediana e variância mínima.

Tabela 2 – Matriz de previsões para combinação de previsões pelos métodos da Média Aritmética Simples, da Mediana e da Variância Mínima.

$$\begin{bmatrix} \hat{Z}_1^1 & \hat{Z}_1^2 & \hat{Z}_1^3 & \hat{Z}_1^4 & \hat{Z}_1^5 & \hat{Z}_1^6 \\ \hat{Z}_2^1 & \hat{Z}_2^2 & \hat{Z}_2^3 & \hat{Z}_2^4 & \hat{Z}_2^5 & \hat{Z}_2^6 \\ \hat{Z}_3^1 & \hat{Z}_3^2 & \hat{Z}_3^3 & \hat{Z}_3^4 & \hat{Z}_3^5 & \hat{Z}_3^6 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{Z}_n^1 & \hat{Z}_n^2 & \hat{Z}_n^3 & \hat{Z}_n^4 & \hat{Z}_n^5 & \hat{Z}_n^6 \end{bmatrix}$$

Fonte: Os Autores.

em que, $\hat{Z}_1^1, \dots, \hat{Z}_n^1, \dots, \hat{Z}_1^6, \dots, \hat{Z}_n^6$ são as observações dos seis modelos ajustados e representados por $\hat{Z}^1, \dots, \hat{Z}^6$.

O procedimento para combinação dos modelos (Tabela 2) é realizada de modo que as combinações devem ser feitas linha por linha, e ao realizar a combinação do Modelo 1 (denotado na matriz da Tabela 2, por \hat{Z}^1) com os demais Modelos, pelo Método da Média Aritmética Simples, o procedimento pode ser feito da seguinte forma: combinando-os dois a dois, três a três, ..., até que se chegue na combinação de todos os modelos. Abaixo são demonstrados os passos para realizar as combinações dos modelos.

- 1) Para combinação dois a dois do Modelo 1 com os Modelo 2, 3, 4, 5 e 6;
- 2) Na combinação de previsões por média aritmética simples é calculada a média aritmética das previsões por linha da matriz (ver Tabela 2) levando em consideração o número de previsões a serem combinadas, por exemplo, ver Apêndice A.

Sendo este método a soma das combinações divididas por suas respectivas quantidades k , a expressão geral do método da média aritmética simples é dada por:

$$\hat{Z}_t^{(1)} = \frac{\hat{Z}_t^1 + \dots + \hat{Z}_t^k}{k}. \quad (70)$$

em que, $t = 1, \dots, n$ representa o número de observações e $k = 2, \dots, 6$ são as combinações dos modelos calculadas por $\hat{\mathbf{Z}}^{(1)} = (\hat{Z}_t^{(1)}, \dots, \hat{Z}_t^{(1)})$.

O método da mediana, assim como o da média é bastante conhecido e utilizado.

3.6.2. Combinação de Previsão pelo Método da Mediana (MED).

O método da mediana é outra forma simples de combinar previsões de modelos, e por não apresentar sensibilidade a *outliers* é um método bastante vantajoso, sendo muito útil em combinações que o método da média aritmética simples não é adequado, ou seja, quando a previsão de algum modelo é tendencioso, o que exige a aplicação de uma medida de localização mais robusta para os valores extremos (Weiss, 2018).

A expressão geral dada para o método da mediana é,

$$\hat{Z}_t^{(2)} = md(\hat{Z}_t^1, \dots, \hat{Z}_t^k). \quad (71)$$

em que, $t = 1, \dots, n$ representa o número de observações e $k = 2, \dots, 6$ são as combinações dos modelos calculadas por $\hat{\mathbf{Z}}^{(2)} = (\hat{Z}_t^{(2)}, \dots, \hat{Z}_t^{(2)})$.

O terceiro método descrito trata da combinação pela variância mínima.

3.6.3. Combinação de Previsão pelo Método da Variância Mínima (VM).

Com a proposta de minimizar a variância da previsão combinada $\hat{\mathbf{Z}}^{(3)}$. Esse procedimento de redução da variância é então denominado de método da Variância Mínima. Portanto, a variância da previsão combinada é dada por:

$$\hat{\mathbf{Z}}^{(3)} = w_1 \hat{\mathbf{Z}}^1 + \dots + w_k \hat{\mathbf{Z}}^k \quad (72)$$

em que, $k = 2, \dots, 6$, são as combinações dos modelos e $\sum_{t=1}^k w_t = 1$ representam os pesos que variam de $0 \leq w \leq 1$.

Para o caso de $k = 2$, temos que a variância da previsão combinada é dada por:

$$\sigma_c^2 = w^2 \sigma_1^2 + (1 - w)^2 \sigma_2^2 + 2\rho \cdot w \sigma_1 (1 - w) \sigma_2. \quad (73)$$

Derivando em relação à w , temos que *menos peso* é destinado às previsões de maior variabilidade dos erros e é dado por:

$$w = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}. \quad (74)$$

em que, ρ representa o valor da correlação linear entre os erros de previsão de \hat{Z}^1 e \hat{Z}^2 , nesta ordem, σ_1^2 e σ_2^2 representam as variâncias dos erros (absolutos) da previsão \hat{Z}^1 e \hat{Z}^2 , nesta ordem.

Em casos que os erros não são correlacionados ($\rho = 0$), o peso w é reduzido ao resultado da expressão abaixo.

$$w = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}. \quad (75)$$

As técnicas de análise de regressão linear são popularmente utilizadas para inferir a respeito de uma variável de interesse, pois por intermédio de equações geradas por formulações matemáticas o método de regressão necessita de estimadores para melhor ajustar seus parâmetros.

A Tabela 3, demonstra o procedimento da combinação dos Modelos e como são calculadas as combinações para realizar as previsões por regressão.

Tabela 3 – Matriz dos Modelos Individuais para Combinação de Previsões pelos Métodos de Regressão.

$$\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \hat{Y}_3 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} 1 & \hat{Z}_1^1 & \hat{Z}_1^2 & \hat{Z}_1^3 & \hat{Z}_1^4 & \hat{Z}_1^5 & \hat{Z}_1^6 \\ 1 & \hat{Z}_2^1 & \hat{Z}_2^2 & \hat{Z}_2^3 & \hat{Z}_2^4 & \hat{Z}_2^5 & \hat{Z}_2^6 \\ 1 & \hat{Z}_3^1 & \hat{Z}_3^2 & \hat{Z}_3^3 & \hat{Z}_3^4 & \hat{Z}_3^5 & \hat{Z}_3^6 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \hat{Z}_n^1 & \hat{Z}_n^2 & \hat{Z}_n^3 & \hat{Z}_n^4 & \hat{Z}_n^5 & \hat{Z}_n^6 \end{bmatrix} \times \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Fonte: Os autores.

A regressão linear geral é expressa por:

$$\hat{Z}_t^{(i)} = \beta_0 + \beta_1 \hat{Z}_t^1 + \dots + \beta_k \hat{Z}_t^k + \varepsilon_t \quad (76)$$

em que, β_0 é a constante, $(\beta_1, \dots, \beta_k)$ representam os coeficientes de regressão, $t = 1, \dots, n$ representa o número de observações e $k = 2, \dots, 6$ são as combinações dos modelos calculadas por $\hat{Z}^{(i)} = (\hat{Z}_t^{(i)}, \dots, \hat{Z}_t^{(i)})$.

Um estimador de parâmetros bastante conhecido e útil na formulação de regressão é o de mínimos quadrados ordinários (MQO), porém em casos que a variável resposta não dispõe de uma distribuição normal ou contém *outliers* este estimador é parcialmente confiável.

No entanto, quando o estimador MQO não apresente confiabilidade necessária o estimado de regressão robusta é o mais apropriado, pois este estimador tem como objetivo minimizar o domínio de pontos discrepantes, que exercem influência na qualidade da estimação dos parâmetros da regressão linear, e como particularidade possui sensibilidade à presença de *outliers*, além de produz estimativas iguais aos do método de mínimos quadrados ordinários quando não a variável resposta segue normalidade e não apresenta *outliers*.

A seguir são descritos os métodos de regressão para combinação de modelos.

3.6.4. Combinação de Previsão pelo Método de Regressão de Mínimos Quadrados (MQO).

No modelo de regressão representada em sua forma geral, e composta de k variáveis $\{\hat{Z}_t^k\}_{t=1}^n$ independentes, com $k = 2, \dots, 6$ e $\{Z_t\}_{t=1}^n$ representando a variável dependente. A estimação de $\beta = (\beta_0, \beta_1, \dots, \beta_k)'$ realizada pelo procedimento de mínimos quadrados (MQ), resulta em um valor estimado $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)' \in \mathbb{R}^{k+1}$, que possibilita a redução da função perda (BISOGNIN, 2007), então

$$\mathcal{L}_1(n) = \sum_{t=1}^n r_t^2. \quad (77)$$

em que, $t = 1, \dots, n$ representa o número de observações e $k = 2, \dots, 6$ são as combinações dos modelos.

E os resíduos r_t escritos como:

$$r_t = Z_t - [\hat{\beta}_0 + \hat{\beta}_1 \hat{Z}_t^1 + \dots + \hat{\beta}_k \hat{Z}_t^k] \quad (78)$$

em que, $\{Z_t\}_{t=1}^n$ é a variável em estudo, $\hat{\beta}_0$ é a constante, $(\hat{\beta}_1, \dots, \hat{\beta}_k)$ representam os coeficientes de regressão, $t = 1, \dots, n$ representa o número de observações e $k = 2, \dots, 6$ são as combinações dos modelos calculadas por $\hat{\mathbf{Z}}^{(4)} = (\hat{Z}_t^{(4)}, \dots, \hat{Z}_t^{(4)})$.

O seguinte método gera combinação de modelos por regressão robusta.

3.6.5. Combinação de Previsão pelo Método de Regressão Robusta de Mínimos Quadrados Podados (MQP).

Em 1984, Rousseeuw sugeriu a estimação robusta por meio do método de mínimos quadrados podados (MQP), onde os estimadores fundamentados na regressão geral seriam alcançados por esse método, resultando em valores estimados para $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k) \in \mathbb{R}^{k+1}$ minimizando a função perda (BISOGNIN, 2007), e sendo denotada por:

$$\mathcal{L}_2(n) = \sum_{j=1}^{m^*} (r^2)_{j:m}. \quad (79)$$

em que, $(r^2)_{1:m} \leq \dots \leq (r^2)_{m^*:m}$ representam os resíduos ao quadrado dado pela expressão (78) e m^* é o número de pontos utilizados no processo de otimização. E, $t = 1, \dots, n$ representa o número de observações e $k = 2, \dots, 6$ são as combinações dos modelos calculadas por $\hat{\mathbf{Z}}^{(5)} = (\hat{Z}_t^{(5)}, \dots, \hat{Z}_t^{(5)})$ para o método de regressão MQO.

Sendo visível a similaridade entre as expressões (79) e a (77), a distinção entre ambas é o procedimento ao quadrado dos resíduos. E quando ao ponto de interrupção e eficiência, o mesmo é determinado pela constante m^* . Realizado o procedimento de estimação de parâmetros do modelo ARFIMA(p, d, q) pelo método de mínimos quadrados podados, caso $m^* = [g(n)/2]$, para maiores detalhes ver Bisognin (2007) Capítulo 5, sendo $[x]$ a parte não fracionaria de x (parte inteira), a interrupção será próxima a 50% (LOPES e MENDES, 2006).

O próximo método também gera combinação de previsões por regressão robusta.

3.6.6. Combinação de Previsão pelo Método de Regressão Robusta (MM).

Outra metodologia de estimador robusto fundamentado na formulação geral de regressão proposto por Yohai em 1987, com denominação conhecida com estimador MM, possui algumas características descritas como:

- I. Altamente eficientes quando os erros seguem distribuição normal,
- II. Qualitativamente robustos e
- III. Apresentam elevado ponto de interrupção.

Estes estimadores *MM* são determinados por estágios, sendo o primeiro resultante do calculo originário, sendo o procedimento qualitativo robusto com elevado ponto de interrupção, porém não evidencia eficiência. No segundo estágio, o estimador *M* da escala de erros é resultante do procedimento calculado pelos resíduos baseados na estimação inicial do primeiro estágio, e no último estágio, um estimador *M* dos parâmetros da regressão é encontrado pela fundamentação em uma função apropriada (BISOGNIN, 2007).

Estimadores embasados na forma geral de regressão e na metodologia *MM* são designados como $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k) \in \mathbb{R}^{k+1}$ que visa reduzir a função perda,

$$\mathcal{L}_3(n) = \sum_{t=1}^n \rho_2 \left(\frac{r_t}{\delta} \right)^2. \quad (80)$$

porém, fica exposta a restrição,

$$\frac{1}{n} \sum_{t=1}^n \rho_1 \left(\frac{r_t}{\delta} \right) \leq b. \quad (81)$$

em que, $\rho_1(\cdot)$ e $\rho_2(\cdot)$ representam as funções simétricas finita e crescente em $[0, \infty)$, tendo $\rho_v(0) = 0$, com $\lim_{u \rightarrow \infty} \rho_v(u) = 1$, para $v = 1, 2$, δ é um parâmetro de escala e b é denominado por $\mathbb{E}_{\emptyset}(\rho_1(u)) = b$, sendo \emptyset a função de distribuição acumulada da distribuição normal padrão e r_t representa os resíduos dados pela expressão (78), para qualquer $t = 1, \dots, n$. O ponto de interrupção dos estimadores *MM* depende somente da

função $\rho_1(\cdot)$ sendo dado por $\min\{b, 1 - b\}$. Portanto, os estimadores apoiados no método *MM* são consistentes e assintoticamente normais.

Então, $t = 1, \dots, n$ representa o número de observações e $k = 2, \dots, 6$ são as combinações dos modelos calculadas por $\hat{\mathbf{Z}}^{(6)} = (\hat{Z}_t^{(6)}, \dots, \hat{Z}_t^{(6)})$ para o método de regressão robusta *MM*.

O método seguinte é o por Regressão Quantílica.

3.6.7. Combinação de Previsão pelo Método de Regressão Quantílica (RQ).

Este método (RQ) é normalmente aplicado a observações que apresentem casos de ocorrência de *outliers*, alta assimetria e heteroscedasticidade (KOENKER, 2005).

Por este método de regressão é possível extrair informações da variável em análise em vários níveis da distribuição, deste modo, é evidente a possibilidade de obter informações estatísticas mais completas da série. Proposta por Koenker e Bassett em 1978, a Regressão Quantílica segue influência do método dos Erros Absolutos Ponderados, que considera além do valor mediano uma ponderação na redução dos erros, para então estimar os quantis que possuem relevância.

Conforme Koenker (2005) abordagens por Regressão Quantílica possuem a capacidade de vincular heteroscedasticidade, que por vez, pode ser identificada a partir da alteração (variação) das estimativas dos coeficientes dos parâmetros para diferentes quantis.

Um fato importante a ser mencionado é que em modelos de Regressão Linear Simples, somente é estimado uma reta para justificar o conjunto de observações, e consequentemente, esse método não registra variações na distribuição da variável dependente. Por outro lado, na Regressão Quantílica é possível estimar retas para cada quantil que possuem relevância, o que torna os modelos de Regressão Quantílica mais adequados para inferir sobre os dados que apresentam e assimetria.

A função perda é dada por:

$$\mathcal{L}_4(n) = \sum_{t=1}^n r_t \cdot \quad (82)$$

em que , $t = 1, \dots, n$, representa o número de observações, r_t são os resíduos dados pela equação (78) e $k = 2, \dots, 6$ são as combinações dos modelos calculadas por $\hat{\mathbf{Z}}^{(7)} = (\hat{Z}_t^{(7)}, \dots, \hat{Z}_t^{(7)})$ para o método de regressão Quantílica.

O método a seguir, gera combinação de previsões por meio de ranqueamento.

3.6.8. Combinação de Previsão pelo Método do Inverso do Rank.

Este método ignora as correlações entre os erros de previsões e considera os pesos, que são proporcionais ao inverso dos valores do MSE dos modelos individuais, MSE_i , para $i = 1, \dots, k$ é o número de modelos gerados pela série e $k = 2, \dots, 6$ são as combinações de modelos de previsões, que resultam da expressão dos pesos,

$$w_i = \frac{MSE_i^{-1}}{\sum_{i=1}^k MSE_i^{-1}}. \quad (83)$$

Uma modalidade deste método foi considerada por Aiolfi e Timmermann em 2016, onde a recomendação de uma estrutura avaliação na forma robusta permitisse que os pesos de combinação fossem inversamente proporcionais, na especificação dos modelos para previsão, $Rank_k$, sendo expresso por:

$$w_i = \frac{Rank_i^{-1}}{\sum_{i=1}^k Rank_i^{-1}}. \quad (84)$$

Neste caso, para o modelo com o menor valor de MSE recebe um *rank* de 1, o seguinte modelo com o segundo menor modelo de MSE recebe um *rank* de 2 e assim sucessivamente, e portanto, esta estrutura combinada desconsidera outra vez as correlações erros os erros de previsões (ELLIOTT e TIMMERMANN, 2016).

Para o método do Inverso do Rank, o procedimento a ser calculado por $\hat{\mathbf{Z}}^{(8)} = (\hat{Z}_t^{(8)}, \dots, \hat{Z}_t^{(8)})$ também tem as observações representadas por $t = 1, \dots, n$ sendo $k = 2, \dots, 6$ as combinações dos modelos.

O último método para combinação de previsões deste estudo gera combinações, por intermédio da correção do Auto Vetor por Viés.

3.6.9. Combinação de Previsão pelo Método de Auto Vetor Corrigida pelo Viés (EIG2).

Inicialmente Hsiao e Wan (2014) propõem uma abordagem de combinação de previsão utilizando auto vetores, onde os pesos são calculados utilizando matriz de erro quadrático médio de predição (EQMP) da seguinte forma: Suponha que $\{\mathbf{Z}_t\}_{t \in \mathbb{Z}}$ é a variável de interesse, e não existem k , $k = 2, \dots, 6$, preditores perfeitamente colineares (iguais), $\mathbf{f}_t = (f_{t1}, \dots, f_{tk})'$, Σ é a matriz de erro quadrática média de previsão (definida positiva) de \mathbf{f}_t e $\boldsymbol{\varepsilon}$ é um vetor $k \times 1$.

Os k autovalores positivos são então organizados em ordem crescente $\phi_{(1)}, \dots, \phi_{(k)}$ e w_j é definido como o auto vetor correspondente a $\phi_{(j)}$. O vetor de pesos da combinação $\mathbf{w} = (w_1, \dots, w_k)'$, denotados por \mathbf{w}^l , são então escolhidos correspondendo ao mínimo de $\left(\frac{\phi_{(1)}}{d_1^2}, \dots, \frac{\phi_{(k)}}{d_k^2}\right)$, em que $d_j = \boldsymbol{\varepsilon}'\mathbf{w}^j$, quando $\mathbf{w} = \frac{1}{d_l}\mathbf{w}^l$.

A combinação das previsões é então obtida por $\widehat{\mathbf{Z}}_t = (\mathbf{f}_t)'\mathbf{w}$.

A diferença para os métodos existentes que minimizam o erro quadrático médio de previsão populacional (por exemplo, Newbold e Granger, 1974) é a função de normalização. Enquanto as abordagens anteriores otimizam o EQMP sob a restrição de $\boldsymbol{\varepsilon}'\mathbf{w} = 1$, Hsiao e Wan (2014) mostram que isso é dominado usando $\mathbf{w}'\mathbf{w} = 1$ como restrição no problema de otimização.

Neste mesmo trabalho, Hsiao e Wan (2014), propõem uma abordagem de auto vetores corrigido pelo viés, a qual baseia-se na ideia de que, se um ou mais dos modelos preditivos produzem análises tendenciosas, a precisão da abordagem padrão do autovetor pode ser melhorada eliminando o viés. O procedimento de otimização para obter pesos combinados coincide com o método padrão do vetor, exceto que é aplicado à matriz EQMP centralizada após a extração de viés (subtraindo a média da coluna do EQMP). Os pesos da combinação de previsões são calculados como:

$$\mathbf{w} = \frac{1}{\tilde{d}_l}\tilde{\mathbf{w}}^l, \quad (85)$$

em que, \tilde{d}_l e $\tilde{\mathbf{w}}^l$ são definidos analogamente a d_l e \mathbf{w}^l na abordagem padrão de autovetor, com a única diferença de que correspondem à decomposição espectral da matriz EQMP centralizada em vez da não centrada em 1.

A previsão combinada é então permitida por:

$$\hat{Z}_t = a + \mathbf{f}'_t \mathbf{W}^{EIG2}. \quad (86)$$

em que, $\mathbf{w}^{EIG2} = \frac{1}{\hat{d}_t} \tilde{\mathbf{w}}^I$ e $a = E\mathbf{y}_t - E(\mathbf{f}'_t) \mathbf{w}^{EIG2}$.

Assim como os métodos anteriores, o método de Auto Vetor Corrigida pelo Viés tem sua combinação feita por $\hat{\mathbf{Z}}^{(9)} = (\hat{Z}_t^{(9)}, \dots, \hat{Z}_t^{(9)})$, com $t = 1, \dots, n$ observações e $k = 2, \dots, 6$ combinações dos modelos.

Na próxima seção, são descritos as medidas de Acurácias utilizadas nos modelos para previsões individuais e para as combinações de previsões.

3.7. MEDIDAS DE ACURÁCIA

O termo acurácia significa exatidão, mais precisão e conceituada na literatura como qualidade do ajuste de distintos modelos encontrados em uma determinada série. Mikhail e Ackermann (1976) determinaram o termo acurácia como sendo a intensidade (ou grau) de proximidade de uma estimativa ao seu parâmetro, ao passo que, precisão indica a intensidade (ou grau) de proximidade das observações à sua média.

Para Saigal e Mehrotra (2012) as medidas de acurácia para previsão, determinam a qualidade do ajuste dos modelos de séries temporais, tendo por finalidade, demonstrar os resultados dos erros de previsão e então comparar seus desempenhos entre os modelos encontrados, e assim, identificar aquele com os menores erros de precisão gerados na previsão de um conjunto de dados selecionados.

A primeira mediada de Acurácia a ser ilustrada, é o erro absoluto médio.

3.7.1. Erro Absoluto Médio – MAE

O termo de erro da medida desta medida depende da escala de precisão da previsão, sendo seus resultados descritos da mesma forma que as unidades de medidas das observações reais da série. Para encontrar o valor do MAE, é necessário utilizar o recurso do valor absoluto, para que os resultados médios finais não sejam negativos (SAIGAL e MEHROTRA, 2012), desta forma temos:

$$MAE = \frac{1}{n} \sum_{t=1}^n |Z_t - \hat{Z}_t|. \quad (87)$$

em que, $t = 1, 2, \dots, n$ é o número de observações no tempo t , $\{Z_t\}_{t=1}^n$ representa a série real ou observados e $\{\hat{Z}_t\}_{t=1}^n$ é a série estimada.

A segunda mediada descrita é a da raiz do erro quadrático médio.

3.7.2. Raiz do Erro Quadrático Médio – RMSE

Esta medida mensura a relevância da média do erro, porém como no erro quadrático médio (RMSE) os erros são elevados ao quadrado antes mesmo de ser calculado o resultado final, erros com valores altos tendem a serem relativamente ainda

mais elevados e conseqüentemente o RMSE é uma medida adequada quando erros elevados não são desejados (SAIGAL e MEHROTRA, 2012).

A formulação do erro quadrático médio é representada por:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Z_t - \hat{Z}_t)^2}. \quad (88)$$

em que, $t = 1, 2, \dots, n$ é o número de observações no tempo t , $\{Z_t\}_{t=1}^n$ representa a série real ou observados e $\{\hat{Z}_t\}_{t=1}^n$ é a série estimada.

Assim como o MAE, o RMSE também é medido nas mesmas unidades que os dados originais.

A terceira mediada descrita é a do erro percentual médio absoluto.

3.7.3. Erro Percentual Médio Absoluto – MAPE

De acordo com Hyndman e Koehler (2006), Saigal e Mehrotra (2012) entre outros autores, recomendam o MAPE por sua objetividade, isto é, o erro percentual médio absoluto (MAPE) é a estatística que melhor expressa à realidade dos dados, pois além de apresentar resultados que estão em percentagem relativa essa medida não é afetado pela unidade de medida da série de previsão. Neste sentido, quanto mais próximo de zero for seu resultado, melhores serão os valores resultantes da previsão.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{(Z_t - \hat{Z}_t)}{Z_t} \right| * 100. \quad (89)$$

em que, $t = 1, 2, \dots, n$ é o número de observações no tempo t , $\{Z_t\}_{t=1}^n$ representa a série real ou observados e $\{\hat{Z}_t\}_{t=1}^n$ é a série estimada.

Como o erro absoluto médio (MAE) também é medido nas mesmas unidades que os dados reais da série e semelhante em magnitude, mas visivelmente menor que o erro quadrático médio da raiz (RMSE), (SAIGAL e MEHROTRA, 2012).

O erro percentual absoluto médio (MAPE) é muito utilizado em elaborações de relatórios, pois sendo uma medida expressa em percentual geral, dispõe de interpretação simples até mesmo para aqueles que não tenham conhecimento de como se constitui um

grande erro em termos monetários. No entanto, esta estatística é calculada exclusivamente por dados positivos.

O erro quadrático médio da raiz (RMSE) é mais sensível do que outras medidas de grande erro eventual. Pois, o processo ao quadrado dá um peso desproporcional a erros muito grandes. Se um grande erro ocasional não for um problema na situação de decisão, o MAE ou o MAPE podem ser o critério mais relevante. Em vários casos, essas estatísticas variam em semelhança e o modelo que é melhor em um deles também será em outros, mas isso pode não ser o caso quando a distribuição de erro tem "outliers" (SAIGAL e MEHROTRA, 2012).

A quarta e última medida de Acurácia deste estudo é a do coeficiente de Theil.

3.7.4. Coeficiente U de Theil

O uso desta estatística tem como finalidade avaliar a qualidade ajustada das previsões. Para Bliemel (1973), a estatística proposta por Henry Theil causa certo embaraço por ter formulado dois coeficientes em épocas distintas, porém com nomes iguais e conhecidas pela simbologia *U*. Como primeira proposição de sua formulação, *Theil* determinou que o resultado estatístico ficasse entre 0 e 1, no entanto, em sua segunda proposição, ele propôs que o limite superior seria um resultado infinito.

Para este estudo foi utilizado à primeira proposição e a expressão é dada por:

$$U \text{ de Theil} = \frac{\sqrt{\frac{1}{n} \sum_{t=1}^n (z_t - \hat{z}_t)^2}}{\left[\sqrt{\frac{1}{n} \sum_{t=1}^n z_t} + \sqrt{\frac{1}{n} \sum_{t=1}^n \hat{z}_t} \right]} \quad (90)$$

em que, $t = 1, 2, \dots, n$ é o número de observações no tempo t , $\{Z_t\}_{t=1}^n$ representa a série real ou observados e $\{\hat{Z}_t\}_{t=1}^n$ é a série estimada.

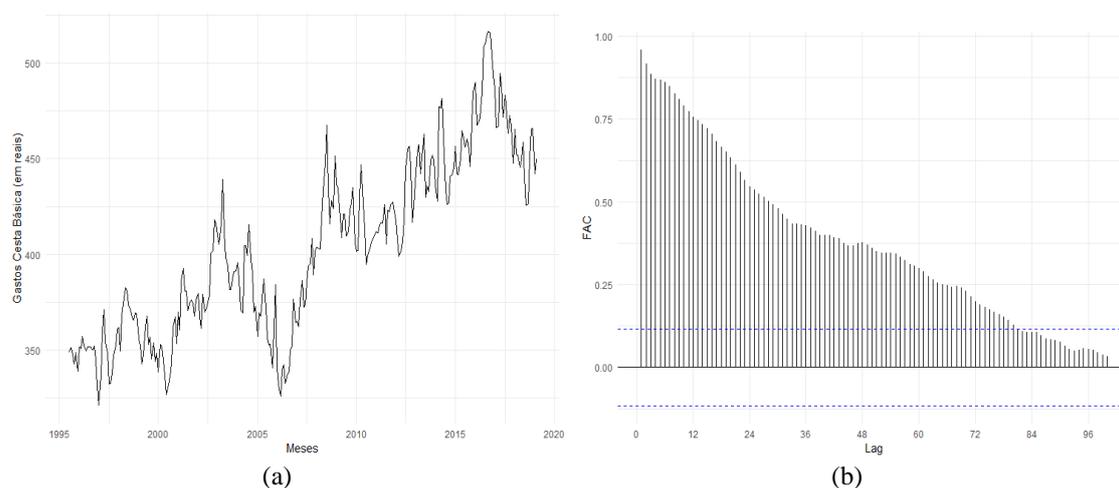
A condição necessária para que esta medida gere um resultado adequado para os erros de previsão, é que a estatística *U de Theil* resulte em um valor muito próximo de zero, demonstrando que o modelo escolhido ou a combinação de modelos obteve erro de previsão menor que de uma previsão ingênua (MAKRIDAKIS, WHEELWRIGHT e HYNDMAN, 1998).

Na próxima seção, são apresentados os resultados estatísticos desta pesquisa.

4. RESULTADOS E DISCUSSÃO

A Figura 1 apresenta o gráfico da série histórica dos gastos médios no consumo da Cesta Básica de Porto Alegre, RS, no período de julho de 1995 a fevereiro de 2019, juntamente com a função de autocorrelação. Por intermédio da série original, é notável o crescimento com variações de picos ascendentes e descendentes em relação aos preços (R\$) praticados no decorrer dos períodos, sendo que no mês de março do ano de 2006 houve a ocorrência de uma expressiva queda de preços dos bens alimentícios, após uma sucessão de altas e baixas. Também se pode verificar a presença de tendência determinística nos preços, e se teve a necessidade de corrigir tal efeito para o uso da mesma, diante desta ocorrência, a série aparentou ter indícios de sazonalidade. Cabe ressaltar que durante o trabalho de análise dos dados, os mesmos foram sendo atualizados conforme o DIEESE e o IBGE atualizavam seus bancos de dados, ou seja, incluímos os valores da Cesta Básica de Porto Alegre para os meses de junho a julho de 2019.

Figura 1 – Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de jul/1995 à fev/2019: (a) gráfico da série temporal; (b) função de autocorrelação amostral.



Fonte: Elaborado pelos autores com base nos dados disponíveis pelo DIEESE e IBGE.

Para verificar a existência de raiz unitária na série temporal, foram aplicados os testes de Dicke-Fuller Aumentado (ADF), de Phillips-Perron e o KPSS.

Tendo os dois primeiros testes, a hipótese alternativa (H_1) de estacionariedade na série, o teste ADF obteve um valor-p [0,2179] maior que o nível de significância de 5%,

demonstrando presença de raiz unitária na série, o que levou a realizar o procedimento de primeira diferença, a qual foi o suficiente para tornar a série estacionária e com valor-p [0,01]. O teste de Phillips-Perron, resultou em um valor-p [0,01] menor que 5%, demonstrando ausência de raiz unitária e estacionariedade na série, também após primeira diferença. No entanto, quando aplicado à hipótese alternativa de tendência explosiva, ambos resultaram em valor-p [0,7821 e 0,99, seguindo a ordem dos testes] maior que o nível de significância testado, evidenciando tendência explosiva na série temporal.

Ao contrário dos testes anteriores, o teste KPSS possui hipótese alternativa de não estacionariedade na série, e o mesmo apresentou um valor-p [0,01] menor que o nível de significância de 5%, apresentando raiz unitária na série por ocorrência da ausência de estacionariedade, e por consequência foi necessário aplicar a primeira diferença nos dados, sendo o suficiente para torná-los estacionários e com valor-p [0,10].

Aplicou-se o teste de Wald-Wolfowitz para verificação de tendência determinista na série temporal, e testando a hipótese alternativa (H_1 : de existência de tendência determinística na série), constatou-se um valor-p [2,2e-16] menor que o nível de significância de 5%, evidenciando presença de tendência determinística na série em estudo.

Por tratar-se de uma série temporal não estacionária e com sazonalidade, e como o objetivo deste trabalho foi o de prever os gastos médios mensais futuros no consumo da Cesta Básica de Alimentos de Porto Alegre, tratamos a tendência desta série temporal como estocástica, utilizando os modelos $SARIMA(p,d,q) \times (P,D,Q)_S$, e como determinística, ou seja, a tendência foi tratada por meio de um modelo linear ajustado a um modelo $SARMA(p,d,q) \times (P,D,Q)_S$, com $d = 0 = D$. Foi ajustado ainda o modelo de Holt Winters, onde a componente sazonal foi tratada como multiplicativa.

A seguir serão geradas as previsões utilizando os modelos e suas respectivas medidas de acurácia. Observamos então que a nossa série é composta por: sazonalidade e tendência.

Modelo 1: Ajuste do modelo $ARIMA(p,d,q)$.

Na Tabela 4, estão presentes os parâmetros $\hat{\phi}_1, \hat{\phi}_2, \hat{\theta}_1$ e $\hat{\theta}_2$ estimados e suas estimativas, na qual, verificou-se que todas são significativas ao nível de significância

de 5%, juntamente com o estimador, com os critérios de informações e também com os testes de resíduo e de normalidade, o ajustamento resultou em um modelo ARIMA(2,1,2).

Tabela 4 – Resultado do ajuste do modelo ARIMA(2,1,2).

Parâmetros	Estimativa	Erro Padrão	Estat. Z	P(> z)
ϕ_1	1,1257	0,1505	7,475	7,515e-14
ϕ_2	-0,6758	0,1094	-6,1787	6,463e-10
θ_1	-1,1991	0,1732	-6,9214	4,473e-12
θ_2	0,5523	0,1483	3,7251	0,0001953

Log Verossimilhança = -1102,89
AIC = 2215,78; AICc = 2216; BIC = 2234,01
Teste de Resíduos (Box-Pierce): df = 15 – p-valor = 0,8791
Teste de Normalidade (Shapiro-Wilk): W=0,9935 – p-valor = 0,2527

Fonte: Os Autores.

Foi realizado o teste de Box-Pierce e de Shapiro-Wilk, para testar os resíduos e a normalidade (Tabela 4), tendo como hipóteses alternativas de que os modelos são correlacionados e que os dados não seguem distribuição normal. Para ambos os testes foi encontrado um p-valor maior que 5%, indicando que não há evidências para rejeitar as hipóteses nulas, de que os resíduos não são correlacionados e os dados seguem distribuição normal, isso demonstra que o modelo ARIMA(2,1,2) proposto é adequado.

Modelo 2: Considerando tendência determinística, foi ajustado o modelo linear $X_t = \beta_0 + \beta_1 t + Y_t$, onde $t = 1, \dots, n$, que obteve um $R^2 = 0,7508$. As estimativas dos coeficientes e suas respectivas estatísticas são apresentadas na Tabela 5.

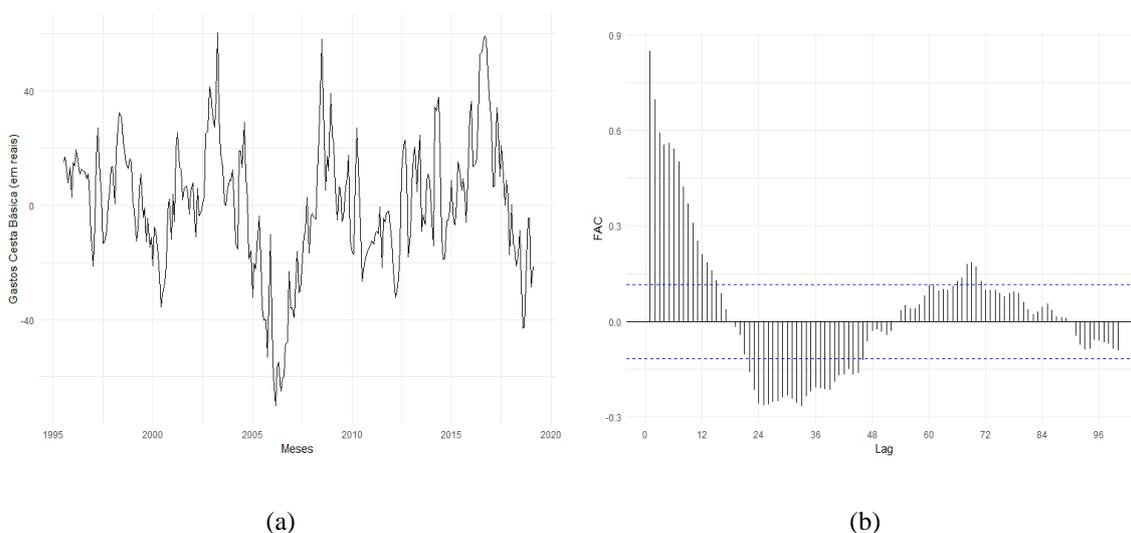
Tabela 5 – Resultado do ajuste da tendência determinística.

Parâmetros	Estimativa	Erro Padrão	Estat. Z	P(> z)
β_0	333,36950	2,74233	121,56	<2e-16
β_1	0,48630	0,01668	29,15	<2e-16

Fonte: Os Autores.

Pela Tabela 5, verificou-se que os parâmetros $\hat{\beta}_0$ e $\hat{\beta}_1$ estimados apresentaram estimativas significativas, ao nível de significância de 5%. Na Figura 2, expõem-se o comportamento da série sem tendência determinística.

Figura 2 – Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de jul/1995 à fev/2019 após retirada da tendência determinística: (a) gráfico da série temporal; (b) função de autocorrelação amostral.



Fonte: Os Autores.

A Figura 2 apresenta a série temporal sem tendência e sua respectiva função de autocorrelação amostral. Foram aplicados os testes de raiz unitária de Dickey-Fuller [valor- $p < 0,01$] e de Phillips-Perron [valor- $p < 0,01$], ou seja, a série temporal ficou estacionária, após retirada a tendência determinística.

A seguir foi ajustado o modelo de Memória Longa ARFIMA(p, d, q), cujos coeficientes encontram-se na Tabela 6.

Tabela 6 – Resultado do ajuste do modelo ARFIMA($1, d, 0$), após retirada da tendência determinística.

Parâmetros	Estimativa	Erro Padrão	Estat. z	P(> z)
d	0,3709	0,0138	26,866	< 2,e-16
ϕ_1	0,5266	0,0525	10,03	< 2,e-16

Log Verossimilhança = -1111, AIC = 2227,838;

Teste de Resíduos (Box-Pierce): df = 15 – p-valor = 0,1133

Teste de Normalidade (Shapiro-Wilk): W=0,9915 – p-valor = 0,0993

Fonte: Os Autores.

A Tabela 6, demonstra que os resultados das estimativas são significativos [valor- $p > 0,05$] para os parâmetros estimados \hat{d} e $\hat{\phi}_1$. E ao testar a hipótese alternativa para análise dos resíduos de Box-Pierce e de normalidade de Shapiro-Wilk, ambos os testes com p-valor maior que 5%, não apresentaram evidências para rejeitar a hipótese nula, e o modelo ARFIMA($1, d, 0$) é adequado, após retirada de tendência determinística.

Modelo 3: Considerando o ajuste de tendência determinística no Modelo 2, após a retirada da mesma foi ajustado também um modelo $SARMA(p,q) \times (P,Q)_s$, cujos coeficientes encontram-se na Tabela 7.

Tabela 7 – Resultado do ajuste do modelo $SARMA(3,1) \times (2,2)_{12}$, após retirada da tendência determinística.

Parâmetros	Estimativa	Erro Padrão	Estat. Z	P(> z)
ϕ_1	1,5135	0,1256	12,0531	< 2,2e-16
ϕ_2	-0,7187	0,1315	-5,4641	4,652e-08
ϕ_3	0,1687	0,0605	2,7901	0,0053695
θ_1	-0,6319	0,1150	-5,4938	3,933e-08
Φ_1	0,3307	0,1666	1,9852	0,0471225
Φ_2	-0,7783	0,2188	-3,5577	0,0003741
Θ_1	-0,3966	0,1878	-2,1118	0,0347044
Θ_2	0,6809	0,2569	2,6500	0,0080484

Log Verossimilhança = -1100,78, AIC = 2219,56
 Teste de Resíduos (Box-Pierce): df = 15 – p-valor = 0,3506
 Teste de Normalidade (Shapiro-Wilk): W=0,9918 – p-valor = 0,1163

Fonte: Os Autores.

Como no modelo 3 os parâmetros estimados (Tabela 7) são todos significativos ao nível de significância de 5%, e não apresentaram evidências para rejeitar as hipóteses nulas dos testes de Box-Pierce e de Shapiro-Wilk, com ausência de tendência determinística o ajuste do modelo $SARMA(3,1) \times (2,2)_{12}$ é adequado.

Modelo 4: Modelo $SARIMA(p,d,q) \times (P,D,Q)_s$. Considerando tendência estocástica, foi ajustado, como apresentado na Tabela 8, o modelo $SARIMA(1,1,1) \times (2,1,1)_{12}$.

A Tabela 8 apresentou significância em todos seus parâmetros estimados, e não houve evidências para aceitar as hipóteses alternativas do teste de resíduos de Box-Pierce e de normalidade de Shapiro-Wilk, e modelo $SARIMA(1,1,1) \times (2,1,1)_{12}$ foi o adequado.

Tabela 8 – Resultado do ajuste do Modelo SARIMA(1,1,1)×(2,1,1)₁₂.

Parâmetros	Estimativa	Erro Padrão	Estat. Z	P(> z)
ϕ_1	0,6225	0,1095	5,6830	1,324e-08
θ_1	-0,8076	0,0813	-9,9393	<2,2e-16
Φ_1	-0,1324	0,0616	-2,1491	0,0316233
Φ_2	-0,2303	0,0618	-3,7268	0,0001939
θ_1	-0,9998	0,0718	-13,9338	<2,2e-16

Log Verossimilhança = -1071,52, AIC = 21,5503
 Teste de Resíduos (Box-Pierce): df = 15 – p-valor = 0,1556
 Teste de Normalidade (Shapiro-Wilk): W=0,9949 – p-valor = 0,4681

Fonte: Os Autores.

Na aplicação do modelo de suavização exponencial, foram analisados os resultados dos três parâmetros estimados para os Modelos 5 e 6, como seguem.

Modelo 5: Modelo de Suavização Exponencial Holt-Winters com efeitos sazonais aditivos: as estimativas para os parâmetros do modelo encontram-se na Tabela 9.

Tabela 9 – Resultado do ajuste do Modelo de Suavização Exponencial Holt-Winters com efeitos sazonais aditivos.

Parâmetro	α	β	γ
Estimativa	0,8340	0,0000	0,7335

Fonte: Os Autores.

Para o modelo de suavização exponencial Aditivo, é importante esclarecer que a amplitude da variação sazonal da série é constante ao decorrer do tempo.

Modelo 6: Modelo de Suavização Exponencial Holt-Winters com efeitos sazonais multiplicativos: as estimativas para os parâmetros do modelo encontram-se na Tabela 10.

Tabela 10 – Resultado do ajuste do Modelo de Suavização Exponencial Holt-Winters com efeitos sazonais multiplicativos.

Parâmetro	α	β	γ
Estimativa	0,8046	0,0000	0,6802

Fonte: Os Autores.

Para o modelo de suavização exponencial Multiplicativo, a amplitude da variação sazonal se expande com o acréscimo do nível médio da série.

Na Tabela 11, encontram-se os resultados gerados pelas medidas de Acurácias para predição dos seis (6) modelos ajustados. Seguindo os critérios ranqueados, utilizando as seguintes medidas de acurácia: em primeiro o MAPE, segundo o U de Theil e terceiro o RMSE. O modelo ajustado que produziu melhores resultados, foi o modelo $SARIMA(1,1,1)\times(2,1,1)_{12}$, representado na Tabela 11 pelo Modelo 4.

Tabela 11 – Medidas de Acurácias para predição dos Modelos 1 a 6.

Modelos	RMSE	MAE	MAPE	U de Theil
Modelo 1	12,0918	9,4714	2,3307	0,0148
Modelo 2	12,2374	9,4032	2,3198	0,0150
Modelo 3	11,7898	9,1177	2,2496	0,0144
Modelo 4	11,5557	9,0532	2,2344	0,0141
Modelo 5	13,8770	10,9274	2,7093	0,0170
Modelo 6	14,0031	11,0154	2,7302	0,0172

Fonte: Os Autores.

Foi demonstrado pela Figura 3 as predições (dentro da amostra) e previsões (passos fora da amostra) para o Modelo 4, e se verificou que o comportamento das predições do modelo $SARIMA(1,1,1)\times(2,1,1)_{12}$ é similar a trajetória dos dados reais.

Figura 3 – Predição e Previsão da Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de jul/1995 à fev/2020, utilizando o Modelo 4.



Fonte: Os Autores.

Nas Tabelas 12 a 16, são apresentados os resultados das combinações de previsões utilizando a predição de um passo à frente, para o período de jul/1995 à fev/2019, sendo que estas são resultados da combinação de 2, 3, 4, 5 e 6 modelos de previsões individuais e para cada método de combinação. Os resultados foram ranqueados utilizando as seguintes medidas de acurácia: em primeiro MAPE, segundo U de Theil e terceiro RMSE. Os coeficientes e pesos das combinações são apresentados no Apêndice A.

Seguindo o ranqueamento das medidas de Acurácias descritos acima, na qual, os melhores resultados são os mais próximos de zero (0), na combinação de predição de dois (2) modelos expostos na Tabela 12, à combinação realizada pelos modelos quatro (4) e seis (6) geraram os melhores resultados quando combinados pelos métodos de Regressão Quantílica (RQ) e de Mínimos Quadrados Ordinários (MQO).

Tabela 12 – Medidas de Acurácia para predição para as Combinações de Previsão utilizando dois (2) modelos de previsões individuais para cada método de combinação.

Combinação de Previsão	Modelos		RMSE	MAE	MAPE	U de Theil
Média Aritmética Simples – MS	3	4	11,5590	9,0150	2,2229	0,0142
Mediana – MED	3	4	11,5590	9,0150	2,2229	0,0142
Variância Mínima – VM	3	4	11,5491	9,0105	2,2218	0,0142
Regressão – MQO	4	6	11,3795	8,9070	2,1960	0,0140
Regressão Robusta – MQP	4	6	11,4149	8,8958	2,1915	0,0140
Regressão Robusta – MM	4	6	11,3814	8,9012	2,1918	0,0140
Regressão Quantílica – RQ	4	6	11,4360	8,8837	2,1880	0,0140
Inverso do Rank – InvR	3	4	11,5032	8,9965	2,2188	0,0141
Auto Vetor cor. p/ Viés – EIG2	3	4	11,5647	9,0179	2,2236	0,0142

Fonte: Os Autores.

Na combinação de predição de três (3) modelos, expostos na Tabela 13, a combinação realizada pelo modelo um (1), quatro (4) e cinco (5) geraram os melhores resultados, quando combinados pelos métodos de Regressão Quantílica (RQ) e de Mínimos Quadrados Ordinários (MQO).

Tabela 13 – Medidas de Acurácia para predição para as Combinações de Previsão utilizando três (3) modelos de previsões individuais para cada método de combinação.

Combinação de Previsão	Modelos			RMSE	MAE	MAPE	U de Theil
Média Aritmética Simples – MS	1	2	4	11,6686	9,0706	2,2340	0,0143
Mediana – MED	1	2	4	11,6834	9,0671	2,2337	0,0143
Variância Mínima – VM	1	2	4	11,6507	9,0595	2,2312	0,0143
Regressão – MQO	1	4	5	11,2618	8,8654	2,1827	0,0138
Regressão Robusta – MQP	3	4	5	11,3563	8,8270	2,1744	0,0139
Regressão Robusta – MM	3	4	5	11,2948	8,8518	2,1775	0,0139
Regressão Quantílica – RQ	1	4	5	11,3205	8,7995	2,1637	0,0139
Inverso do <i>Rank</i> – InvR	1	3	4	11,5160	9,0053	2,2188	0,0141
Auto Vetor cor. p/ Viés – EIG2	1	2	4	11,6793	9,0810	2,2375	0,0143

Fonte: Os Autores.

Para a predição de quatro (4) modelos, Tabela 14, a combinação realizada pelo modelo um (1), quatro (4), cinco (5) e seis (6) geraram bons resultados quando combinados pelos métodos de Regressão Quantílica (RQ). No entanto, foi gerada uma segunda combinação entre os modelos um (1), três (3), quatro (4) e cinco (5) pelo método de Mínimos Quadrados Ordinários (MQO).

Tabela 14 – Medidas de Acurácia para predição para as Combinações de Previsão utilizando quatro (4) modelos de previsões individuais para cada método de combinação.

Combinação de Previsão	Modelos				RMSE	MAE	MAPE	U de Theil
Média Aritmética Simples – MS	1	2	3	4	11,7273	9,1127	2,2456	0,0144
Mediana – MED	1	2	3	4	11,7995	9,1313	2,2515	0,0145
Variância Mínima – VM	1	2	3	4	11,7103	9,1030	2,2432	0,0144
Regressão – MQO	1	3	4	5	11,2618	8,8652	2,1826	0,0138
Regressão Robusta – MQP	2	3	4	5	11,2917	8,8335	2,1742	0,0139
Regressão Robusta – MM	1	3	4	5	11,2657	8,8544	2,1781	0,0138
Regressão Quantílica – RQ	1	4	5	6	11,2984	8,7922	2,1609	0,0139
Inverso do <i>Rank</i> – InvR	1	2	3	4	11,5606	9,0132	2,2210	0,0142
Auto Vetor cor. p/ Viés – EIG2	1	2	3	4	11,7411	9,1235	2,2488	0,0144

Fonte: Os Autores.

Para a predição de cinco (5) modelos, Tabela 15, a combinação realizada pelo modelo um (1), dois (2) quatro (4), cinco (5) e seis (6) geraram bons resultados quando combinados pelos métodos de Regressão Quantílica (RQ). E para a segunda combinação entre os modelos um (1), três (3), quatro (4), cinco (5) e seis (6) o método utilizado para gerar bons resultados foi o de Mínimos Quadrados Ordinários (MQO).

Tabela 15 – Medidas de Acurácia para predição para as Combinações de Previsão utilizando cinco (5) modelos de previsões individuais para cada método de combinação.

Combinação de Previsão	Modelos					RMSE	MAE	MAPE	U de Theil
Média Aritmética Simples – MS	1	2	3	4	5	11,9112	9,3074	2,2972	0,0146
Mediana – MED	1	2	3	4	6	11,8335	9,2138	2,2718	0,0145
Variância Mínima – VM	1	2	3	4	6	11,8355	9,2479	2,2819	0,0145
Regressão – MQO	1	3	4	5	6	11,2074	8,8731	2,1847	0,0138
Regressão Robusta – MQP	2	3	4	5	6	11,2478	8,8455	2,1779	0,0138
Regressão Robusta – MM	1	2	3	4	5	11,2584	8,8602	2,1796	0,0138
Regressão Quantílica – RQ	1	2	4	5	6	11,2908	8,7906	2,1606	0,0139
Inverso do Rank – InvR	1	2	3	4	5	11,6544	9,1031	2,2449	0,0143
Auto Vetor cor. p/ Viés – EIG2	1	2	3	4	5	11,9373	9,3306	2,3045	0,0146

Fonte: Os Autores.

Na de predição dos seis (6) modelos, Tabela 16, predominaram o método de RQ e o MQO, gerando os melhores resultados entre as combinações.

Tabela 16 – Medidas de Acurácia para predição para as Combinações de Previsão utilizando os seis (6) modelos de previsões individuais para cada método de combinação.

Combinação de Previsão	RMSE	MAE	MAPE	U de Theil
Média Aritmética Simples – MS	12,1043	9,4840	2,3436	0,0149
Mediana – MED	11,9128	9,3492	2,3081	0,0146
Variância Mínima – VM	11,9929	9,3832	2,3173	0,0147
Regressão – MQO	11,2052	8,8816	2,1869	0,0137
Regressão Robusta – MQP	11,2579	8,8760	2,1881	0,0138
Regressão Robusta – MM	11,2132	8,8565	2,1800	0,0138
Regressão Quantílica – RQ	11,3051	8,7872	2,1601	0,0139
Inverso do Rank – InvR	11,7322	9,1806	2,2652	0,0144
Auto Vetor cor. p/ Viés – EIG2	12,1611	9,5347	2,3583	0,0149

Fonte: Os Autores.

Foram gerados cinco gráficos representados pelas Figuras 4 a 8. E por essas Figuras foram esboçadas as predições e previsões combinadas geradas pelo método de Regressão Quantílica (QR) e de Mínimos Quadrados Ordinários (MQO).

Figura 4 – Predição e Previsão da Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de jul/1995 à fev/2020, utilizando as Combinações de Previsão, com dois (2) modelos, por Regressão Quantílica e Mínimos Quadrados.



Fonte: Os Autores.

Figura 5 – Predição e Previsão da Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de jul/1995 à fev/2020, utilizando as Combinações de Previsão, com três (3) modelos, por Regressão Quantílica e Mínimos Quadrados.



Fonte: Os Autores.

Figura 6 – Predição e Previsão da Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de jul/1995 à fev/2020, utilizando as Combinações de Previsão, com quatro (4) modelos, por Regressão Quantílica e Mínimos Quadrados.



Fonte: Os Autores.

Figura 7 – Predição e Previsão da Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de jul/1995 à fev/2020, utilizando as Combinações de Previsão, com cinco (5) modelos, por Regressão Quantílica e Mínimos Quadrados.



Fonte: Os Autores.

Figura 8 – Predição e Previsão da Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de jul/1995 à fev/2020, utilizando as Combinações de Previsão, com seis (6) modelos, por Regressão Quantílica e Mínimos Quadrados.



Fonte: Os Autores.

Nas Figuras 4 a 8 (vistas anteriormente), as predições (dentro da amostra) e previsões (passos à frente fora da amostra) combinadas pelo método de Regressão Quantílica (RQ) e de Mínimos Quadrados Ordinários (MQO), obtiveram comportamentos similares na trajetória das predições. E as combinações feitas pelo método MQO foram as que tiveram seus comportamentos mais próximos da série original.

As Tabelas 17 a 19 apresentam as medidas de acurácia para previsão utilizando o modelo e as combinações de previsão que obtiveram melhores resultados. Ressaltamos que para o cálculo de tais medidas foi utilizado o período de março a junho de 2019.

Tabela 17 – Medidas de Acurácia de Previsão para o Modelo 4.

RMSE	MAE	MAPE	U de Theil
23,283	22,829	4,6984	0,0246

Fonte: Os Autores.

Tabela 18 – Medidas de Acurácia de Previsão para a Combinação de Previsão utilizando Regressão Quantílica - RQ.

Modelos	RMSE	MAE	MAPE	U de Theil
4 e 6	11,4360	8,8837	2,1880	0,0140
1, 4 e 5	11,3205	8,7995	2,1637	0,0139
1, 4, 5 e 6	11,2984	8,7922	2,1609	0,0139
1, 2, 4, 5 e 6	11,2908	8,7906	2,1606	0,0139
Todos	11,3051	8,7872	2,1601	0,0139

Fonte: Os Autores.

Tabela 19 – Medidas de Acurácia de Previsão para a Combinação de Previsão utilizando Regressão por Mínimos Quadrados - MQO.

Modelos	RMSE	MAE	MAPE	U de Theil
4 e 6	11,3795	8,9070	2,1960	0,0140
1, 4 e 5	11,2618	8,8654	2,1827	0,0138
1, 3, 4 e 5	11,2618	8,8652	2,1826	0,0138
1, 3, 4, 5 e 6	11,2074	8,8731	2,1847	0,0138
Todos	11,2052	8,8816	2,1869	0,0137

Fonte: Os Autores.

Pelos resultados acurados presentes nas Tabelas anteriores (17 a 19) foi possível verificar que os resultados gerados para o Modelo 4 foram bastante elevados. Para os modelos combinados pelo método de Regressão Quantílica (RQ) e de Mínimos Quadrados Ordinários (MQO), os modelos foram melhores ajustados.

Nas Tabelas 20 a 21 são apresentados os resultados das combinações de previsões, para o período de mar/2019 à jun/2019, sendo que estas são resultados da combinação de 2, 3, 4, 5 e 6 modelos de previsões individuais e para cada método de combinação. Os resultados foram ranqueados utilizando as seguintes medidas de acurácia: em primeiro MAPE, segundo U de Theil e terceiro RMSE.

Tabela 20 – Previsão da Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de mar/2019 à fev/2020 utilizando Combinação de Previsões, por meio da Regressão Quantílica - RQ.

Real		Modelos				
Data	Valor	4 e 6	1, 4 e 5	1, 4, 5 e 6	1, 2, 4, 5 e 6	Todos
mar/2019	473,94	456,5368	453,9037	451,0411	452,5583	454,9686
abr/2019	489,49	458,1663	454,8314	453,9629	453,2648	455,1795
mai/2019	485,72	460,8958	457,4912	455,9007	456,7888	459,3166
jun/2019	489,97	462,9289	460,1252	453,5893	460,1278	461,9879
jul/2019	482,26	456,7007	456,2132	453,5872	456,5476	458,2067
ago/2019		463,8645	462,1244	450,7726	462,1778	464,5935
set/2019		465,8816	463,2932	452,0170	463,4515	467,5326
out/2019		463,6932	460,2644	453,6174	460,3817	463,9505
nov/2019		464,3348	461,0675	457,6117	461,9238	468,6138
dez/2019		471,6459	468,0368	454,4702	469,1672	476,5142
jan/2020		466,4333	464,2145	458,7176	465,5370	474,1404
fev/2020		464,5704	463,8889	456,8652	465,2848	473,3257

Fonte: Os Autores.

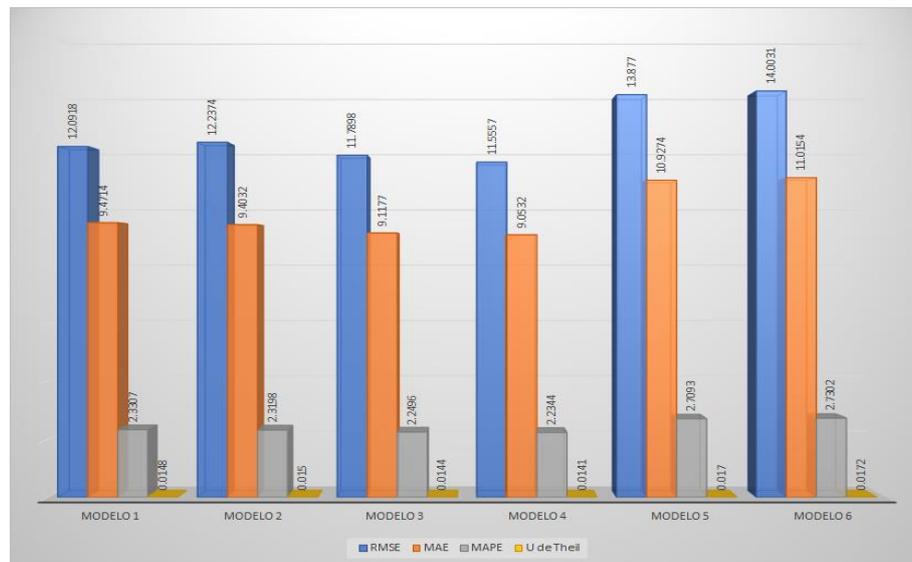
Tabela 21 – Previsão da Série Temporal dos Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, RS, no período de mar/2019 à fev/2020 utilizando Combinação de Previsões, por meio da Regressão por Mínimos Quadrados - MQO.

Real		Modelos				
Data	Valor	4 e 6	1, 4 e 5	1, 3, 4 e 5	1, 3, 4, 5 e 6	Todos
mar/2019	473,94	458,3344	455,6546	451,1758	453,0683	454,8392
abr/2019	489,49	459,9142	456,7525	456,7505	454,0876	455,2072
mai/2019	485,72	462,6598	459,5109	459,3322	458,4541	459,6509
jun/2019	489,97	464,7905	462,0835	454,9499	462,2955	462,2810
jul/2019	482,26	458,4749	457,9244	453,4285	458,2358	456,8771
ago/2019		465,9637	463,7867	446,0622	463,3455	462,5363
set/2019		468,0358	465,0003	446,9483	464,5361	464,7499
out/2019		465,6254	462,1787	453,3631	462,1618	461,5660
nov/2019		466,2975	462,9609	456,5413	463,6525	465,2638
dez/2019		473,9660	469,8133	447,6558	470,6421	472,8441
jan/2020		468,6081	465,9005	452,8665	466,6008	469,3600
fev/2020		466,8185	465,3491	447,8149	465,8001	467,5704

Fonte: Os Autores.

A Figura 9, expressa os resultados originados pelas medidas de Acurácia para os seis modelos individuais que passaram no teste de resíduos e obtiveram seus parâmetros significativos, sendo os mesmos elaborados por meio das observações propostas neste estudo.

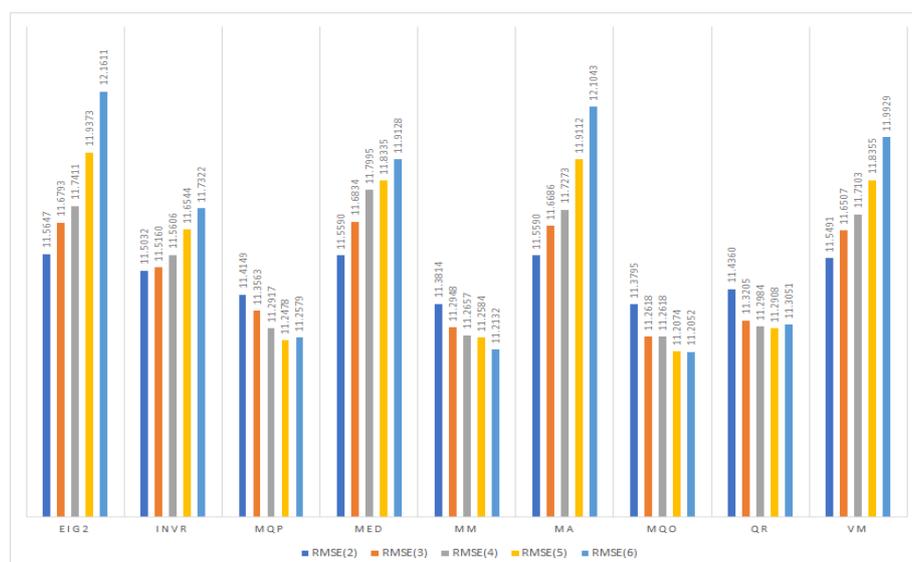
Figura 9 – Medidas de Acurácias dos seis modelos ajustados para os procedimentos de combinações de previsões.



Fonte: Os Autores.

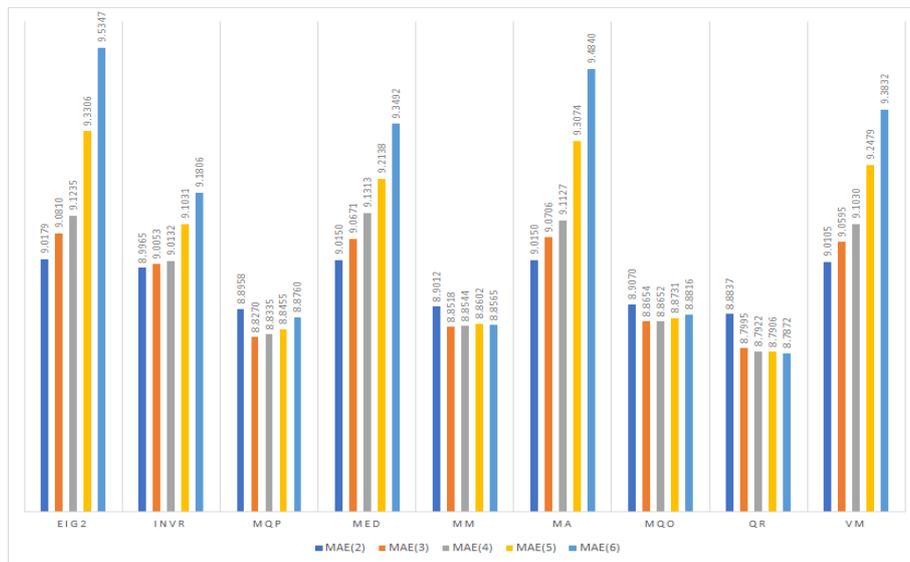
Os gráficos representados pelas Figuras 10 a 13, demonstram os resultados dos nove métodos de combinações de previsões, por meio das quatro medidas de Acurácia (RMSE, MAE, MAPE e U de Theil). Maiores detalhes sobre as medidas, incluindo os modelos que foram utilizados na combinação podem ser encontrados no Apêndice B.

Figura 10 – Menores RMSE por métodos de combinações de previsão e por número de previsões combinadas.



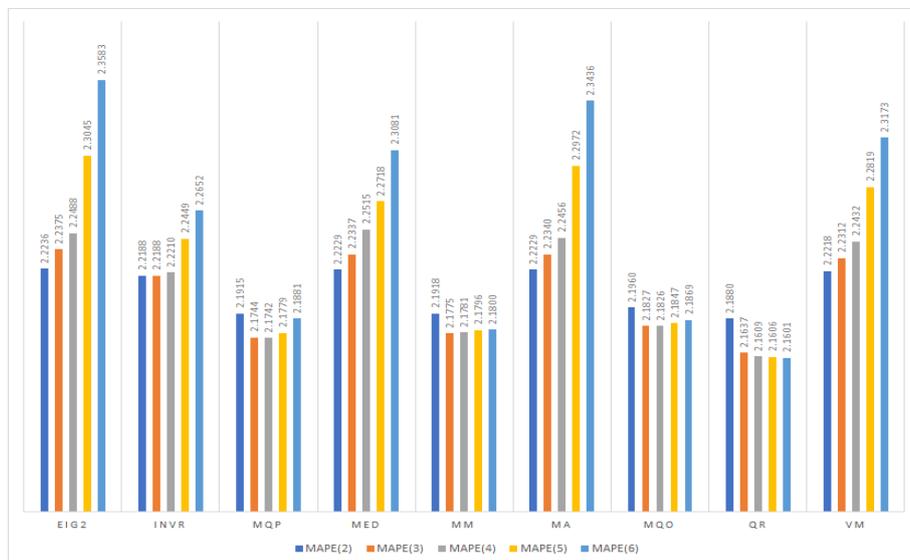
Fonte: Os Autores.

Figura 11 – Menores MAE por métodos de combinações de previsão e por número de previsões combinadas.



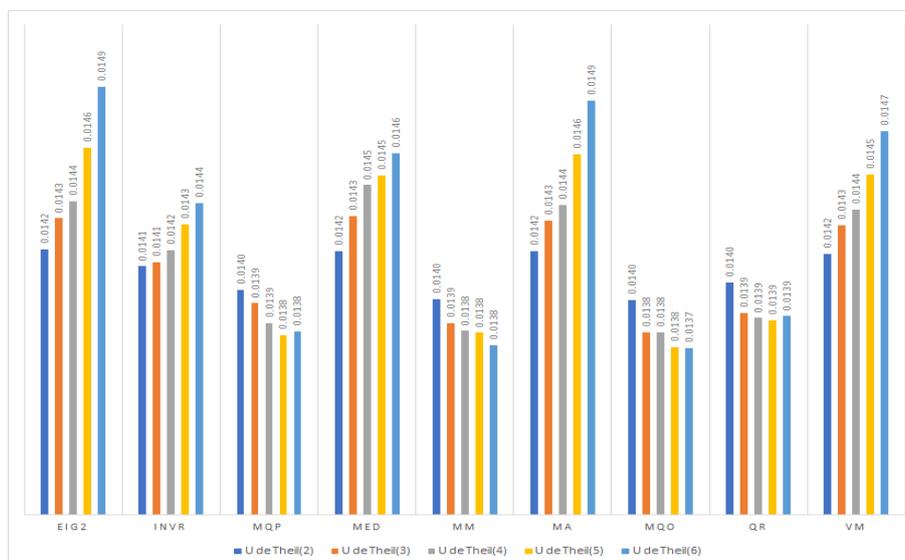
Fonte: Os Autores.

Figura 12 – Menores MAPE por métodos de combinações de previsão e por número de previsões combinadas.



Fonte: Os Autores.

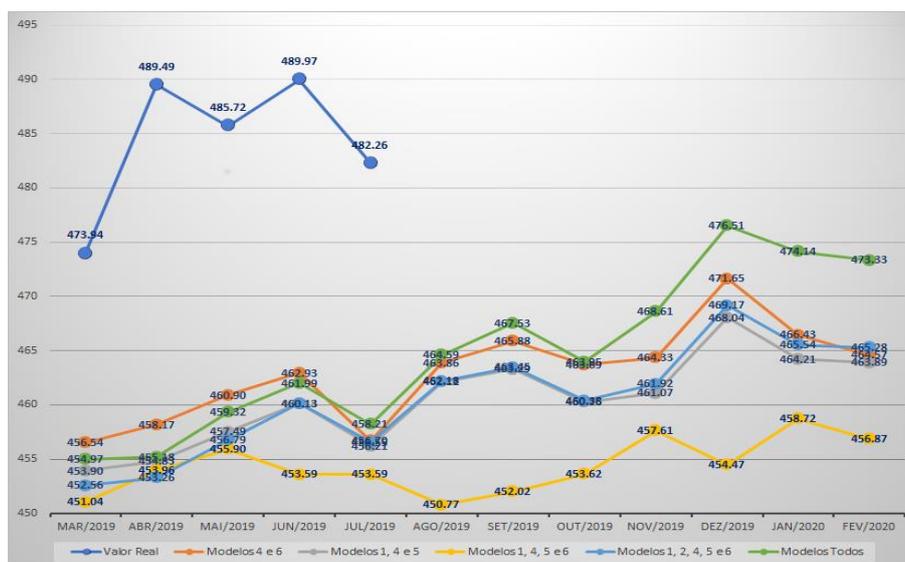
Figura 13 – Menores U de Theil por métodos de combinações de previsão e por número de previsões combinadas.



Fonte: Os Autores.

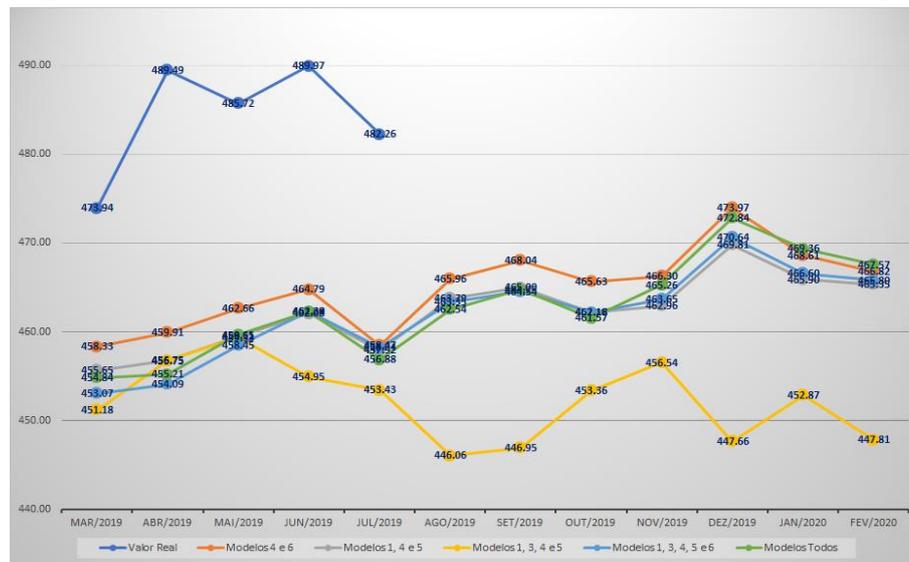
As figuras 14 e 15 apresentam cinco meses (mar/2019 a jul/2019) da série original da variável Gastos Médios Mensais no Consumo da Cesta Básica de Porto Alegre, para comparar com os resultados de previsões combinadas dos dois melhores métodos (Regressão Quantílica e Mínimos Quadrados) gerados pelos valores mais próximos de zero, resultantes das medidas de Acurácia.

Figura 14 – Previsões utilizando o método Combinações de Previsões por Regressão Quantílica.



Fonte: Os Autores.

Figura 15 – Previsões utilizando o método Combinações de Previsões por Mínimos Quadrados.



Fonte: Os Autores.

Na seção seguinte, foram feitas as conclusões e considerações do estudo proposto realizado sobre os gastos médios mensais no consumo da Cesta Básica de Porto Alegre, RS.

5. CONSIDERAÇÕES FINAIS

Ao concluir a análise dos resultados estatísticos para este estudo, verificou-se que entre a seleção do melhor modelo individual o que apresentou resultados acurados mais próximos à zero foi o Modelo 4: SARIMA(1,1,1)×(2,1,1)₁₂, que demonstrou-se presente nas combinações de previsões de todos os melhores modelos calculados. Pelas medidas de Acurácia para predição das combinações de previsão utilizando dois, três, quatro, cinco e os seis modelos de previsões individuais para cada método de combinação, os métodos combinados que obtiveram os melhores valores foram o por Regressão Quantílica (RQ) e por Mínimos Quadrados Ordinários (MQO).

Em relação às medidas de Acurácias para previsões, constatou-se que o Modelo 4 não apresentou resultados favoráveis para calcular as previsões, pois obteve resultados bastante elevados em relação aos das combinações de previsões.

Por meio das medidas de Acurácia de previsão para a combinação de previsão, o método por Regressão Quantílica foi um dos que apresentou dois modelos com os resultados mais próximos à zero entre os demais. Sendo uma das combinações representadas pelos Modelos 1, 2, 4, 5 e 6 e a outra pela combinação dos seis modelos.

Outro método que obteve medidas de Acurácia de previsão para a combinação de previsão duas combinações de modelos com resultados também próximos à zero, foi o método por Regressão de Mínimos Quadrados Ordinários. O qual obteve uma combinação pelos Modelos 1, 3, 4 e 5 e a outra também foi composta pela combinação de seis modelos (assim como pelo método RQ).

Analisando os maiores valores (em R\$) das previsões realizadas pelas combinações dos modelos por meio do método de Regressão Quantílica e de Mínimos Quadrados Ordinários, os resultados previstos que mais se aproximaram do valor real da série, foram os das combinações dos Modelos 4 e 6 para ambos os métodos.

No entanto, em relação aos dois métodos (RQ e MQO) calculados, quando verificado as predições com a série de dados reais da série temporal, gastos médios mensais no consumo da Cesta Básica de Alimentos de Porto Alegre, o método que obteve comportamento mais próximo da trajetória dos dados reais da série foi o método de Mínimos Quadrados Ordinários (MQO).

Diante desses resultados, podemos considerar que as combinações de previsões produzidas pelos seus respectivos métodos, ofereceram contribuições muito importantes para o presente estudo, pois sendo essa uma variável que expressa à variação dos preços

de produtos alimentícios, e que depende de tantas outras variáveis que podem sofrer variações periódicas em um ano, como por exemplo, o clima, as tributações, o combustível, os fretes, etc., os resultados previstos e apresentados neste estudo, quando comparados aos preços cobrados de mar/2019 a jul/2019 foram satisfatórios.

Um fator importante para trabalhos futuros pode ser o estudo da série histórica de cada um dos treze produtos que compõem a Cesta Básica de Alimentos, para identificar: 1) Qual dessas 13 variáveis (produtos) obteve maior impacto (“peso”) no comportamento da série de Índice de preços? e 2) Qual das 13 variáveis (produtos) obteve maior volatilidade (variação) na série Índice de Preços? Essas questões podem ser analisadas e avaliadas pelos modelos da família ARCH (GAGRCH, APARCH, TGARCH).

Isso demonstra que, ainda há muito que explorar (estudar) a respeito dessa variável, pois sendo ela dependente de muitas outras, que não foram tratados nesta pesquisa, seria interessante desenvolver futuros estudos por combinações de previsões, na tentativa de se chegar a resultados de previsões mais próximos possíveis dos preços cobrados na realidade.

REFERENCIAS

AMORIM, A. L.; SOUSA, E. P.; CORONEL, D. A. **Preço da cesta básica na Região Sul do Brasil: testando a integração espacial**. Revista de Economia e Administração, v.12, n.3, 321-348p, jul./set. 2013.

AMORIM, H. R. **EM TEMPOS DE GUERRA: Josué de Castro e as políticas públicas de alimentação no Estado Novo**. CLIO: Revista de Pesquisa Histórica - CLIO (Recife), ISSN: 2525-5649, n°. 35, p. 51-75, Jan-Jun, 2017.

ANTUNES, J. L. F.; CARDOSO, M. R. A. **Uso da análise de séries temporais em estudos epidemiológicos**. Epidemiol. Serv. Saúde, Brasília, 24(3):565-576, jul-set 2015.

BANERJEE, Anindya *et al.* **Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data**, Oxford University Press, Oxford, 1993.

BARRETO, Ricardo C.; HOLANDA, Fabrício José, C.. **Variáveis Determinantes do Custo da Cesta Básica em Fortaleza**. Revista Econômica do Nordeste, Fortaleza, v. 38, n° 2, abr-jun. 2007.

BARTLETT, M. S. **“On the theoretical specification of sampling properties of autocorrelated time series”**. Journal of the Royal Statistical Society, Série B, 1946. v. 27, p. 27-41.

BATES, J. M.; GRANGER, C. W. J. **The combination of forecasts. Operational Research Quarterly**, v. 20, 1969, p. 451-468.

BERAN, Jan. **Statistics for Long-Memory Process**. Chapman & Hall. New York, 1994.

BISOGNIN, C. **ESTIMAÇÃO E PREVISÃO EM PROCESSOS SARFIMA $(p,d,q) \times (P,D,Q)_S$ NA PRESENÇA DE OUTLIERS**. 2007. 222 f. Tese (Doutorado em Matemática)–Universidade Federal do Rio Grande do Sul, Porto Alegre, 2007.

BLIEMEL, Friedhelm. **Theil's Forecast Accuracy Coefficient: A Clarification**. JOURNAL OF MARKETING RESEARCH, NOVEMBER 1973.

BOX, G.E.P.; JENKINS, G.M.; REINSEL, G. **Time Series Analysis: Forecasting and Control**. Third Edition. Englewood: Prentice Hall, 1994.

BRASIL. **DECRETO-LEI N. 399 DE 30 DE ABRIL DE 1938**. Disponível em: <https://www2.camara.leg.br/legin/fed/decllei/1930-1939/decreto-lei-399-30-abril-1938-348733-publicacaooriginal-1-pe.html>>. Acesso em: 17 jan. 2018.

BRESSAN, Aurelino, Angel. **TOMADA DE DECISÃO EM FUTUROS AGROPECUÁRIOS COM MODELOS DE PREVISÃO DE SÉRIES TEMPORAIS**. RAE-eletrônica, v. 3, n. 1, Art. 9, jan./jun. 2004.

BROCKWELL, P. J.; DAVIS, R. A. **Time Series: Theory and Methods**. Springer series in Statistics, 3^o editon, 2002.

BUENO, Rodrigo De. Losso da. Silveira. **Econometria de Séries Temporais**. São Paulo: Cengage Learning, 2008.

CARVALHO, P. L. C.; FERRAZ, M. I. F.; SAFADI, T.. **VERIFICAÇÃO DA EXISTÊNCIA DA COMPONENTE SAZONAL NA SÉRIE DO CUSTO DA CESTA BÁSICA DO MUNICÍPIO DE LAVRAS, MG, UTILIZANDO O MÉTODO X-12 ARIMA**. Organizações Rurais & Agroindustriais, Lavras, v. 8, n. 2, p. 268-278, 2006.

CARVALHO, R. L. da Silva *et al.* **COMPORTAMENTO DAS SÉRIES TEMPORAIS DE TEMPERATURA DO AR, UMIDADE E PRECIPITAÇÃO PLUVIOMÉTRICA NO MUNICÍPIO DE ARIQUEMES (RONDÔNIA-BRASIL)**. Revista Brasileira de Climatologia. Ano 12 – Vol. 18 – JAN/JUN 2016.

CHATFIELD, C. **Time-series forecasting**. Boca Raton London, New York Washington, D.C., Chapman and Hall/CRC, 2000.

COSTA, J. V.; FERREIRA, M.; CORDEIRO, M. T. A. **Análise de séries temporais climáticas**. Revista de Ciências Agroveterinárias, Lages, v.14, n.2, p.169-177, 2015.

COSTANTINI, M.; PAPPALARDO, C. **A hierarchical procedure for the combination of forecasts**. International journal of forecasting, 26(4):725–743, 2010.

DIEESE, Instituto. **Metodologia da Cesta Básica de Alimentos**. São Paulo, 1993.

DIEESE, Instituto. **Metodologia da Pesquisa Nacional da Cesta Básica de Alimentos Janeiro de 2016**. São Paulo, 2016.

DIEESE, Instituto. **Pesquisa Nacional da Cesta Básica de Alimentos**. Disponível em: <<https://www.dieese.org.br/analisecestabasica/analiseCestaBasica201903.html>>. Acesso em: 18 mar. 2019.

ELLOTT, Graham; TIMMERMANN, Allan. **Economic Forecasting**. Princeton University Press, 2016.

GONÇALVES, L., Resende; SÁFADI, Thelma. **Modelagem de Séries Representativas do Setor Energético Brasileiro**. Rev. bras. Estat., Rio de Janeiro, v. 69, n. 231, p.7-32 32, jul./dez. 2008.

GUJARATI, D. N.; PORTER, D. C. **Econometria Básica**. 5^o ed. – Porto alegre: AMHG, 2011.

HAMILTON, D. James. **Time Series Analysis**. Princeton University Press, New Jersey, 1994.

HOSKING, J. R. **Fractional differencing**. Biometrika, v.68, n.1, p.165–176, 1981.

HSIAO, Cheng; WAN, S. Ki. **Is there an optimal forecast combination?** Journal of Econometrics 178 (2014) 294–309.

HYNDMAN, R. J.; KOEHLE, A. B. **Another look at measures of forecast accuracy.** International Institute of Forecasters. Elsevier, 2006.

IBGE. **Índice Nacional de Preços ao Consumidor Amplo - IPCA.** Download. Disponível em: <<https://www.ibge.gov.br/estatisticas/economicas/precos-e-custos/9256-indice-nacional-de-precos-ao-consumidor-amplo.html?t=downloads>>. Acesso em 15 ago. 2019.

KOENKER, R. **Quantile Regression.** Cambridge University Press: Cambridge, 2005.

KOENKER, R; BASSETT, G. **Regression Quantiles.** Econometrica, Vol. 46, Nº. 1. (Jan., 1978), pp. 33-50.

LAVINAS, Lena. **ACESSIBILIDADE ALIMENTAR E ESTABILIZAÇÃO ECONÔMICA NO BRASIL NOS ANOS 90.** TEXTO PARA DISCUSSÃO Nº 591. Rio de Janeiro, IPEA, 1998.

LOPES, C. M. Carvalho *et al.* **O uso de modelos de séries temporais no estudo da produção de álcool no Brasil.** Rev. bras. Estat., Rio de Janeiro, v. 70, n. 232, p.71-88, jan./jun. 2009.

MAKRIDAKIS, S.; WHEELWRIGHT, S. C.; HYNDMAN, R. J. **Forecasting: methods and applications.** John Wiley & Sons, 3rd edition, 1998.

MIKHAIL, E.; ACKERMAN, F. **Observations and Least Squares.** University Press of America, 1976. 497 p.

MINE, Tais; et al. **Análise do Valor e da Variação do Preço da Cesta Básica de Alimentos na cidade de Jacareí: um Estudo Quantitativo de Painel de Respondentes Varejistas.** Natal – RN. Congresso Internacional de Administração. 2016.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais.** São Paulo: E. Blucher, 2004.

PAIVA, E. C. *et al.* **CONVERGÊNCIA DOS PREÇOS: UMA ANÁLISE PARA AS CESTAS BÁSICAS NAS CAPITAIS DO BRASIL.** Rev. Econ. NE, Fortaleza, v. 48, n. 4, p. 91-105, out./dez., 2017.

QUENOUILLE, M. H. **“Approximate Tests of Correlation in Time Series.”** Journal of the Royal Statistical Society: Series B, 11, 68-84, 1949.

R CORE TEAM (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

ROSSI, José, W.; NEVES, Cesar das. **Econometria e Séries Temporais com Aplicação de Dados da Economia Brasileira**. Rio de Janeiro: Editora LTC, 2014.

ROUSSEEUW, P. J. "**Least Median of Square Regression**". Journal of the American Statistical Association, Vol. 79, pp. 871-880, 1984.

ROYSTON, J. P. **An Extension of Shapiro and Wilk's W Test for Normality to Large Samples**. Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 31, N^o. 2 (1982), pp. 115-124.

SAIGAL, S.; MEHROTRA, D. **PERFORMANCE COMPARISON OF TIME SERIES DATA USING PREDICTIVE DATA MINING TECHNIQUES**. Advances in Information Mining, Volume 4, Issue 1, 2012.

SARAIVA, M. Balduino *et at.* **Políticas agrícolas de garantia de preço mínimo: impactos sobre os preços do arroz no Rio Grande do Sul**. Revista Nexos Econômicos, v.8, n. 2, jul-dez., 2014, 67-79p.

SHUMWAY R.H., STOFFER D.S. **Time Series Analysis and Its Applications: With R Examples**. 3rd Edition. - Springer, 2010.- 609 p.

SILVA, Alberto Carvalho. **De Vargas a Itamar: políticas e programas de alimentação e nutrição**. Estudos Avançados 9 (23) 1995.

SILVA, M. D. V.; PIRES, M. M.; FERRAZ, M. I. F. **Análise da interdependência entre os preços da cesta básica das capitais do nordeste**. Reflexões Econômicas. v. 1, n. 1, abr./set. 2015, p. 219-247.

Timmermann, A. **Forecast combinations**. In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, Vol. 1. Elsevier, Amsterdam, 2006.

WEI, Willhiam, S. W. **Time Series Analysis: Univariate and Multivariate Methods**. Pearson Addison Wesley, Boston, 2^o ed., 2005.

WEISS, Christoph. **Essays in Hierarchical Time Series Forecasting and Forecast Combination**. 2018. 200 f. Dissertação (Doutorado em Filosofia)-Cambridge Judge Business School. Reino Unido, 2018.

WERNER, L.; BISOGNIN, C.; MENEGOTTO, L.. **PREVISÃO DA UMIDADE RELATIVA DO AR DIÁRIA DE PORTO ALEGRE**. Rio de Janeiro. Cadernos do IME – Série Estatística, 2017.

WOOLDRIDGE, Jeffrey, M. **Introdução a Econometria: uma abordagem moderna**. Tradução - Rogério César de Souza e José Antônio Ferreira. São Paulo. Thomson Learning, 2007.

ANEXO A

A Tabela A, deste anexo, apresenta os Números Índices do Índice Nacional de Preços ao Consumidor Amplo (IPCA) da Região Metropolitana de Porto Alegre do período de jul/1995 a jul/2019, os quais foram utilizados para deflacionar as observações dos gastos médios mensais no consumo da Cesta Básica de Porto Alegre.

Tabela A1 – Números Índices do Índice Nacional de Preços ao Consumidor Amplo (IPCA) da Região Metropolitana de Porto Alegre do período de jul/1995 a jul/2019.

Mês/Ano	Índices	Mês/Ano	Índices	Mês/Ano	Índices
Jul.1995	1126,76	Dez.1998	1370,54	Mai. 2002	1815,08
Ago.1995	1135,55	Jan.1999	1377,12	Jun. 2002	1819,07
Set.1995	1142,36	Fev.1999	1397,09	Jul. 2002	1843,45
Out.1995	1159,15	Mar.1999	1425,03	Ago. 2002	1851,01
Nov.1995	1173,41	Abr.1999	1445,84	Set. 2002	1865,63
Dez.1995	1187,49	Mai.1999	1456,11	Out. 2002	1893,24
Jan.1996	1195,09	Jun.1999	1460,04	Nov. 2002	1944,93
Fev.1996	1202,26	Jul.1999	1477,85	Dez. 2002	1984,22
Mar.1996	1209,95	Ago.1999	1485,83	Jan.2003	2024,30
Abr.1996	1227,25	Set.1999	1491,33	Fev. 2003	2055,47
Mai.1996	1240,75	Out.1999	1507,29	Mar. 2003	2088,15
Jun.1996	1253,65	Nov.1999	1520,70	Abr. 2003	2114,46
Jul.1996	1260,42	Dez.1999	1527,24	Mai. 2003	2130,53
Ago.1996	1263,32	Jan.2000	1536,40	Jun. 2003	2122,86
Set.1996	1268,88	Fev. 2000	1545,62	Jul. 2003	2122,86
Out.1996	1271,42	Mar. 2000	1547,32	Ago. 2003	2131,78
Nov.1996	1273,58	Abr. 2000	1555,37	Set. 2003	2145,21
Dez.1996	1277,02	Mai. 2000	1559,88	Out. 2003	2152,93
Jan.1997	1287,11	Jun. 2000	1568,77	Nov. 2003	2157,45
Fev.1997	1292,39	Jul. 2000	1600,15	Dez. 2003	2172,77
Mar.1997	1300,53	Ago. 2000	1614,87	Jan.2004	2187,33
Abr.1997	1314,45	Set. 2000	1615,84	Fev. 2004	2199,58
Mai.1997	1321,29	Out. 2000	1614,55	Mar. 2004	2219,16
Jun.1997	1328,16	Nov. 2000	1622,14	Abr. 2004	2233,81
Jul.1997	1330,42	Dez. 2000	1628,14	Mai. 2004	2253,69
Ago.1997	1332,02	Jan.2001	1637,42	Jun. 2004	2276,00
Set.1997	1332,82	Fev. 2001	1643,15	Jul. 2004	2290,11
Out.1997	1336,69	Mar. 2001	1653,34	Ago. 2004	2299,27
Nov.1997	1340,17	Abr. 2001	1666,57	Set. 2004	2310,08
Dez.1997	1344,73	Mai. 2001	1681,07	Out. 2004	2319,78
Jan.1998	1349,30	Jun. 2001	1685,95	Nov. 2004	2331,84
Fev.1998	1350,78	Jul. 2001	1702,30	Dez. 2004	2358,19
Mar.1998	1359,70	Ago. 2001	1708,26	Jan.2005	2378,47
Abr.1998	1366,09	Set. 2001	1713,38	Fev. 2005	2391,31
Mai.1998	1374,42	Out. 2001	1728,80	Mar. 2005	2402,79
Jun.1998	1376,34	Nov. 2001	1740,73	Abr. 2005	2447,72
Jul.1998	1381,02	Dez. 2001	1753,26	Mai. 2005	2456,29

Ago.1998	1374,94	Jan.2002	1761,50	Jun. 2005	2459,97
Set.1998	1371,50	Fev. 2002	1768,02	Jul. 2005	2467,60
Out.1998	1370,68	Mar. 2002	1790,30	Ago. 2005	2466,37
Nov.1998	1368,35	Abr. 2002	1806,05	Set. 2005	2473,52
Out. 2005	2486,63	Out. 2009	2928,18	Out. 2013	3637,98
Nov. 2005	2493,10	Nov. 2009	2945,75	Nov. 2013	3660,17
Dez. 2005	2503,82	Dez. 2009	2946,63	Dez. 2013	3684,69
Jan.2006	2513,84	Jan.2010	2962,84	Jan.2014	3704,22
Fev. 2006	2518,36	Fev. 2010	2984,77	Fev. 2014	3721,26
Mar. 2006	2531,46	Mar. 2010	3008,65	Mar. 2014	3755,87
Abr. 2006	2543,61	Abr. 2010	3025,50	Abr. 2014	3796,43
Mai. 2006	2549,97	Mai. 2010	3027,32	Mai. 2014	3824,90
Jun. 2006	2546,66	Jun. 2010	3022,78	Jun. 2014	3835,99
Jul. 2006	2550,23	Jul. 2010	3028,22	Jul. 2014	3837,91
Ago. 2006	2550,74	Ago. 2010	3035,49	Ago. 2014	3843,67
Set. 2006	2553,55	Set. 2010	3041,26	Set. 2014	3859,43
Out. 2006	2561,47	Out. 2010	3060,42	Out. 2014	3887,99
Nov. 2006	2564,80	Nov. 2010	3080,62	Nov. 2014	3903,15
Dez. 2006	2570,70	Dez. 2010	3097,87	Dez. 2014	3934,38
Jan.2007	2568,90	Jan.2011	3112,43	Jan.2015	3981,20
Fev. 2007	2567,87	Fev. 2011	3138,26	Fev. 2015	4026,19
Mar. 2007	2581,99	Mar. 2011	3161,80	Mar. 2015	4099,06
Abr. 2007	2592,83	Abr. 2011	3194,68	Abr. 2015	4123,65
Mai. 2007	2605,53	Mai. 2011	3210,65	Mai. 2015	4163,65
Jun. 2007	2611,52	Jun. 2011	3215,14	Jun. 2015	4194,88
Jul. 2007	2625,62	Jul. 2011	3221,57	Jul. 2015	4228,86
Ago. 2007	2634,28	Ago. 2011	3226,08	Ago. 2015	4240,70
Set. 2007	2631,12	Set. 2011	3242,53	Set. 2015	4264,45
Out. 2007	2637,43	Out. 2011	3274,31	Out. 2015	4295,58
Nov. 2007	2652,73	Nov. 2011	3293,30	Nov. 2015	4339,82
Dez. 2007	2666,26	Dez. 2011	3299,56	Dez. 2015	4375,41
Jan.2008	2669,46	Jan.2012	3309,79	Jan.2016	4443,67
Fev. 2008	2679,07	Fev. 2012	3324,35	Fev. 2016	4486,77
Mar. 2008	2698,63	Mar. 2012	3332,00	Mar. 2016	4516,83
Abr. 2008	2723,19	Abr. 2012	3357,99	Abr. 2016	4559,29
Mai. 2008	2747,43	Mai. 2012	3372,43	Mai. 2016	4601,24
Jun. 2008	2772,16	Jun. 2012	3374,45	Jun. 2016	4600,32
Jul. 2008	2789,35	Jul. 2012	3394,70	Jul. 2016	4626,54
Ago. 2008	2794,93	Ago. 2012	3410,99	Ago. 2016	4643,66
Set. 2008	2808,35	Set. 2012	3430,77	Set. 2016	4652,48
Out. 2008	2816,78	Out. 2012	3447,92	Out. 2016	4664,11
Nov. 2008	2837,34	Nov. 2012	3462,06	Nov. 2016	4681,37
Dez. 2008	2841,31	Dez. 2012	3483,52	Dez. 2016	4679,50
Jan.2009	2848,13	Jan.2013	3513,83	Jan.2017	4687,92
Fev. 2009	2859,24	Fev. 2013	3526,13	Fev. 2017	4699,17
Mar. 2009	2863,53	Mar. 2013	3542,35	Mar. 2017	4710,45
Abr. 2009	2889,59	Abr. 2013	3553,69	Abr. 2017	4720,81
Mai. 2009	2908,37	Mai. 2013	3571,46	Mai. 2017	4743,47
Jun. 2009	2912,15	Jun. 2013	3577,53	Jun. 2017	4730,19
Jul. 2009	2915,64	Jul. 2013	3581,11	Jul. 2017	4724,51
Ago. 2009	2916,51	Ago. 2013	3595,43	Ago. 2017	4740,10
Set. 2009	2924,09	Set. 2013	3618,08	Set. 2017	4743,42

Out. 2017	4758,60	Jun. 2018	4965,80	Jan.2019	5023,46
Nov. 2017	4784,77	Jul. 2018	4968,28	Fev. 2019	5031,00
Dez. 2017	4798,17	Ago. 2018	4963,31	Mar. 2019	5090,37
Jan.2018	4830,80	Set. 2018	4991,60	Abr. 2019	5132,62
Fev. 2018	4834,66	Out. 2018	5027,54	Mai. 2019	5138,78
Mar. 2018	4839,98	Nov. 2018	5006,42	Jun. 2019	5117,71
Abr. 2018	4859,34	Dez. 2018	5019,44	Jul. 2019	5145,35
Mai. 2018	4895,79				

Fonte: Site IBGE (<http://bit.do/fdRnY>).

APÊNDICE A

A Tabela A1, deste apêndice, apresenta a formulação de todas as combinações de previsões de dois modelos pelo método da Média Aritmética Simples (MA), representada por $\hat{Z}_t^{(1)}$. A Tabela A1, representa as combinações de previsões apenas para dois modelos, porém utilizando a mesma metodologia chega-se aos resultados das combinações para três, quatro, cinco e os seis modelos.

Tabela A1 – Combinações de previsões de dois modelos pelo método da Média Aritmética Simples (MS).

$\hat{Z}_t^{(1)} = \frac{\hat{Z}_t^1 + \hat{Z}_t^2}{2}$	$\hat{Z}_t^{(1)} = \frac{\hat{Z}_t^2 + \hat{Z}_t^3}{2}$	$\hat{Z}_t^{(1)} = \frac{\hat{Z}_t^3 + \hat{Z}_t^5}{2}$
$\hat{Z}_t^{(1)} = \frac{\hat{Z}_t^1 + \hat{Z}_t^3}{2}$	$\hat{Z}_t^{(1)} = \frac{\hat{Z}_t^2 + \hat{Z}_t^4}{2}$	$\hat{Z}_t^{(1)} = \frac{\hat{Z}_t^3 + \hat{Z}_t^6}{2}$
$\hat{Z}_t^{(1)} = \frac{\hat{Z}_t^1 + \hat{Z}_t^4}{2}$	$\hat{Z}_t^{(1)} = \frac{\hat{Z}_t^2 + \hat{Z}_t^5}{2}$	$\hat{Z}_t^{(1)} = \frac{\hat{Z}_t^4 + \hat{Z}_t^5}{2}$
$\hat{Z}_t^{(1)} = \frac{\hat{Z}_t^1 + \hat{Z}_t^5}{2}$	$\hat{Z}_t^{(1)} = \frac{\hat{Z}_t^2 + \hat{Z}_t^6}{2}$	$\hat{Z}_t^{(1)} = \frac{\hat{Z}_t^4 + \hat{Z}_t^6}{2}$
$\hat{Z}_t^{(1)} = \frac{\hat{Z}_t^1 + \hat{Z}_t^6}{2}$	$\hat{Z}_t^{(1)} = \frac{\hat{Z}_t^3 + \hat{Z}_t^4}{2}$	$\hat{Z}_t^{(1)} = \frac{\hat{Z}_t^5 + \hat{Z}_t^6}{2}$

Fonte: Os Autores.

APÊNDICE B

As Tabelas B1 a B5, deste apêndice, apresentam as menores medidas de acurácia (os mais próximos de zero) calculadas por intermédio dos nove métodos de combinações de previsões dos seis modelos abordados no presente estudo, sendo visíveis os melhores resultados, pelos métodos de Regressão de Mínimos Quadrados

Ordinários (MQO) e de Regressão Quantílica (RQ). As tabelas foram padronizadas com quatro casas após a vírgula e os (*) apresentam os menores valores para U de Theil.

Tabela B1 – Menores medidas de acurácia dos métodos de combinações de previsões utilizando dois modelos.

Combinação	Modelos		RMSE	MAE	MAPE	U de Theil
Média Aritmética Simples – MS	3	4	11,5590	9,0150	2,2229	0,0142
Mediana – MED	3	4	11,5590	9,0150	2,2229	0,0142
Variância Mínima – VM	3	4	11,5491	9,0105	2,2218	0,0142
Regressão – MQO	4	6	11,3795	8,9070	2,1960	0,0140*
Regressão Robusta – MQP	4	6	11,4149	8,8958	2,1915	0,0140
Regressão Robusta – MM	4	6	11,3814	8,9012	2,1918	0,0140
Regressão Quantílica – RQ	4	6	11,4360	8,8837	2,1880	0,0140
Inverso do Rank – InvR	3	4	11,5032	8,9965	2,2188	0,0141
Auto Vetor cor. p/ Viés – EIG2	3	4	11,5647	9,0179	2,2236	0,0142

Fonte: Os Autores.

Tabela B2 – Melhores resultados das medidas de acurácias pelos métodos de combinações de previsões de três modelos.

Combinação	Modelos			RMSE	MAE	MAPE	U de Theil
Média Aritmética simples – MS	1	2	4	11,6686	9,0706	2,2340	0,0143
Mediana – MED	1	2	4	11,6834	9,0671	2,2337	0,0143
Variância Mínima – VM	1	2	4	11,6507	9,0595	2,2312	0,0143
Regressão – MQO	1	4	5	11,2618	8,8654	2,1827	0,0138
Regressão Robusta – MQP	3	4	5	11,3563	8,8270	2,1744	0,0139
Regressão Robusta – MM	3	4	5	11,2948	8,8518	2,1775	0,0139
Regressão Quantílica – RQ	1	4	5	11,3205	8,7995	2,1637	0,0139
Inverso do Rank – InvR	1	3	4	11,5160	9,0053	2,2188	0,0141
Auto Vetor cor. p/ Viés – EIG2	1	2	4	11,6793	9,0810	2,2375	0,0143

Fonte: Os Autores.

Tabela B3 – Melhores resultados das medidas de acurácias pelos métodos de combinações de previsões de quatro modelos.

Combinação	Modelos				RMSE	MAE	MAPE	U de Theil
Média Aritmética Simples – MS	1	2	3	4	11,7273	9,1127	2,2456	0,0144
Mediana – MED	1	2	3	4	11,7995	9,1313	2,2515	0,0145
Variância Mínima – VM	1	2	3	4	11,7103	9,1030	2,2432	0,0144
Regressão – MQO	1	3	4	5	11,2618	8,8652	2,1826	0,0138*
Regressão Robusta – MQP	2	3	4	5	11,2917	8,8335	2,1742	0,0139
Regressão Robusta – MM	1	3	4	5	11,2657	8,8544	2,1781	0,0138
Regressão Quantílica – RQ	1	4	5	6	11,2984	8,7922	2,1609	0,0139
Inverso do Rank – InvR	1	2	3	4	11,5606	9,0132	2,2210	0,0142
Auto Vetor cor. p/ Viés – EIG2	1	2	3	4	11,7411	9,1235	2,2488	0,0144

Fonte: Os Autores.

Tabela B4 – Melhores resultados das medidas de acurácias pelos métodos de combinações de previsões de cinco modelos.

Combinação	Modelos					RMSE	MAE	MAPE	U de Theil
Média Aritmética Simples – MS	1	2	3	4	5	11,9112	9,3074	2,2972	0,0146
Mediana – MED	1	2	3	4	6	11,8335	9,2138	2,2718	0,0145
Variância Mínima – VM	1	2	3	4	6	11,8355	9,2479	2,2819	0,0145
Regressão – MQO	1	3	4	5	6	11,2074	8,8731	2,1847	0,0138*
Regressão Robusta – MQP	2	3	4	5	6	11,2478	8,8455	2,1779	0,0138
Regressão Robusta – MM	1	2	3	4	5	11,2584	8,8602	2,1796	0,0138
Regressão Quantílica – RQ	1	2	4	5	6	11,2908	8,7906	2,1606	0,0139
Inverso do <i>Rank</i> – InvR	1	2	3	4	5	11,6544	9,1031	2,2449	0,0143
Auto Vetor cor. p/ Viés – EIG2	1	2	3	4	5	11,9373	9,3306	2,3045	0,0146

Fonte: Os Autores.

Tabela B5 – Melhores resultados das medidas de acurácias pelos métodos de combinações de previsões de todos os modelos.

Combinação	RMSE	MAE	MAPE	U de Theil
Média Aritmética Simples – MS	12,1043	9,4840	2,3436	0,0149
Mediana – MED	11,9128	9,3492	2,3081	0,0146
Variância Mínima – VM	11,9929	9,3832	2,3173	0,0147
Regressão – MQO	11,2052	8,8816	2,1869	0,0137
Regressão Robusta – MQP	11,2579	8,8760	2,1881	0,0138
Regressão Robusta – MM	11,2132	8,8565	2,1800	0,0138
Regressão Quantílica – RQ	11,3051	8,7872	2,1601	0,0139
Inverso do <i>Rank</i> – InvR	11,7322	9,1806	2,2652	0,0144
Auto Vetor cor. p/ Viés – EIG2	12,1611	9,5347	2,3583	0,0149

Fonte: Os Autores.

APÊNDICE C

As Tabelas C1 a C10, neste apêndice, apresentam as estimativas dos coeficientes dos métodos de combinação de previsões utilizando os métodos de regressão: Regressão Quantílica – RQ, Regressão Robusta de Mínimos Quadrados Podados – MQP, Regressão Robusta – MM e Regressão de Mínimos Quadrados Ordinários – MQO, e as estimativas dos pesos das combinações de previsões pelos métodos: Inverso do *Rank* – InvR, Variância Mínima – VM e Auto Vetor Corrigido pelo Viés – EIG2.

Tabela C1 – Modelos de combinação de previsão: estimativas dos coeficientes β_0 , β_1 e β_2 pelos métodos de regressão.

Combinação de Previsão	Modelos		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
Regressão – MQO	4	6	4,7125	1,2965	-0,3082
Regressão Robusta – MQP	4	6	11,4370	1,1945	-0,2230
Regressão Robusta – MM	4	6	6,1440	1,3059	-0,3213
Regressão Quantílica – RQ	4	6	13,4481	1,2466	-0,2810

Fonte: Os Autores.

Tabela C2 – Modelos de combinação de previsão: estimativas dos pesos ω_1 e ω_2 atribuídos às predições.

Combinação de Previsão	Modelos		$\hat{\omega}_1$	$\hat{\omega}_2$
Variância Mínima – VM	3	4	0,4793	0,5207
Inverso do Rank – InvR	3	4	0,3333	0,6667
Auto Vetor cor. p/ Viés – EIG2	3	4	0,5112	0,4888

Fonte: Os Autores.

Tabela C3 – Modelos de combinação de previsão: estimativas dos coeficientes β_0 , β_1 , β_2 e β_3 pelos métodos de regressão.

Combinação de Previsão	Modelos			$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Regressão – MQO	1	4	5	4,5077	0,3049	1,0540	-0,3700
Regressão Robusta – MQP	3	4	5	8,3711	0,3263	1,0700	-0,4193
Regressão Robusta – MM	3	4	5	2,2501	0,3230	1,0538	-0,3830
Regressão Quantílica – RQ	1	4	5	11,3185	0,3116	1,0553	-0,3967

Fonte: Os Autores.

Tabela C4 – Modelos de combinação de previsão: estimativas dos pesos ω_1 , ω_2 e ω_3 atribuídos às predições.

Combinação de Previsão	Modelos			$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$
Variância Mínima – VM	1	2	4	0,3256	0,3179	0,3565
Inverso do Rank – InvR	1	3	4	0,1818	0,2727	0,5455
Auto Vetor cor. p/ Viés – EIG2	1	2	4	0,3381	0,3436	0,3183

Fonte: Os Autores.

Tabela C5 – Modelos de combinação de previsão: estimativas dos coeficientes $\beta_0, \beta_1, \beta_2, \beta_3$ e β_4 pelos métodos de regressão.

Combinação de Previsão	Modelos				$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
Regressão – MQO	1	3	4	5	4,4295	0,2993	0,0073	1,0526	-0,3701
Regressão Robusta – MQP	2	3	4	5	6,4869	-0,2496	0,5541	1,0310	-0,3525
Regressão Robusta – MM	1	3	4	5	5,0074	0,3381	0,0175	1,0274	-0,3955
Regressão Quantílica – RQ	1	4	5	6	10,0403	0,3655	1,0384	-1,0191	0,5880

Fonte: Os Autores.

Tabela C6 – Modelos de combinação de previsão: estimativas dos pesos $\omega_1, \omega_2, \omega_3$ e ω_4 atribuídos às predições.

Combinação de Previsão	Modelos				$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_4$
Variância Mínima – VM	1	2	3	4	0,2451	0,2393	0,2471	0,2684
Inverso do <i>Rank</i> – InvR	1	2	3	4	0,1600	0,1200	0,2400	0,4800
Auto Vetor cor. p/ Viés – EIG2	1	2	3	4	0,2523	0,2577	0,2547	0,2352

Fonte: Os Autores.

Tabela C7 – Modelos de combinação de previsão: estimativas dos coeficientes $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ e β_5 pelos métodos de regressão.

Combinação de Previsão	Modelos					$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
Regressão – MQO	1	3	4	5	6	3,0757	0,3266	0,0900	1,0038	-1,7026	1,2746
Regressão Robusta – MQP	2	3	4	5	6	4,0169	-0,1747	0,5582	1,0388	-1,5639	1,1303
Regressão Robusta – MM	1	2	3	4	5	4,9716	0,3143	-0,1372	0,1675	1,0314	-0,3884
Regressão Quantílica – RQ	1	2	4	5	6	8,8736	0,3340	0,0879	1,0031	-1,1924	0,7431

Fonte: Os Autores.

Tabela C8 – Modelos de combinação de previsão: estimativas dos pesos $\omega_1, \omega_2, \omega_3, \omega_4$ e ω_5 atribuídos às predições.

Combinação de Previsão	Modelos					$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_4$	$\hat{\omega}_5$
Variância Mínima – VM	1	2	3	4	5	0,2067	0,2018	0,2083	0,2263	0,1569
Inverso do <i>Rank</i> – InvR	1	2	3	4	5	0,1460	0,1095	0,2190	0,4380	0,0876
Auto Vetor cor. p/ Viés – EIG2	1	2	3	4	5	0,1965	0,2009	0,1980	0,1871	0,2174

Fonte: Os Autores.

Tabela C9 – Modelos de combinação de previsão: estimativas dos coeficientes $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ e β_6 pelos métodos de regressão.

Combinação de Previsão	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$
Regressão – MQO	3,0591	0,2976	-0,1628	0,2654	1,0076	-1,6182	1,2027
Regressão Robusta – MQP	3,7130	0,5019	-0,0150	0,0220	1,0492	-2,2683	1,6997
Regressão Robusta – MM	4,1799	0,3730	0,0549	0,0254	0,9830	-1,6524	1,2055
Regressão Quantílica – RQ	9,5965	0,2486	0,0452	0,1908	0,9352	-1,6733	1,2270

Fonte: Os Autores.

Tabela C10 – Modelos de combinação de previsão: estimativas dos pesos $\omega_1, \omega_2, \omega_3, \omega_4, \omega_5$ e ω_6 atribuídos às predições.

Combinação de Previsão	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_4$	$\hat{\omega}_5$	$\hat{\omega}_6$
Variância Mínima – VM	0,1791	0,1748	0,1805	0,1961	0,1360	0,1335
Inverso do <i>Rank</i> – InvR	0,1361	0,1020	0,2041	0,4082	0,0816	0,0680
Auto Vetor cor. p/ Viés – EIG2	0,1585	0,1621	0,1594	0,1534	0,1834	0,1832

Fonte: Os Autores.