

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

**ESTUDO DO CONSUMO E DO NÚMERO DE
CONSUMIDORES DE ENERGIA ELÉTRICA DO RS
POR MEIO DE COMPONENTES PRINCIPAIS E
MODELOS DE PREVISÃO**

DISSERTAÇÃO DE MESTRADO

Francisca Mendonça Souza

Santa Maria, RS, Brasil

2011

**ESTUDO DO CONSUMO E DO NÚMERO DE CONSUMIDORES DE
ENERGIA ELÉTRICA DO RS POR MEIO DE COMPONENTES
PRINCIPAIS E MODELOS DE PREVISÃO**

Francisca Mendonça Souza

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

Orientador: Prof. Dr. Luis Felipe Dias Lopes

Santa Maria, RS, Brasil

2011

**Universidade Federal de Santa Maria
Centro de Tecnologia
Programa de Pós-Graduação em Engenharia de Produção**

A Comissão Examinadora, abaixo assinada,
aprova a Dissertação de Mestrado

**ESTUDO DO CONSUMO E DO NÚMERO DE CONSUMIDORES DE
ENERGIA ELÉTRICA DO RS POR MEIO DE COMPONENTES
PRINCIPAIS E MODELOS DE PREVISÃO**

elaborada por
Francisca Mendonça Souza

como requisito parcial para obtenção do grau de
Mestre em Engenharia de Produção

COMISSÃO EXAMINADORA:

Luis Felipe Dias Lopes, Dr.
(Presidente/Orientador)
(UFSM)

Roselaine Ruviano Zanini, Dra.
(UFSM)

Castelar Braz Garcia, Dr.
(UFPEL)

Santa Maria, 20 de janeiro de 2011

Dedico este trabalho aos meus pais, a quem devo toda esta caminhada, a eles o meu reconhecimento e carinho.

Ao Adriano, pelo estímulo, experiências e ensinamentos.

Ao Acássio, o meu reconhecimento, pelo companheirismo e incentivo.

AGRADECIMENTOS

A Deus, por iluminar meu caminho e me dar forças para seguir sempre em frente;

Ao Prof. Dr. Luis Felipe Dias Lopes, pela orientação disponível dispensada a este trabalho;

Aos professores do PPGEF – UFSM e aos colegas pelo companheirismo e aprendizado em equipe;

À banca examinadora, pela contribuição e sugestões dadas a esta pesquisa;

À CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pelo auxílio financeiro.

Aos familiares, pelo incentivo e apoio, que se fizeram sempre presentes; aos colegas e amigos, que foram imprescindíveis, pois sempre se mostraram solidários e capazes de tornar as dificuldades mais amenas;

Aos meus pais que compartilharam de meus ideais e os alimentaram, pelo exemplo de dedicação e perseverança, pela compreensão, carinho e amor.

RESUMO

Dissertação de Mestrado
Programa de Pós-Graduação em Engenharia de Produção
Universidade Federal de Santa Maria

ESTUDO DO CONSUMO E DO NÚMERO DE CONSUMIDORES DE ENERGIA ELÉTRICA DO RS POR MEIO DE COMPONENTES PRINCIPAIS E MODELOS DE PREVISÃO

AUTORA: FRANCISCA MENDONÇA SOUZA
ORIENTADOR: PROF. DR. LUIS FELIPE DIAS LOPES
Data e Local da Defesa: Santa Maria, 20 de janeiro de 2011.

O crescimento econômico do estado do Rio Grande do Sul, assim como do restante do país, está intimamente ligado à distribuição e geração de energia elétrica. Desta forma, para que o sistema energético do estado não entre em colapso (falta/excesso) nas etapas de geração e fornecimento, é necessário o estudo da distribuição da energia gerada e do número de consumidores. Portanto, o objetivo deste estudo é determinar uma variável de referência, por meio da análise de componentes principais, de forma a analisar globalmente o comportamento do consumo em MWh e dos consumidores de energia elétrica do Rio Grande do Sul por meio de modelos de previsão lineares (ARIMA) e não lineares (ARCH). A análise refere-se ao número de consumidores e ao consumo de energia elétrica no estado do RS, realizado pelas três maiores distribuidoras de energia elétrica, nomeadamente a AES-Sul, CEEE e RGE, no período de janeiro de 1998 a dezembro de 2009. Os modelos matemáticos encontrados para as variáveis estudadas são capazes de modelar satisfatoriamente os dados, possibilitando uma análise de seu comportamento, capaz de acompanhar e de determinar o movimento das séries em estudo de forma global, proporcionando medidas de gerenciamento, tanto das empresas distribuidoras de fornecimento quanto do número de consumidores de cada empresa, permitindo uma análise sobre a estabilidade do processo produtivo.

Palavras-chave: Previsão; ARIMA; ARCH; Componentes Principais; Energia Elétrica.

ABSTRACT

Mastership Dissertation
Post-graduation Program in Production Engineering
Federal University of Santa Maria

STUDY OF CONSUMPTION AND NUMBER OF CONSUMERS OF ELECTRICITY THROUGH THE RS AND PRINCIPAL COMPONENTS MODELS FOR PREDICTING

AUTHOR: FRANCISCA MENDONÇA SOUZA
ADVISER: PROF. DR. LUIS FELIPE DIAS LOPES
Santa Maria, 20th january, 2011.

The economic growth of the state of Rio Grande do Sul as well as the rest of the country is closely connected with the distribution and power generation. Thus for the state's energy system does not collapse (shortage / surplus) in steps of generation and supply, it is necessary to study the distribution of energy generated and the number of consumers. Therefore, the purpose of this study is to determine a reference variable by means of principal component analysis in order to analyze the overall behavior of consumption and consumers in MWh of electricity in the Rio Grande do Sul, by means of linear models of forecast (ARIMA) and nonlinear (ARCH). The analysis refers to the number of consumers and to the consumption of electrical energy to the RS, conducted by the three largest electricity distributors, including AES-Sul, CEEE and GER from January 1998 to December 2009. The mathematical models for the variables studied are able to model satisfactorily the data, enabling an analysis of their behavior, able to monitor and determine the movement of the series under study as a whole, providing management measures, both companies supply distributor as the number of customers of each company, allowing an analysis on the stability of the production process.

Key-words: Forecasting; ARIMA; ARCH; Principal Components; Electrical Energy.

LISTA DE TABELAS

Tabela 1 – Distribuidoras e regiões de abrangência	42
Tabela 2 – Autovalores e variância explicada de cada componente para o número de consumidores de energia elétrica	51
Tabela 3 – Cargas fatoriais companhia CEEE-RS	52
Tabela 4 – Autovalores e variância explicada de cada componente para o fornecimento mensal de energia por classe de consumo	57
Tabela 5 – Cargas Fatoriais constantes nas variáveis residencial, comercial, industrial, rural, serviços públicos e outros para as três distribuidoras em estudo.....	58
Tabela 6 – Estatísticas descritivas das componentes principais para o número de consumidores e consumo de energia elétrica para as três distribuidoras	62
Tabela 7 – Teste de raiz unitária ADF, KPSS aplicados às componentes principais que representam respectivamente as distribuidoras CEEE, AES-Sul, e RGE, para o número de consumidores	69
Tabela 8 – Teste de raiz unitária ADF, KPSS aplicados às componentes principais que representam respectivamente as distribuidoras CEEE, AES-Sul, e RGE, para o fornecimento de energia (MWh)	69
Tabela 9 – Correlação entre número de consumidores das CP para CEEE, AES-Sul e RGE componente principal representativa desta distribuidora .	70
Tabela 10 – Correlação entre variáveis originais para o consumo de energia em MWh e a componente principal representativa de cada distribuidora	71
Tabela 11 – Modelo ARIMA, dos critérios AIC para distribuidora CP1 _{CEEE}	72
Tabela 12 – Resultados das estatísticas <i>F</i> e ARCH-LM	74
Tabela 13 – Modelo com heterocedasticidade do fator CP1 _{CEEE}	75
Tabela 14 – Resultados das estatísticas <i>F</i> e ARCH-LM	77
Tabela 15 – Modelo ARIMA, dos critérios AIC para distribuidora CP1 _{AES-Sul}	78
Tabela 16 – Testes para presença de heterocedasticidade	79
Tabela 17 – Modelo ARIMA, dos critérios AIC para distribuidora CP1 _{RGE}	81
Tabela 18 – Testes para presença de heterocedasticidade	82
Tabela 19 – Modelo ARIMA, dos critérios AIC para distribuidora CP2 _{CEEE}	83
Tabela 20 – Testes para presença de heterocedasticidade	85
Tabela 21 – Modelo SARIMA, dos critérios AIC para distribuidora CEEE	87
Tabela 22 – Resultados das estatísticas <i>F</i> e ARCH-LM	89
Tabela 23 – Modelo ARIMA, dos critérios AIC para distribuidora CP2 _{CEEE}	90
Tabela 24 – Resultados das estatísticas <i>F</i> e ARCH-LM	91
Tabela 25 – Modelo SARIMA, dos critérios AIC para distribuidora AES-Sul	92
Tabela 26 – Resultados das estatísticas <i>F</i> e ARCH-LM	94
Tabela 27 – Modelo SARIMA, dos critérios AIC para distribuidora AES-Sul	95
Tabela 28 – Resultados das estatísticas <i>F</i> e ARCH-LM	97
Tabela 29 – Modelo ARIMA, dos critérios AIC para distribuidora CP2 _{RGE}	98
Tabela 30 – Resultados das estatísticas <i>F</i> e ARCH-LM	100

LISTA DE FIGURAS

Figura 1 – Comportamento das funções de autocorrelação e autocorrelação parcial de um modelo AR (1)	21
Figura 2 – Comportamento das funções de autocorrelação parcial e autocorrelação de um modelo MA (1)	22
Figura 3 – Fluxograma do ciclo iterativo de Box & Jenkins	29
Figura 4 – Representação gráfica dos autovalores e percentual de variância explicada CEEE (a), AES-Sul (b) e RGE(c) respectivamente	50
Figura 5 – Gráfico das cargas fatoriais para as distribuidoras CEEE (a), AES-Sul (b) e RGE (c) para o número de consumidores de energia elétrica para o Rio Grande do Sul	54
Figura 6 – Círculo de correlação unitário para as CEEE (a), AES-Sul (b) e RGE (c) respectivamente	55
Figura 7 – Representação gráfica dos autovalores e percentual de variância explicada CEEE, AES-Sul e RGE respectivamente	56
Figura 8 – Gráfico das cargas fatoriais para as distribuidoras CEEE (a), AES-Sul (b) e RGE (c) para o número de consumidores de energia elétrica para o Rio Grande do Sul	60
Figura 9 – Círculo de correlação unitário para as CEEE (a), AES-Sul (b) e RGE (c) respectivamente	61
Figura 10 – Fatores que representam as três companhias para o número de consumidores de energia elétrica	63
Figura 11 – Fatores que representam as três companhias para o fornecimento de energia elétrica	64
Figura 12 – Correlograma, estatística de Ljung-Box - Q(K) e <i>p-valor</i> da Componente principal para número de consumidores da distribuidora CP1 _{CEEE/RS}	65
Figura 13 – Correlograma, estatística de Ljung-Box - Q(K) e <i>p-valor</i> da Componente principal para número de consumidores da distribuidora AES-Sul/RS	65
Figura 14 – Correlograma, estatística de Ljung-Box - Q(K) e <i>p-valor</i> da Componente principal para número de consumidores da para distribuidora CP1 _{RGE}	66
Figura 15 – Correlograma, estatística de Ljung-Box - Q(K) e <i>p-valor</i> da Componente principal para número de consumidores da distribuidora CP2 _{RGE} ...	66
Figura 16 – Correlograma, estatística de Ljung-Box - Q(K) e <i>p-valor</i> da Componente principal para o fornecimento de energia em MWh da distribuidora CP1 _{CEEE}	66
Figura 17 – Correlograma, estatística de Ljung-Box - Q(K) e <i>p-valor</i> da Componente principal para o fornecimento de energia em MWh da distribuidora CP2 _{CEEE}	67
Figura 18 – Correlograma, estatística de Ljung-Box - Q(K) e <i>p-valor</i> da Componente principal para o fornecimento de energia em MWh da distribuidora CP1 _{AES-Sul}	67
Figura 19 – Correlograma, estatística de Ljung-Box - Q(K) e <i>p-valor</i> da Componente principal para o fornecimento de energia em MWh da distribuidora CP1 _{RGE}	68

Figura 20 – Correlograma, estatística de Ljung-Box - Q(K) e <i>p-valor</i> da Componente principal para o fornecimento de energia em MWh da distribuidora CP2 _{RGE}	68
Figura 21 – Série residual do modelo ARIMA (1,1,0)	73
Figura 22 – Correlograma, estatística de Ljung-Box (Q(K)) e <i>p-valor</i> da série dos resíduos quadráticos do modelo ARIMA (1,1,0) para a componente principal que explica a concessionária CP1 _{CEEE}	73
Figura 23 – Comportamento residual da FAC e FACP do modelo ARCH (1)	75
Figura 24 – Correlograma, estatística de Ljung-Box (Q(K)) e <i>p-valor</i> da série dos resíduos quadráticos do modelo ARCH (1) para a componente principal que explica a concessionária CP1 _{CEEE} ...	76
Figura 25 – Valores originais, valores estimados e os resíduos para o modelo ARCH(1)	77
Figura 26 – Série residual do modelo ARIMA (1,1,1)	78
Figura 27 – Correlograma, estatística de Ljung-Box (Q(K)) e <i>p-valor</i> da série dos resíduos quadráticos do modelo ARIMA (1,1,1) para a componente principal que explica a concessionária CP1 _{AES-Sul} .	79
Figura 28 – Valores originais, valores estimados e os resíduos para o modelo ARCH(1) – CP1 _{AES-Sul}	80
Figura 29 – Comportamento residual da FAC e FACP SARIMA (1,1,0) (1,0,0) ₁₂ ...	81
Figura 30 – Correlograma, estatística de Ljung-Box (Q(K)) e <i>p-valor</i> da série dos resíduos quadráticos do modelo SARIMA (1,1,0) (1,0,0) ₁₂ para a componente principal que explica a concessionária CP1 _{RGE}	82
Figura 31 – Valores originais, valores estimados e os resíduos para o modelo ARCH(1) – CP1 _{RGE}	83
Figura 32 – Comportamento residual da FAC e FACP ARIMA (1,1,0)	84
Figura 33 – Correlograma, estatística de Ljung-Box (Q(K)) e <i>p-valor</i> da série dos resíduos quadráticos do modelo ARIMA (1,1,0) para a componente principal que explica a concessionária CP2 _{RGE}	84
Figura 34 – Valores originais, valores estimados e os resíduos para o modelo ARIMA (1, 1, 0)	85
Figura 35 – Comportamento residual da FAC e FACP SARIMA (1,1,1) (1,1,0) ₁₂ ...	87
Figura 36 – Correlograma, estatística de Ljung-Box (Q(K)) e <i>p-valor</i> da série dos resíduos quadráticos do modelo SARIMA (1,1,1) (1,1,0) ₁₂ para a componente principal que explica a concessionária CP1 _{CEEE}	88
Figura 37 – Valores originais, valores estimados e os resíduos para o modelo SARIMA (1,1,1) (1,1,0) ₁₂	89
Figura 38 – Comportamento residual da FAC e FACP ARIMA (1,1,1)	90
Figura 39 – Correlograma, estatística de Ljung-Box (Q(K)) e <i>p-valor</i> da série dos resíduos quadráticos do modelo ARIMA (1,1,1) para a componente principal que explica a concessionária CP2 _{CEEE}	91
Figura 40 – Valores originais, valores estimados e os resíduos para o modelo ARIMA (1,1,1)	92
Figura 41 – Comportamento residual da FAC e FACP do modelo SARIMA (0,1,1) (1,1,0) ₁₂	93
Figura 42 – Correlograma, estatística de Ljung-Box (Q(K)) e <i>p-valor</i> da série dos resíduos quadráticos do modelo SARIMA (0,1,1) (1,1,0) ₁₂ para a componente principal que explica a concessionária CP1 _{AES-Sul}	93

Figura 43 – Valores originais, valores estimados e os resíduos para o modelo SARIMA (0,1,1) (1,1,0) ₁₂	95
Figura 44 – Correlograma, estatística de Ljung-Box (Q(K)) e p-valor da série dos resíduos do modelo SARIMA (0,1,1) (1,1,1) ₆ para a componente principal que explica a concessionária CP1 _{RGE}	96
Figura 45 – Correlograma, estatística de Ljung-Box (Q(K)) e p-valor da série dos resíduos quadráticos do modelo SARIMA (0,1,1) (1,1,1) ₆ para a componente principal que explica a concessionária CP1 _{RGE}	96
Figura 46 – Valores originais, valores estimados e os resíduos para o modelo SARIMA (0,1,1) (1,1,1) ₆	97
Figura 47 – Correlograma, estatística de Ljung-Box (Q(K)) e p-valor da série dos resíduos do modelo SARIMA (0,1,0) (1,1,1) ₆ para a componente principal que explica a concessionária CP2 _{RGE}	98
Figura 48 – Correlograma, estatística de Ljung-Box (Q(K)) e p-valor da série dos resíduos quadráticos do modelo SARIMA (0,1,0) (1,1,1) ₆ para a componente principal que explica a concessionária CP2 _{RGE}	99
Figura 49 – Valores originais, valores estimados e os resíduos para o modelo SARIMA (0,1,1) (1,1,1) ₆	100

SUMÁRIO

RESUMO	5
ABSTRACT	6
LISTA DE TABELAS	7
LISTA DE FIGURAS	8
1 INTRODUÇÃO	12
1.1 Tema da pesquisa	13
1.2 Justificativa e importância do trabalho	13
1.3 Problema de pesquisa	14
1.4 Objetivos	14
1.4.1 Objetivo geral	14
1.4.2 Objetivos específicos	14
1.5 Delimitação do tema	15
1.6 Estrutura do trabalho	15
2 REFERENCIAL TEÓRICO	16
2.1 Análise de componentes principais	16
2.2 Modelos lineares e não lineares de previsão	18
2.2.1 Modelos não lineares de previsão	30
3 METODOLOGIA	35
4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS	39
4.1 Análise de componentes principais	40
4.2 Estatísticas descritivas das componentes principais	53
4.3 Verificação da estacionariedade das componentes principais	56
4.4 Correlação linear entre as variáveis originais e as CP selecionadas	62
4.5 Modelos de Previsão das Componentes Principais selecionadas	63
4.5.1 Modelo para o número de consumidores de energia elétrica do RS	63
4.5.2 Modelo para o consumo de energia elétrica em MWh por classe de consumo no Rio Grande do Sul para CEEE, AES-Sul e RGE	79
5 CONSIDERAÇÕES FINAIS	95
REFERÊNCIAS	97

LISTA DE ABREVIATURAS E SIGLAS

ADF	Teste de Dickey e Fuller Aumentado
AES-Sul	Empresa do grupo americano AES Corporation
ANEEL	Agência Nacional de Energia Elétrica.
AIC	Critério de Informação Akaike
AR	Autorregressivo
ARCH	Heterocedasticidade Condicional Autorregressiva
ARCH-LM	Teste Multiplicador de Lagrange
ARIMA	Autorregressivo Integrado de Média Móvel
ARIMA-ARCH	Modelo ARIMA considerando Heterocedasticidade Autorregressiva
ARMA	Autorregressivo de Média Móvel
CEEE	Companhia Estadual de Distribuição de Energia Elétrica
CP	Componente principal
EGARCH	GARCH Exponencial
Estatística F	Estatística F-Snedecor
FAC	Função de Autocorrelação
FACP	Função de Autocorrelação Parcial
GARCH	Heterocedasticidade Condicional Autorregressiva Generalizada
KPSS	Teste proposto pelos autores Kwiatkowski, Phillips, Schmidt e Shin
MA	Média Móvel
MV	Método de Máxima Verossimilhança
Q	Estatística Q de Ljung-Box
RB	Ruído Branco
RGE	Rio Grande Energia

1 INTRODUÇÃO

O emprego da energia elétrica no país teve como marco pioneiro a instalação da Usina Hidrelétrica Ribeirão do Inferno, em 1883, destinada ao fornecimento de força motriz a serviços de mineração em Diamantina, Minas Gerais; a Usina Hidrelétrica da Companhia Fiação e Tecidos São Silvestre, de 1885, no município de Viçosa, também em Minas Gerais; a Usina Hidrelétrica Ribeirão dos Macacos, em 1887, no mesmo estado; a Usina Termelétrica Velha Porto Alegre, em 1887, no Rio Grande do Sul; e a Usina Hidrelétrica Marmelos, instalada em 1889, em Juiz de Fora, Minas Gerais, por iniciativa do industrial Bernardo Mascarenhas.

Até a primeira década do século XX, foi construído no país um grande número de pequenas usinas geradoras de energia elétrica. A produção visava ao atendimento dos serviços públicos instalados nas cidades, sendo empregada, predominantemente, na iluminação pública e particular, nos bondes utilizados para o transporte coletivo e no fornecimento de força motriz a unidades industriais, sobretudo do setor têxtil.

Eram frequentes as instalações autoprodutoras nas indústrias e em unidades de consumo doméstico, no setor agrícola. Essas empresas municipais constituíram-se como resultado da iniciativa do empresariado nacional ligado à agricultura de exportação, aos serviços urbanos, principalmente iluminação e transportes, e à indústria.

A energia elétrica como objeto de estudo tornou-se imprescindível nos dias de hoje, pois o tema repercute em todos os setores da economia. Pode-se discutir a produção de energia, o planejamento energético, a previsão de demanda, os temas que envolvem o meio ambiente, a sustentabilidade, a inovação, a tecnologia, a logística, a economia, a sociedade.

Ao longo dos últimos anos, o setor elétrico brasileiro passou por mudanças estruturais em sua regulamentação, a exemplo do que ocorreu em diversos países de todo o mundo. O objetivo foi o estabelecimento de um modelo de máxima eficiência econômica, caracterizado pela existência de um ambiente competitivo que viabilizasse investimentos em expansão e garantisse o atendimento ao mercado consumidor, conforme os critérios de confiabilidade e qualidade estabelecidos pela ANEEL - Agência Nacional de Energia Elétrica.

As distribuidoras utilizam subdivisões denominadas por setores, seja este de distribuição ou de consumo, e estes setores são denominados: industrial, rural, comercial, residencial e outros, que é uma subdenominação de consumos.

Após a reestruturação da distribuição de energia elétrica, o que deixou clara a competitividade entre as companhias, houve a adoção de novas políticas para orientar suas decisões estratégicas, que determinam o sucesso no curto e no longo prazo. Em um ambiente competitivo e com incertezas, um dos maiores desafios é a alocação de recursos e incentivos sujeitos a riscos inerentes à distribuição e manutenção da potência demandada de modo que não afete o mercado consumidor. Essa incerteza induz as distribuidoras à procura da máxima eficiência, contribuindo para o planejamento e para o atendimento da demanda com qualidade e confiabilidade, indo ao encontro de soluções de problemas, resultantes da diferença entre as metas da empresa e a situação constatada.

1.1 Tema da pesquisa

O tema desta pesquisa é a utilização de modelos de previsão lineares e não lineares, aplicados às componentes principais representativas do número de consumidores, e ao volume de consumo de energia elétrica das três maiores empresas distribuidoras para o Rio Grande do Sul.

1.2 Justificativa e importância do trabalho

O crescimento econômico do estado do Rio Grande do Sul, assim como do restante do país, está intimamente ligado à distribuição e geração de energia elétrica. Desta forma, para que o sistema energético do estado não entre em colapso (falta/excesso) nas etapas de geração e fornecimento, é necessário o estudo da distribuição da energia gerada e do número de consumidores.

Este mapeamento, feito por meio de técnicas de previsão, auxilia os gestores na verificação da tendência geral do consumo e de consumidores. Assim, é possível acompanhar os movimentos dessa distribuição.

Desta forma, justifica-se o trabalho apresentado, em que é possível prever o nível das séries em estudo e, quando estas apresentarem características de volatilidade, é possível estudar o seu comportamento em relação a sua variabilidade.

A sua importância está no fato de ser possível verificar os períodos de maior e menor volatilidade – cluster de volatilidade, identificando períodos em que se deve tomar o maior cuidado, evitando excessos e/ou restrições nos períodos voláteis identificados tanto na distribuição quanto no consumo pode acarretar problemas de falta ou excesso de na distribuição.

1.3 Problema de pesquisa

O problema de pesquisa é a utilização da análise de componentes principais para prever o nível das séries em estudo e, quando possível, a volatilidade da mesma. Desta forma, procura-se reduzir a dimensionalidade do problema e estudar o comportamento geral das séries.

1.4 Objetivos

Neste item, descreve-se o objetivo geral e os objetivos específicos que nortearão a pesquisa.

1.4.1 Objetivo geral

Determinar variáveis de referência por meio da análise de componentes principais de forma a analisar globalmente o comportamento do consumo em MWh e do número de consumidores de energia elétrica do Rio Grande do Sul por meio de modelos de previsão lineares e não lineares.

1.4.2 Objetivos específicos

- Determinar a combinação linear representativa das distribuidoras de energia elétrica, por meio da análise de componentes principais, de modo a englobar todos os setores de distribuição e consumo;
- Estimar modelos lineares, utilizando a metodologia Box & Jenkins para prever o nível da combinação linear em estudo;
- Verificar a presença de volatilidade e estimá-la, por meio de modelos não lineares, da classe geral Autorregressivos heteroscedásticos condicionais (ARCH), da

- combinação linear das variáveis quando necessário;
- Mostrar a aplicabilidade das previsões das combinações lineares, de modo a revelarem o comportamento das variáveis individuais.

1.5 Delimitação do tema

Esta pesquisa está delimitada geograficamente pelas companhias e usuários do Rio Grande do Sul e pela utilização de técnicas de modelos lineares da classe geral ARIMA e pelos modelos não lineares ARCH. Além destas delimitações citadas, deve-se ressaltar que os modelos de previsões aqui utilizados serão capazes de fornecer o movimento geral da curva, sem que seja possível especificar valores previstos individualmente, pois o que se procura é detectar os movimentos ascendentes e descendentes das curvas em estudo.

1.6 Estrutura do trabalho

A pesquisa foi estruturada em cinco capítulos. O primeiro traz a introdução, o tema da pesquisa, a justificativa e a importância do trabalho, o problema de pesquisa, o objetivo geral e os específicos, a delimitação do tema e a estrutura do trabalho.

No capítulo dois, encontra-se a Revisão de Literatura, em que se apresenta uma revisão de alguns conceitos sobre modelos lineares e não lineares, assim como a análise de componentes principais, pois esses serão os tópicos necessários para o desenvolvimento da pesquisa.

Na metodologia, no capítulo três, são apresentadas as etapas necessárias para a coleta, tabulação e modelagem das variáveis e dos dados em estudo, representados pelas componentes.

Já, no capítulo quatro, são demonstrados os resultados das análises realizadas neste estudo.

As considerações finais e as sugestões constam no capítulo cinco e, após esse, estão relacionadas às referências bibliográficas.

2 REFERENCIAL TEÓRICO

Descrevem-se, neste item, os assuntos necessários para a elaboração da pesquisa, de modo a atingir o objetivo geral e os específicos. Aborda-se a análise de componentes principais, modelos lineares, por meio dos modelos ARIMA, e os modelos não lineares por meio dos modelos ARCH.

2.1 Análise de componentes principais

O método de Análise de Componentes Principais (ACP) foi desenvolvido por Pearson (1901) e, posteriormente, por Hotelling (1933). Essa técnica vem sendo estudada por autores como Morrison (1976), Seber (1984), Reinsel (1993), Jackson (1980, 1981) e Johnson e Wichern (1992, 1998). A idéia central é a redução do conjunto de dados a ser analisado, principalmente quando os dados são constituídos de um grande número de variáveis inter-relacionadas.

Essa redução é feita transformando-se o conjunto de variáveis originais em um novo conjunto de variáveis que mantém, ao máximo, a variabilidade do conjunto de dados. As novas variáveis, denominadas de Componentes Principais (CP), são *independentes e não correlacionadas*, favorecendo a análise, principalmente quando muitas variáveis devem ser analisadas.

Partindo-se de um vetor $X = [(X_1 X_2 \dots X_p)]$, um vetor de variáveis aleatórias com vetor média μ e matriz de variância-covariância Σ , procura-se encontrar um novo conjunto de variáveis $Y_1 Y_2 \dots Y_p$, as quais são não correlacionadas entre si e cujas variâncias de cada nova variável é descrita em ordem decrescente de valores. Esta combinação linear é denominada de componente principal (REIS, 2001).

Cada componente principal será representada por $Y_j = l_{1j}X_1 + l_{2j}X_2 + \dots + l_{pj}X_{p1} = a_j X$, em que a_j corresponde ao vetor de constantes

que deve manter a condição de normalidade, isto é, $\sum_{j=1}^p a_{ij}^2 = 1$ para que o sistema

tenha solução única e de ortogonalidade, isto é, $\sum_{j=1}^p a_{ij}a_{kj} = 0$ para $i \neq k$ com $i, k =$

$1, \dots, p$, que garante que as componentes principais sejam independentes.

Para satisfazer as condições de normalidade e ortogonalidade, utiliza-se o método do multiplicador de Lagrange, resultando na definição da equação característica de Σ , em que:

$$|\Sigma - \lambda I| = 0 \quad (1)$$

Para que se encontre a solução desse sistema, é necessário que $|\Sigma - \lambda I|$ seja singular, isto é, quando igualada ao vetor nulo, deverá ter uma solução não nula para a_1 , se e somente se λ for um autovalor da matriz Σ .

Conforme Corrar, Paulo e Dias Filho (2007), tomando os elementos do vetor \vec{x}_i , assim determinados como os coeficientes de Y_i , tem-se que o i -ésimo componente principal:

$$i) \text{Vâr}(Y_i) = \hat{\Lambda}_i \text{ logo } \text{Vâr}(Y_1) > \text{Vâr}(Y_2) > \dots > \text{Vâr}(Y_p); \quad (2)$$

$$ii) \sum \text{Vâr}(X_i) = \sum \hat{\Lambda}_i = \sum \text{Vâr}(Y_i); \quad (3)$$

$$iii) \text{Côv}(Y_i, Y_j) = 0, \text{ desde que } \sum_{j=i}^p a_{ij} a_{kj} = 0. \quad (4)$$

A explicação de cada componente principal Y_i é medida em termos de variância. Logo, tem-se que o quociente é expresso em percentagem:

$$\frac{\hat{\Lambda}_i}{\sum_{i=1}^p \hat{\Lambda}_i} \cdot 100 = \frac{\hat{\Lambda}_i}{\sum_{i=1}^p \hat{\Lambda}_i} \cdot 100 = \frac{\hat{\Lambda}_i}{\text{traço}(\Sigma)} \cdot 100, \quad (5)$$

sendo que esta expressão representa a proporção da variância total explicada por cada componente.

No caso de se trabalhar com dados amostrais, a metodologia de análise continua a mesma, apenas se substitui a matriz de variância-covariância populacional Σ pela matriz de variância-covariância amostral S .

É possível encontrar tantas componentes principais quantas forem as variáveis envolvidas na análise. Mas, como um dos principais objetivos da análise é a redução de dimensionalidade, é necessário selecionar o número de componentes a serem consideradas no estudo.

Alexander (2005) diz que a análise de componentes principais trata de um método usual na extração das mais importantes fontes não autocorrelacionadas de variação de um sistema multivariado. Não somente as estruturas a termo, mas também as volatilidades implícitas das diversas opções de um mesmo ativo. Desta forma, a técnica torna-se útil para a aplicação do estudo proposto.

2.2 Modelos lineares e não lineares de previsão

Tradicionalmente, são utilizados os modelos lineares – autorregressivos integrados e de médias móveis (ARIMA), para eliminar a autocorrelação, os quais são adequados para modelar a média condicional de uma série, cuja variância dos erros é assumida constante ao longo do tempo.

Modelos ARIMA

Uma das técnicas usadas pela econometria contemporânea para a previsão do comportamento de variáveis se constitui no emprego de modelos univariados. A construção dos modelos de séries temporais univariados é fundamentada na teoria de que existe uma grande quantidade de informação presente na série de dados, sendo esses dados capazes de fornecer estimativas sobre o comportamento futuro da variável. Assim, será apenas o próprio comportamento da variável que responderá pela sua dinâmica futura. Este tipo de modelo é conhecido na literatura como o modelo autorregressivo integrado de médias móveis (ARIMA), ou definido simplesmente por modelo (ARIMA, BOX e JENKINS, 1970).

A construção do modelo ARIMA parte da concepção de que as séries temporais envolvidas na análise são geradas por um processo estocástico estacionário, cuja natureza pode ser representada a partir de um modelo matemático. Segundo Box, Jenkins e Reinsel (1994), um processo é estacionário quando ele oscila em torno de uma média constante, com variância também constante, conhecida por homocedástica, e é definido como estocástico por ser controlado por leis probabilísticas.

Segundo Morettin (2006), os modelos ARIMA são modelos responsáveis por capturar e modelar processos de memória curta, pois a FAC decresce rapidamente para zero. Caso o processo não tenha esta característica, outros modelos devem ser empregados.

Um processo $\{x_t, t \in Z\}$ segue um processo ARIMA (p,d,q) se $\Delta^d x_t$ seguir um modelo ARMA(p,q), ou seja,

$$\phi(B)\Delta^d x_t = \theta(B)e_t \quad (6)$$

onde B é operador retroativo, ou operador de defasagens, d é inteiro positivo e representa a diferença, conforme a ordem de integração; ϕ e θ são os parâmetros dos processos autorregressivos e de média móvel de ordem p e q , respectivamente, e $e_t \sim RB(0, \sigma^2)$.

A construção de modelos ARIMA segue um ciclo iterativo da metodologia de Box e Jenkins, que é composto por quatro etapas: a) identificação; b) estimação; c) verificação do diagnóstico e d) previsão (GUJARATI, 2000). Esses modelos também são discutidos pelos autores Box, Jenkins e Reinsel (1994), Makridakis, Wheelwright e Hyndman (1998), Morettin e Tolo (2004) e Morettin (2008).

A seguir, descrevem-se os modelos estacionários AR, MA e ARMA.

Modelo Autorregressivo de Ordem p (AR(p))

O nome autorregressivo é pelo fato de que Z_t , no instante t é função dos Z 's nos instantes anteriores a t , adicionado de um erro no instante t . O modelo que contém $\phi(B)=1$ é chamado modelo autorregressivo, sendo denotado por AR (p), onde " p " indica a ordem do modelo, isto é, o número de defasagens.

$$\tilde{Z}_t = \varphi_1 \tilde{Z}_{t-1} + \varphi_2 \tilde{Z}_{t-2} + \dots + \varphi_p \tilde{Z}_{t-p} + a_t \quad (7)$$

$$\varphi(B)\tilde{Z}_t = a_t \quad (8)$$

O modelo autoregressivo de ordem 1 e 2 pode ser representado pelas expressões (9) e (11), respectivamente

$$AR(1) \rightarrow \tilde{Z}_t = \varphi_1 \tilde{Z}_{t-1} + a_t \quad (9)$$

$$AR(2) \rightarrow \tilde{Z}_t = \varphi_1 \tilde{Z}_{t-1} + \varphi_2 \tilde{Z}_{t-2} + a_t \quad (10)$$

onde a notação \tilde{Z}_t corresponde a valores subtraídos da média do processo.

$$\tilde{Z}_t = Z_t - \mu \quad (11)$$

O modelo será considerado estacionário se $\phi(B)$ convergir para $|B| \leq 1$ como $\phi(B) = 0$, condição chamada de condição de estacionariedade. A condição de estacionariedade do AR(p) estabelece que todas as raízes da equação devem cair fora do círculo unitário e, sendo o processo de ordem finita, será sempre inversível.

Vê-se, portanto, que as autocovariâncias não dependem do t , e sim da ordem p . Como $|\phi| < 1$, pela condição de estacionariedade, quanto maior o valor de p , ou seja, quanto maior a distância entre as observações, menor a autocovariância.

A identificação do modelo será feita através das funções de autocorrelação e autocorrelação parcial, que indicam a ordem do modelo, conforme a Figura 1.

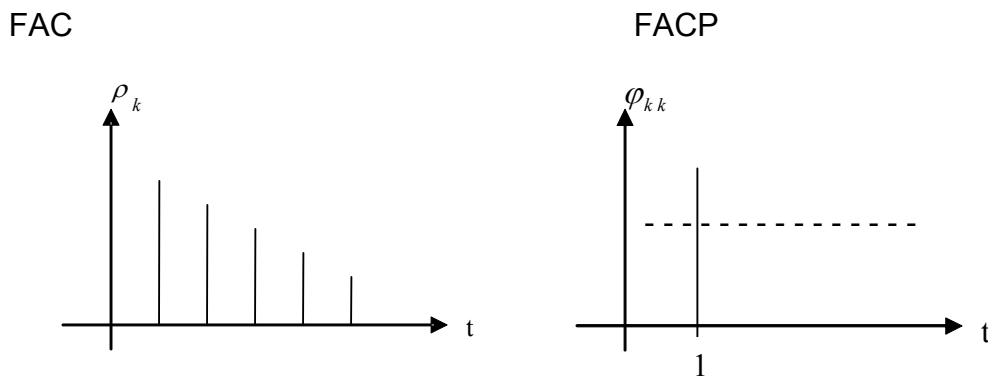


Figura 1 – Comportamento das funções de autocorrelação e autocorrelação parcial de um modelo AR (1).

A função de autocorrelação (FAC) decai exponencialmente, alternando ou não de sinal. Em geral é uma mistura de exponenciais e ondas senóides amortecidas e, a autocorrelação parcial (FACP) apresentará um corte rápido no *lag* significativo, indicando a ordem p do modelo.

Modelo de Médias Móveis de Ordem q (MA(q))

O nome do modelo vem do fato de que Z_t é uma função soma algébrica ponderada dos a_t que se movem no tempo. O nome, em si não é exato, pois Z_t não é média, visto que os parâmetros θ 's não somam, obrigatoriamente, a unidade.

Notação: MA (q) em que “ q ” indica a ordem do modelo, e θ é um parâmetro.

$$\tilde{Z}_t = a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}; \quad (12)$$

$$\tilde{Z}_t = \theta(B)a_t \quad (13)$$

A notação \tilde{Z}_t corresponde a valores do processo subtraídos de sua média. Pode-se observar os modelos MA(1) e MA(2) nas equações 14 e 15.

$$MA(1) \rightarrow \tilde{Z}_t = a_t + \theta_1 a_{t-1} \quad (14)$$

$$MA(2) \rightarrow \tilde{Z}_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} \quad (15)$$

O modelo é considerado estacionário se o processo for de ordem finita. Logo, no AR(1) e no AR(2), o número de defasagens é finito, portanto o modelo será sempre estacionário.

A inversibilidade é dada em função dos parâmetros, se $\pi(B)$ convergir para $|B| < 1$ com $\pi(B) = \theta^{-1} B$, as raízes da equação deverão cair fora do círculo unitário.

A identificação do modelo é feita através das funções de autocorrelações e autocorrelações parciais. A função de autocorrelação fornece a ordem do modelo, a qual deverá apresentar um corte rápido no *lag* significativo, indicando a ordem do modelo. A autocorrelação parcial decairá exponencialmente se todos os parâmetros forem positivos, caso contrário, formará uma senóide amortecida (SOUZA, 1991).

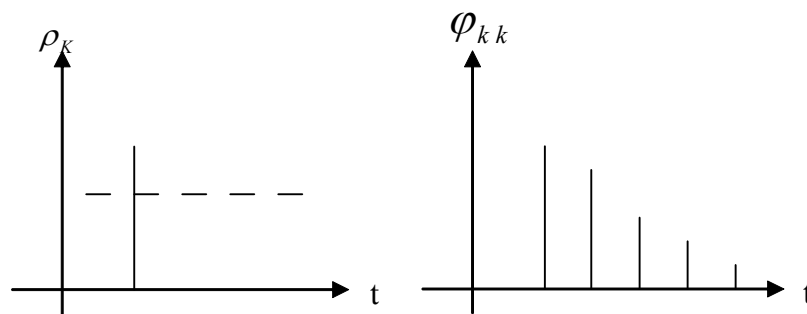


Figura 2 – Comportamento das funções de autocorrelação parcial e autocorrelação de um modelo MA (1).

Modelo Autorregressivo de médias móveis ARMA (p, q)

No modelo ARMA assim como nos modelos puros AR(p) e MA(q), ajustam-se somente as séries temporais estacionárias, sendo os modelos definidos pelas equações 16 e 17.

$$Z_t = c + \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (16)$$

ou

$$z_t - \phi_1 z_{t-1} - \dots - \phi_p z_{t-p} = c + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (17)$$

Há restrições específicas aos valores que possam assumir os parâmetros das equações (2.10) e (2.11). Para $p=1$, $-1 < \phi_1 < 1$. Para $p = 2$, $-1 < \phi_2 < 1$, $\phi_2 + \phi_1 < 1$ e $\phi_2 - \phi_1 < 1$. Para $p > 3$, condições mais elaboradas prevalecem. Similarmente, para $q=1$, $-1 < \theta_1 < 1$. Para $q = 2$, $-1 < \theta_2 < 1$, $\theta_2 + \theta_1 < 1$ e $\theta_2 - \theta_1 < 1$. Para $q > 3$, também prevalecem condições mais elaboradas (MAKRIDAKIS et al., 1998, p. 334).

Um dispositivo de notação útil é a utilização do operador de deslocamento retroativo B , conforme 18.

$$BX_t = X_{t-1}, \quad (18)$$

ou, em outras palavras, B operando sobre Z_t tem o efeito de deslocar os dados para trás em período. Assim,

$$B(BX_t) = B^2 X_t = X_{t-2} \quad (19)$$

e, eventualmente.

$$B^{12} X_t = X_{t-12} \quad (20)$$

Utilizando-se o operador B , a equação (2.14) pode ser escrita como segue:

$$(1 - \phi_1 B - \dots - \phi_p B^p) X_t = c + (1 - \theta_1 B - \dots - \theta_q B^q) a_t \quad (21)$$

Pois é da expressão (2.15) que o modelo em questão recebe sua denominação específica de ARMA (p , q), isto é, modelo autorregressivo de média móvel, de ordem p e q , conforme assinalado na expressão (21).

$$\underbrace{(1 - \phi_1 B - \dots - \phi_p B^p)}_{\text{AR}(p)} X_t = c + \underbrace{(1 - \theta_1 B - \dots - \theta_q B^q)}_{\text{MA}(q)} a_t \quad (22)$$

Os coeficientes $\phi_1, \phi_2, \dots, \phi_p$, e $\theta_1, \theta_2, \dots, \theta_q$ são estimados pelo ajuste da série temporal que está sendo modelada pelos métodos de mínimos quadrados ou da máxima verossimilhança.

A aplicação do método dos mínimos quadrados é semelhante à da regressão linear, entretanto, sem as formas simples desta aplicação, sendo as estimativas

feitas interativamente, por *softwares*, até que a soma dos erros quadráticos seja minimizada.

No segundo método, a verossimilhança de um conjunto de observações, denominada L, é proporcional à probabilidade de se obter as observações, dado o modelo, sendo, portanto, uma medida da plausibilidade do atual conjunto de observações, dado um particular conjunto de valores dos parâmetros. O método da máxima verossimilhança estima, interativamente, através de *softwares*, os valores dos parâmetros que maximizem a verossimilhança (MAKRIDAKIS et al., 1998).

O modelo ARMA, como qualquer modelo de previsão adequadamente ajustado a uma série temporal, deve ter seus erros, e_t , distribuídos aleatoriamente, isto é, comportando-se como uma série de ruído branco. Teoricamente, todos os coeficientes de autocorrelação de uma série de números aleatórios devem ser zero, pois eles não guardam qualquer relação entre si. Entretanto, a distribuição amostral dos coeficientes de autocorrelação de um conjunto de n amostras, retiradas de uma população de números aleatórios, pode ser aproximada por uma curva normal de média zero e desvio padrão de $1/\sqrt{n}$. Portanto, espera-se que 95% das n amostras de coeficientes de autocorrelação de uma série aleatória estejam entre $\pm 1,96/\sqrt{n}$. Se não for este o caso, a série provavelmente não é de ruído branco (MAKRIDAKIS et al., 1998).

Uma abordagem alternativa à verificação dos coeficientes de autocorrelação dos resíduos da aplicação do modelo ARMA é a utilização dos testes de Portmanteau, que considera o conjunto dos k primeiros coeficientes de autocorrelação, ao contrário de considerar cada um separadamente. É correto concluir que os resíduos não são ruído branco e a estatística Q, sugerida por Box-Pierce, ou Q^* , sugerida por Ljung-Box, se posicionam no extremo de 5% da cauda direita da distribuição do χ^2 . As expressões, para o cálculo das estatísticas Q, de Box-Pierce, ou Q^* , de Ljung-Box, são as seguintes:

$$Q = n \sum_{k=1}^h r_k^2 \quad (23)$$

e

$$Q^* = n + (n + 2) \sum_{k=1}^h \frac{r_k^2}{(n - k)} \quad (24)$$

Nas expressões (23) e (24), n é o número de observações da série, e h é a máxima defasagem dos coeficientes de correlação dos resíduos. Alguns autores aconselham que seja tomado $h = 20$ (MAKRIDAKIS et al., 1998), outros, $h = n/4$ (Box et al., 1994, p. 32). O valor de h deve ser utilizado como número de graus de liberdade da distribuição χ^2 .

Quando nem a função de autocorrelação nem a função de autocorrelação parcial caem bruscamente a zero, a recomendação de uso de um modelo ARIMA é reforçada.

Como vimos anteriormente, os processos $AR(p)$, $MA(q)$ e $ARMA(p,q)$ apresentam a função de autocorrelação FAC com características especiais. Assim:

- (i) um processo $AR(p)$ tem FAC que decai de acordo com exponenciais e/ou senóides amortecidas, infinita em extensão;
- (ii) um processo $MA(q)$ tem FAC finita, no sentido que ela apresenta um corte após o “lag” q ;
- (iii) um processo $ARMA(p, q)$ tem FAC infinita em extensão, a qual decai de acordo com exponenciais e/ou senóides amortecidas após o “lag” $q-p$.

Essas observações serão úteis no procedimento de identificação do modelo aos dados observados; calculando-se as estimativas das FAC que se acredita reproduzir, adequadamente, as verdadeiras FAC's desconhecidas, e comparando seu comportamento com o descrito acima. Para cada modelo, deve-se escolher um modelo que descreva o processo estocástico.

Modelos não estacionários ARIMA (p, d, q)

Como os processos encontrados na prática são raramente estacionários, deve-se encontrar algum tipo de operador que, por transformações matemáticas, transforme o conjunto de dados em estacionários.

Até agora, obteve-se, a partir de um processo de ruído branco a_t , um processo Z_t estacionário, porém com relação de dependência. Como os processos encontrados são raramente estacionários, deve-se encontrar algum tipo de operador que produza, a partir de Z_t , um processo estacionário.

Consideram-se, nesta modelagem, somente os processos chamados não estacionários homogêneos, isto é, aqueles que podem se tornar estacionários por meio de diferenças. Dada uma série Z_t , a primeira diferença de Z_t é definida por:

$$\Delta Z_t = Z_t - Z_{t-1} \quad (25)$$

e a segunda diferença é dada por:

$$\Delta^2 Z_t = \Delta(Z_t - Z_{t-1}); \quad (26)$$

$$\Delta^2 Z_t = Z_t - 2 Z_{t-1} + Z_{t-2}. \quad (27)$$

genericamente, a d -ésima diferença de Z_t é definida por:

$$\Delta^d Z_t = \Delta(\Delta^{d-1} Z_t). \quad (28)$$

Essas séries são não estacionárias homogêneas, mas, com um número finito de diferenças, consegue-se estacionarizá-las.

Dada Z_t , não estacionária, se $W_t = \Delta^d Z_t$, for estacionária pode-se representar W_t por um modelo ARMA (p, q) , isto é:

$$\phi(B)W_t = \theta(B)a_t. \quad (29)$$

Diz-se que Z_t segue um modelo autorregressivo integrado de médias móveis ARIMA (p, d, q) .

$$\phi(B)\Delta^d Z_t = \theta(B)a_t. \quad (30)$$

Como $\Delta Z_t = (1-B) Z_t$, vê-se que $\Delta = 1-B$ e (2.24) pode ser escrito como:

$$\phi(B)(1-B)^d Z_t = \theta(B)Z_t \quad (31)$$

Na maioria dos casos, é suficiente tomar $d = 1$ ou 2 diferenças para que $\Delta^d Z_t$ seja estacionária.

Outro fator que deve ser levado em consideração no momento da modelagem de séries temporais é a periodicidade que pode aparecer nos dados. Esta periodicidade que ocorre dentro do período de um ano é denominada de

sazonalidade e é representada por um período de repetição S . A periodicidade pode ser incorporada ao modelo por meio da utilização de um modelo ARIMA sazonal, como representado na equação 32.

$$\phi(B) \Phi(B^S) \Delta^d \Delta_S^D Z_t = \theta(B) \Theta(B^S) a_t \quad (32)$$

Onde, S é o período da sazonalidade, que pode ser trimestral, ($s=3$) semestral ($s=6$) e assim consecutivamente.

Caso seja detectada uma não estacionariedade do tamanho da sazonalidade S , pode-se realizar diferenças sazonais, de modo que a série se torne estacionária.

$$\Delta_S = 1 - B^S. \quad (33)$$

A equação 2.27 representa o operador diferença sazonal, D da equação (2.26) é o número de diferenças sazonais, $\phi(B^S) = 1 + \phi_1 B^S + \dots + \phi_P^{SP}$ é o operador autorregressivo sazonal de ordem P , e $\Theta(B^S) = 1 + \Theta_1 B^S + \dots + \Theta_Q B^{SQ}$ é o operador de médias móveis sazonal de ordem Q .

A equação (2.26) implica que devemos tomar D diferenças sazonais da série Z_t , onde d representa a ordem de diferenciação não estacionária de modo que a série

$$Y_t = \Delta^d \Delta_S^D Z_t = (1 - B)^d (1 - B^S)^D Z_t \quad (34)$$

seja estacionária, o que possibilitará a aplicação da metodologia Box & Jenkins. (VICINI e SOUZA, 2007).

O modelo (34) é denominado ARIMA sazonal ou SARIMA (p, d, q) (P, D, Q)_s.

Veja-se alguns casos particulares do modelo:

- (i) ARIMA (0, 1, 1): $\Delta Z_t = (1 - \theta B) a_t$;
- (ii) ARIMA (1, 1, 1): $(1 - \phi B) \Delta Z_t = (1 - \theta B) a_t$;
- (iii) ARIMA ($p, 0, 0$) = AR(p); ARIMA (0, 0, q) = MA(q);
- (iv) ARIMA ($p, 0, q$) = ARMA (p, q).

O item (i) é um caso importante, e é também chamado de modelo integrado

de médias móveis, IMA (1, 1).

$$Z_t = (Z_{t-1} + a_t - \theta B)a_t. \quad (35)$$

Pode-se demonstrar que esse modelo pode ser escrito na forma autorregressiva, conforme equação 36.

$$Z_t = \lambda Z_{t-1} + \lambda(1-\lambda)Z_{t-2} + \lambda(1-\lambda)^2 Z_{t-3} + \dots + a_t \quad (36)$$

onde Z_t seja dado em termos de seu passado, através de uma ponderação exponencial.

A seguir, apresentam-se as etapas para a efetivação da metodologia Box & Jenkins que será utilizada para a modelagem das séries em estudo.

A metodologia de Box-Jenkins refere-se ao método sistemático de identificação, ajuste, checagem e uso de modelos autorregressivos integrados à média móvel ou, simplesmente, modelos ARIMA.

Como uma série temporal tem dados coletados sequencialmente ao longo do tempo, espera-se que ela apresente correlação seriada no tempo. Os modelos Box-Jenkins, genericamente conhecidos por ARIMA (*Autoregressive Integrated Moving Average*), em português são denominados Autorregressivos Integrados de Médias Móveis. São modelos matemáticos que visam a captar o comportamento da correlação seriada ou autocorrelação entre os valores da serie temporal e, com base nesse comportamento, realizar previsões futuras.

A estratégia utilizada para a construção de modelos pela metodologia Box & Jenkins é baseada no ciclo iterativo, Figura 3, no qual a estrutura do modelo é formada pelos próprios elementos da série.

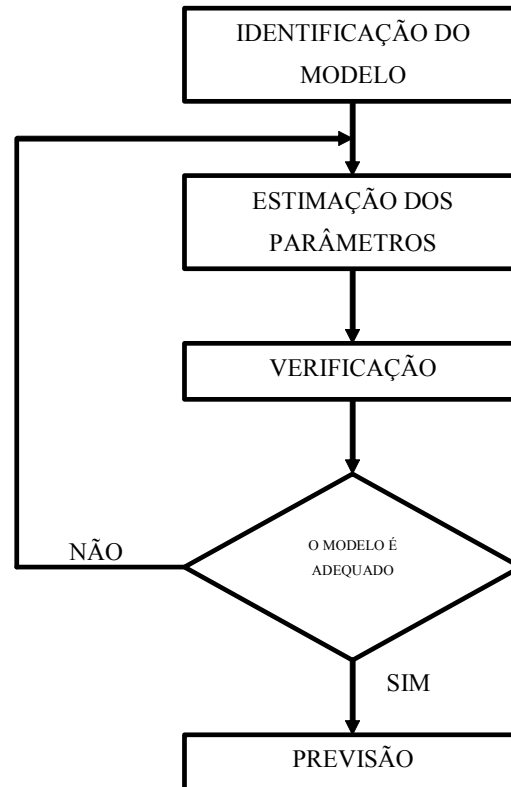


Figura 3 – Fluxograma do ciclo iterativo de Box & Jenkins.

O processo de identificação consiste em determinar quais dos filtros AR, MA, ARMA ou ARIMA, compõem a série. A realização do processo de identificação necessita, portanto, de outros instrumentos, que são a função de autocorrelação e a função de autocorrelação parcial.

Função de Autocorrelação (FAC) é o coeficiente de autocorrelação ou correlação serial de ordem k , ou seja, a autocorrelação entre y_t e y_{t-k} , é dada por:

$$\rho_k = \frac{\text{cov}(y_t, y_{t-k})}{V(y_t)} = \frac{\gamma_k}{\gamma_0}. \quad (37)$$

A sequência de pares (k, ρ_k) , $k = 1, 2, \dots$ é denominada função de autocorrelação. Valores negativos de k não são considerados explicitamente, tendo em vista que $\rho_k = \rho_{-k}$. O coeficiente de autocorrelação ρ_k envolve parâmetros geralmente desconhecidos. A análise estatística da FAC amostral requer o conhecimento da distribuição de r_k . Se $\rho_k = 0$ para $k \neq 0$, onde a variância de r_k é aproximadamente igual a $1/n$. Se n for grande, a distribuição de r_k é aproximadamente, normal:

$$k_k \rightarrow N(0;1/n). \quad (38)$$

Bartellet (1946) obteve o seguinte resultado, para a variância de r_k , quando $\rho_k = 0$ para $k > q$:

$$V(r_k) = \frac{1}{n}(1 + 2\rho_1^2 + \dots + 2\rho_q^2), k = q + 1, q + 2, \dots \quad (39)$$

Nesse caso, o cálculo efetivo da variância de $V(k)$ requer a substituição de ρ por r . Uma vez concluída a distribuição por r_k , pode-se construir os intervalos de confiança e realizar os testes de hipótese usuais, para verificar se cada coeficiente de autocorrelação (ρ_1, ρ_2, \dots) é nulo. Também é possível testar se os k , primeiros coeficientes de autocorrelação são conjuntamente iguais a zero. Nesse caso, utiliza-se o teste de Ljung-Box, cuja estatística é dada por:

$$Q^*(k) = n(n+2) \sum_{k=1}^k \frac{r_k^2}{n-k} \quad (40)$$

$Q(k)$ tem distribuição χ^2 , com k graus de liberdade. Se $Q(k) > \chi_k^2$, rejeita-se a hipótese de que os k primeiros coeficientes de autocorrelação são nulos.

Função de Autocorrelação Parcial (FACP), o coeficiente de autocorrelação parcial de ordem k , usualmente representado por ϕ_{kk} , mede a correlação entre y_t e y_{t-k} , depois que a influência de $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$, sobre Y_t for descontada. O coeficiente ϕ_{jj} , $j = 1, 2, \dots$ é dado pelo último coeficiente, β_{jj} , de cada uma das autorregressões a seguir:

$$\begin{aligned} y_t &= \beta_{11}y_{t-1} + \varepsilon_t, \rightarrow \phi_{11} = \beta_{11} \\ y_t &= \beta_{11}y_{t-1} + \beta_{22}y_{t-2} + \varepsilon_t, \rightarrow \phi_{22} = \beta_{22} \\ &\vdots \\ y_t &= \beta_{k1}y_{t-1} + \beta_{k2}y_{t-2} + \dots + \beta_{kk}y_{t-k} + \varepsilon_t, \rightarrow \phi_{kk} = \beta_{kk} \end{aligned} \quad (41)$$

Os valores de ϕ_{jj} , podem também, ser obtidos a partir da solução do sistema de equações de Yule-Walker para sucessivos valores de j :

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_k \rho_{k-1} \\
\rho_2 &= \phi_1 \rho_1 + \phi_2 + \dots + \phi_k \rho_{k-2} \rightarrow \phi_{jj} = \phi_j, j = 1, 2, \dots \\
&\vdots \\
\rho_k &= \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_k
\end{aligned}
\tag{42}$$

A sequência de pares (j, ϕ_{jj}) constitui a função de autocorrelação parcial. Estimativas de $\phi_{jj} - \hat{\phi}_{jj}$ podem ser obtidas aplicando-se mínimos quadrados ordinários às autorregressões, para estimar β_{jj} , ou substituindo por ρ_j por r_j , no sistema de equações de Yule-Walker. Segundo Quenouille (1949), se o processo for autorregressivo de ordem p , a variância de $\hat{\phi}_{jj}$ para $j > p$ é aproximadamente igual a $1/n$. Adicionalmente, para n moderado, $\hat{\phi}_{jj}$ tem-se distribuição normal. Portanto:

$$\hat{\phi}_{jj} = N(0; 1/n) \text{ para } j > p \tag{43}$$

Assim, os testes de hipótese e os intervalos de confiança usuais podem ser empregados para avaliar a significância da estatística de $\hat{\phi}_{jj}$.

A estimação dos parâmetros é a próxima etapa, consiste em determinar se a série estacionária, original ou transformada, é gerada por um processo autorregressivo ($p \neq 0$), de médias móveis ($q \neq 0$), e/ou mista. Neste caso é necessário determinar o valor de p e/ou q .

As características da FAC e da FACP, em cada caso, é que indicarão qual o possível processo gerador da série.

2.2.1 Modelos não lineares de previsão

A modelagem linear representada pelos modelos ARIMA será útil para estimar os modelos que representarão a série em estudo. Os resíduos oriundos da modelagem ARIMA serão investigados em relação à homocedasticidade de sua variância, caso os resíduos apresentarem característica de variância não constante, serão analisados por meio dos modelos não lineares ARCH.

Os Modelos Autorregressivos de Heterocedasticidade Condicional (ARCH) tiveram o seu surgimento com o pesquisador americano Robert Engle

(1982), que observou que os dois primeiros momentos de uma série temporal (média e variância) não são captados conjuntamente pelos modelos lineares. Geralmente, os modelos ARIMA, embora produzindo resíduos do tipo ruído branco, quando analisados na sua forma quadrática, não apresentam um comportamento não aleatório puro.

Os modelos lineares são responsáveis por modelar e prever a média condicional da série em estudo, assumindo que o resíduo a_t de tais modelos seja ruído branco, com média zero e variância constante. Os modelos propostos por Engle dão um passo a mais e verificam a existência de heterocedasticidade nos resíduos. Esta heterocedasticidade residual é denominada de volatilidade, o que a princípio poderia ser um problema a ser enfrentado na modelagem de séries temporais, pois não apresentaria propriedades desejáveis na modelagem e na estimação dos parâmetros, passa a ser uma ferramenta a mais para descobrir o comportamento de uma variável (SÁFADI e ANDRADE FILHO, 2007).

Como a variabilidade pode ser explicada pela volatilidade existente entre os erros oriundos do modelo de previsão linear, pode-se observar que a variância destes erros não é constante ao longo do tempo, mas varia de um período para outro, existindo uma autocorrelação na variância dos erros de previsão (GUJARATI, 2000).

A idéia básica desses modelos é a de que os componentes da série e_t não sejam correlacionados serialmente, mas são dependentes, e essa dependência pode ser descrita por uma função quadrática de seus valores passados. Desta forma, a série de resíduos e_t pode possuir média zero e não ser autocorrelacionada, mas a série dos seus valores quadráticos e_t^2 pode apresentar autocorrelação serial e, por isso, são denominados de modelos não lineares.

Geralmente séries temporais com presença de heterocedasticidade apresentam caudas mais pesadas, portanto, a curtose (K) é uma medida que pode auxiliar na identificação da presença de volatilidade. Esta medida é obtida pela razão entre o quarto momento de e_t ($E(e_t^4)$) e o quadrado de sua variância. Desenvolvendo algebricamente este quociente, tem-se que a curtose de um modelo ARCH (1), por exemplo, será dada por meio de (44).

$$K(e_t) = \frac{3(1-\alpha_1^2)}{(1-3\alpha_1^2)} > 3. \quad (44)$$

Portanto, ao admitir-se que a_t segue uma modelo ARCH, as caudas são mais pesadas do que a da distribuição normal que possui curtose igual a 3. Com isso, pode-se dizer que se há excesso de curtose, então há volatilidade.

O modelo ARCH, propriamente dito, assume que, em uma regressão, a variância do erro se correlaciona com a variável independente (erro da regressão). Desta forma, a variância é dita condicional (não constante) e seu valor depende da observação do próprio erro (ÉLIN DUXUS, 2009).

A idéia principal do modelo ARCH é o fato de que a variância de e_t , no período de tempo t , depende de e_{t-1}^2 . O termo de erro e_t , condicionado à informação disponível no período $(t-1)$ seria distribuído conforme a seguinte notação:

$$e_t \sim N[0, (\alpha_0 + \alpha_1 e_{t-1}^2)] \quad (45)$$

em que α_0 e α_1 são parâmetros explicativos da variância do termo de erro e_t (CAMPOS, 2007).

Um modelo ARCH(m), em que m denota a ordem do modelo, expressa a variância condicional do modelo anterior para a média condicional como uma função das inovações quadráticas passadas (SILVA, SÁFADI e CASTRO JÚNIOR, 2005).

Segundo Morettin (2008), o modelo ARCH(m) é definido por:

$$r_t = \sqrt{h_t} e_t \quad (46)$$

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \dots + \alpha_m r_{t-m}^2 \quad (47)$$

em que e_t é i.i.d com média zero, e h_t é a variância condicional.

Para assegurar que a variância condicional seja positiva e fracamente estacionária, as seguintes restrições paramétricas são necessárias: $\alpha_0 > 0$, $\alpha_t \geq 0$ para todo $t = 2 \dots m$ e $\sum \alpha_t < 1$.

Sob a condição de estacionariedade, a variância será dada por:

$$Var(r_t) = \frac{\alpha_0}{1 - \sum_{t=1}^m \alpha_t} \quad (48)$$

Pela própria definição, valores grandes de r_t são seguidos por outros valores grandes da série. Se admitirmos que r_t siga um modelo ARCH, as caudas serão mais pesadas do que as da distribuição Normal, o que é uma propriedade característica do modelo.

Na prática, seguindo o modelo original de Engle, assume-se que e_t tem distribuição Normal ou *t-Student* padronizada (SILVA, SÁFADI e CASTRO JÚNIOR, 2005).

Campos (2007) afirma que, se a assimetria é diferente de zero e a curtose é maior que três, isso significa que a série pode exibir sinais de volatilidade e de *clusters* de volatilidade. A assimetria retrata a forma de distribuição dos dados, e a curtose indica que a série apresenta uma distribuição leptocúrtica em relação à normal, isto é, os dados estão agrupados no centro juntamente com algumas observações nos extremos das caudas.

Apresentam-se, a seguir, as etapas pertencentes aos modelos ARCH. Maiores detalhes encontram-se em Morettin (2006), Bueno (2008), Enders (1995) e Hamilton (1994).

Desta forma, Engle (1982) provou que os modelos lineares são incompletos em função da incapacidade de explicar algumas características dessas séries. Na área das finanças a volatilidade serve como uma medida de velocidade de reação do mercado financeiro, pois mercados com baixas volatilidades são considerados mercados lentos, pois a tendência é se manterem estáveis. E mercados com alta velocidade são considerados rápidos, pois possuem grande variabilidade e suas ações e reações podem ser inesperadas. Os modelos ARCH são estimados pelo método de máxima verossimilhança e seguem as etapas de identificação do modelo, estimação dos parâmetros, verificação ou checagem e, por fim, faz a análise dos resíduos para validar o modelo não linear encontrado. Estas etapas são similares às etapas da metodologia de Box & Jenkins.

Neste sentido, o estudo do comportamento da volatilidade tem sido considerado uma peça fundamental, pois serve como medidor da vulnerabilidade dos mercados financeiros e da economia. Deseja-se analisar-se nas

empresas distribuidoras de energia elétrica para o Rio Grande do Sul o conhecimento individual da volatilidade. Se o comportamento da volatilidade dessas empresas for significativo, suas volatilidades devem ser estimadas, sendo possível, assim, analisar o seu efeito.

Desde o Prêmio de Nobel de Economia que Engle recebeu em 2003 pela sua contribuição com métodos estatísticos de modelagem e previsão da volatilidade de séries financeiras, os modelos ARCH tem sido amplamente estudados, ampliados e utilizados com diferentes propósitos. Issler (1999) e Bollerslev, Chou e Kroner (1992) discutem brevemente as vantagens que a modelagem ARCH oferece em relação às medidas que frequentemente são usadas por agentes do mercado.

Salienta-se o estudo aqui desenvolvido em um processo do número de consumidores e consumo de energia elétrica. Verificar se estes setores possuem características heterocedásticas, verificar a sua volatilidade ao longo do tempo seria de extrema importância para manter o processo de consumo e o número de consumidores sob controle. Antecipando os comportamentos dessas variáveis, evitam-se inúmeros problemas, tais como os blackouts, o excesso de geração de energia e/ou distribuição dentre e gargalos no setor o que podem levar a impactos negativos como custos elevados, más decisões de investimentos, precificação de títulos e gerenciamento de risco.

Testes de diagnósticos

No processo de modelagem de séries temporais, têm-se diversos pressupostos, já mencionados, que devem ser satisfeitos e, portanto, testados.

Na seção 4.4, apresentam-se estatística Ljung-Box para detectar autocorrelação e os testes de raiz unitária para analisar a estacionariedade; na seção 4.5, os testes de normalidade; os testes para detectar a heterocedasticidade condicional residual.

i) Testes de normalidade

O pressuposto da normalidade precisa ser atendido com o objetivo de oferecer confiabilidade aos testes estatísticos, uma vez que, em geral, eles se baseiam na distribuição normal do termo residual. Sua violação afeta as qualidades

de não tendenciosidade e variância mínima do método de mínimos quadrados ordinários.

Vários testes estatísticos têm sido usados para avaliar a normalidade dos resíduos do modelo estimado, dentre os quais se destacam:

- Jarque-Bera

É uma estatística que tem a vantagem de ser simples na sua operacionalização, em termos de compreensão intuitiva. Depende de dois parâmetros, assimetria e curtose, que podem ser calculados a partir de dados históricos.

A estatística Jarque-Bera possui um poder de teste elevado, não erra na presença de normalidade. Há uma baixa probabilidade de indicar a normalidade na presença de não normalidade das distribuições, ou seja, pode-se cometer o Erro Tipo II, aceitando-se a hipótese de normalidade quando a hipótese alternativa de não normalidade é a verdadeira.

É mensurada pela seguinte equação, a qual segue uma distribuição qui-quadrada com dois graus de liberdade:

$$JB = N \left[\frac{A_3^2}{6} + \frac{(K_4 - 3)^2}{24} \right] \sim \chi_2^2 (\alpha = 5\%) \geq 5,99 \quad (49)$$

onde: A_3 é coeficiente de assimetria, K_4 é o coeficiente de curtose excessivo e N é o tamanho da amostra.

As hipóteses a serem testadas são:

$H_0 =$ Assimetria = zero e curtose = a três (normalidade);

$H_1 =$ Assimetria \neq zero e curtose \neq três (não normalidade).

Regra de decisão: o teste baseia-se nos coeficientes de assimetria e curtose. Se p-valor for maior que o nível de significância de 5% cujo valor crítico é 5,99, a hipótese nula de normalidade não será violada.

ii) Testes para autocorrelação

A autocorrelação indica se existe uma dependência temporal entre os valores sucessivos dos resíduos, ou seja, os resíduos são correlacionados entre si. A presença de autocorrelação significa que as estimativas através do método MQO dos parâmetros não são tão eficientes, ou seja, não apresentam variância mínima. Ela provoca um viés nos erros-padrão, conduzindo a testes e intervalos de confiança incorretos.

A estatística Ljung-Box, em conjunto com os correlogramas FAC e FACP, é aplicada nas seguintes situações: inicialmente, nos dados originais para analisar, em primeira instância, a estacionariedade da série; em seguida, nos resíduos do modelo ARIMA com o propósito de verificar se os resíduos do modelo estimado formam um ruído branco; posteriormente, nos resíduos ao quadrado desse modelo para identificar se existe autocorrelação significativa. Se houver, é necessário aplicar testes para verificar se existe heterocedasticidade condicional na série quadrática. Se a heterocedasticidade for confirmada, a série é modelada através dos modelos ARCH.

- Estatística Ljung-Box

Segundo Ehlers (2007), ao invés de verificar as autocorrelações residuais individualmente, pode-se testar se um grupo de autocorrelações é significativamente diferente de zero por meio das chamadas estatísticas Q.

Segundo Morettin (2006), Box e Pierce (1970) sugeriram, um teste de hipótese para as autocorrelações dos resíduos estimados, apesar de não detectar quebras específicas no comportamento de ruído branco, pode indicar se esses valores são muito altos. Uma modificação desse teste foi proposta por Ljung e Box (1978), na qual a estatística de teste é dado pela expressão (50)

$$Q(K) = n(n+2) \sum_{k=1}^K \frac{\hat{r}_k^2}{(n-k)} \quad ((50))$$

em que n representa o número de observações, e k representa o número de parâmetros estimados, seguindo uma distribuição χ^2 com $K - p - q$ graus de liberdade.

As hipóteses testadas são:

$$H_0 = \text{Ruído branco (RB);}$$

$$H_1 = \text{Não é ruído branco (RB).}$$

Regra de decisão: se a estatística calculada excede o valor crítico da tabela de χ^2 para determinado nível de significância, podemos rejeitar a hipótese nula de que todos os (\hat{r}_k) são iguais a zero. Isso implica aceitar a hipótese alternativa de que pelo menos uma autocorrelação (\hat{r}_k) é estatisticamente diferente de zero. Nesse caso, o resíduo não é ruído branco. No caso da aceitação da hipótese alternativa, rejeita-se o modelo, pois há uma estrutura de correlação serial significativa nos resíduos (GUJARATI, 2000; SÁFADI, 2004).

Na prática, utilizam-se entre as 15 e 20 primeiras autocorrelações (\hat{r}_k) .

Embora, em amostras grandes, tanto a estatística Q como a Q(K) sigam a distribuição χ^2 , a estatística Q(K) de Ljung-Box possui as melhores propriedades em amostras pequenas do que a estatística Q de Box e Pierce.

Testes para heterocedasticidade condicional

Apresentam-se, a seguir, os testes para comprovação da presença de heterocedasticidade condicional na série residual do modelo estimado.

Até pouco tempo atrás, a heterocedasticidade era considerada um problema muito comum em dados *cross section*, embora também esteja presente em algumas séries temporais, mas estudos de Engle (1982) e por Bollerslev (1986) provaram que essa anomalia também pode ser modelada.

Na prática, é desejável que a variância dos resíduos do modelo estimado seja constante, ou seja, $Var(\mu_i) = \sigma^2$. Caso isso ocorra, diz-se que os resíduos do modelo estimado são homocedásticos. A violação desse pressuposto revela que a variância irá diferir ao longo do tempo, isto é, $Var(\mu_i) = \sigma_i^2$, e os resíduos serão heterocedásticos.

Na presença de heterocedasticidade, o método de mínimos quadrados ordinários não produz estimativas eficientes para seus parâmetros nem variância

mínima. A não minimização da variância conduz a erros-padrão viesados e à incorreção dos testes t-student e F-Snedecor, além dos intervalos de confiança. Isso justifica a necessidade de aplicar modelos heterocedásticos para modelar essa característica.

Existem diversos testes estatísticos para detectar a presença de heterocedasticidade nos resíduos ao quadrado do modelo estimado. Apresentam-se a seguir, os testes de White aplicado no caso multivariado, o ARCH-LM empregado nos casos univariados e, posteriormente, a estatística F.

- ARCH-LM

O teste ARCH-LM, do tipo multiplicador de Lagrange proposto por Engle (1982), testa a hipótese de que os erros resultantes da estimação do modelo são dependentes dos erros $r_{t-1}^2, r_{t-2}^2, \dots, r_{t-m}^2$, e baseia-se na estimação da regressão auxiliar (BEM, 1998).

Silva et al. (2005) afirma que, se os coeficientes de autocorrelações estimados revelam um padrão previsível para a média condicional da série analisada e se as autocorrelações estimadas para os resíduos ao quadrado também exibem sinais de previsibilidade, há uma evidência de efeitos ARCH.

O teste ARCH-LM utiliza a seguinte expressão:

$$r_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \dots + \alpha_m r_{t-m}^2 + u_t \quad ((51))$$

onde $t = m+1, \dots, T$. A estatística do teste $T \times R^2$, também chamada de Obs*R-squared, tem distribuição assintótica $\chi^2(m)$ sob H_0 . Nesse caso, R^2 é o quadrado do coeficiente de correlação múltipla da regressão (51).

Para comprovar a presença de efeitos ARCH, testam-se as seguintes hipóteses:

$$H_0 = \text{Não existem efeitos ARCH};$$

$$H_1 = \text{Existem efeitos ARCH}.$$

Regra de decisão: se p-valor é menor que o nível de significância adotado, no caso específico deste trabalho adotou-se um nível de 5%, rejeita-se a hipótese nula, portanto, existe heterocedasticidade condicional.

- Estatística F

De acordo com Morettin (2008), um teste assintoticamente equivalente, que pode ter propriedades melhores para amostras pequenas, consiste em utilizar a estatística

$$F = \frac{(SQR_0 - SQR_1) / m}{SQR_1 / (T - 2m - 1)} \sim F(m, T - 2m - 1) \quad ((52))$$

na qual $SQR_0 = \sum_{t=m+1}^T (r_t^2 - \bar{r})^2$ e $SQR_1 = \sum_{t=m+1}^T \hat{u}_t^2$, com \bar{r} a média amostral dos r_t^2 e \hat{u}_t os resíduos de MQO da regressão (52). Se o valor de F for significativo, diz-se que há heterocedasticidade condicional na série.

Assim, o teste ARCH-LM é aplicado nos resíduos do modelo estimado em dois momentos: primeiro, aplica-se o teste à série de resíduos ao quadrado do modelo ARIMA para média condicional. Se esta série possui heterocedasticidade condicional, necessita-se estimar um modelo por meio dos modelos ARCH. Em um segundo momento, aplica-se o teste na série dos resíduos padronizados do modelo ARCH para variância condicional e na série ao quadrado destes resíduos, em que se pretende verificar a não rejeição da hipótese nula nos dois casos, indicando que a série residual do modelo estimado é puramente aleatória, sem dependência na média e na variância.

- Testes de raiz unitários para analisar a estacionariedade

Os testes a serem utilizados para testar a estacionariedade das componentes em estudo são o Augmented Dickey-Fuller (1984) - (ADF), que possui como hipótese nula que a série é não estacionária e integrada de ordem d ($d > 0$). Na grande maioria dos casos de séries não estacionárias, basta uma diferenciação para tornar a série resultante estacionária, isto é, I(0).

Como uma forma de validação dos testes ADF, utiliza-se o teste Kwiatkowski-Phillips-Schmidt-Shin (1992) - (KPSS), em que as hipóteses apresentadas são inversas dos testes ADF, ou seja, H_0 postula que a série é I(0) contra a alternativa que a série seja I(1). Estes testes são amplamente discutidos na literatura de econometria como em Enders (1995) e Maddala (1992).

Como as hipóteses nulas dos testes ADF e KPSS são distintas, espera-se

que os dois testes sinalizem a mesma decisão em relação à estacionariedade da série. Mas Baillie et al. (1996) salienta que a utilização desses testes pode gerar situações de conflitos, chegando-se a uma inconclusão em relação à estacionariedade da séries.

3 METODOLOGIA

A análise refere-se ao número de consumidores e ao consumo de energia elétrica no estado do RS, realizado pelas três maiores distribuidoras de energia elétrica, nomeadamente a AES-Sul, CEEE e RGE. Não são considerados, neste estudo, as cooperativas de energia elétrica ou geração própria de energia, que são responsáveis por menos de 6% de geração e distribuição.

Os dados foram coletados no período de janeiro de 1998 a dezembro de 2009, com observações mensais e medidas em unidades mega watts hora (MWh). No setor de consumo e de número de consumidores, as observações são mensais.

Os setores analisados dessas distribuidoras são: industrial, comercial, rural, serviços públicos, residencial e outros. A coleta de dados foi realizada na Fundação de Economia e Estatística do Rio Grande do Sul, que reúne informações de natureza socioeconômica relativas ao estado e a seus municípios.

Será necessário realizar a modelagem linear aplicada às componentes selecionadas, utilizando a classe geral dos modelos ARIMA e verificando a adequação dos modelos concorrentes por meio da análise dos resíduos, assim como selecionando aqueles que melhor representarem a série em estudo por meio dos critérios de penalidades AIC (Critério de Informação *Akaike*).

Caso seja verificada a presença de volatilidade residual na variável em estudo, esses resíduos serão modelados utilizando a classe de modelos ARCH e novamente será investigada a adequação do modelo via análise dos resíduos.

Foram consideradas, como variáveis deste estudo, o volume de consumo (em MWh) e o número de consumidores nos setores e distribuidoras apresentados na Tabela 1. Tabela 1 – Distribuidoras e regiões de abrangência.

Companhias Estaduais de Distribuição de Energia Elétrica para o Rio Grande do Sul		
Distribuidoras	Setores	Área de Concessão
CEEE	Residencial Comercial	Regiões metropolitanas de Porto Alegre, litoral e campanha gaúcha.
AES-Sul	Industrial Rural	Regiões metropolitanas e centro-oeste do estado.
RGE	Serviços Públicos Outros	Região norte-nordeste do estado do Rio Grande do Sul.

Para melhor visualização da abrangência de cada distribuidora de energia elétrica para o Rio Grande do Sul, em estudo, observa-se a Figura I.

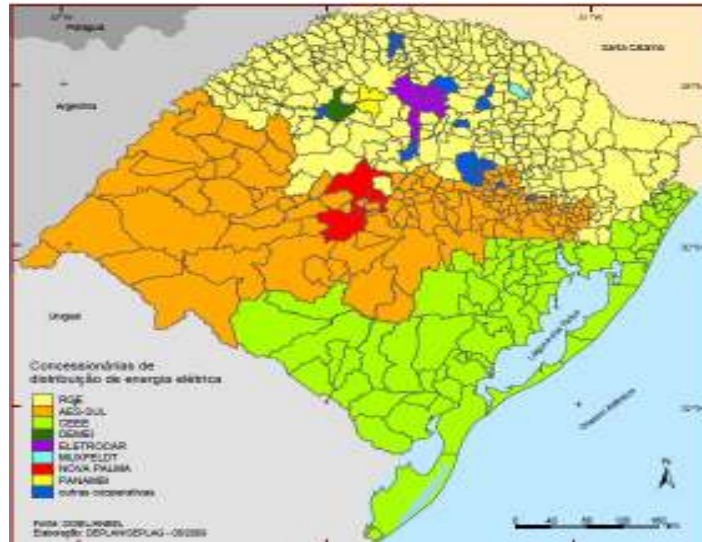


Figura I: Distribuidoras de energia elétrica para o Rio Grande do Sul.

Fonte: SIGEL/ANNEL

Elaboração: DEPLAN/SEPLAG – 08/2009.

As etapas metodológicas a serem seguidas são as seguintes:

Etapa 1: Análise de Componentes Principais – (ACP)

Extração das componentes principais e seleção das mesmas variáveis do número de consumidores e do fornecimento de energia, nos diversos setores e diversas companhias, de modo que sejam usadas apenas CP selecionadas pelo método de Kaiser (1960). Objetiva-se nesta etapa a ausência de correlação entre as componentes principais, o que proporciona uma melhor eficácia para se tratar os problemas de multicolinearidade, além de reduzir a dimensão estudada.

Etapa 2: Estatísticas descritivas

Inicialmente, um gráfico com a variável original é traçado para se ter uma ideia do seu comportamento. Também, verificam-se as estatísticas descritivas que revelem o comportamento da variável em estudo, tais como estacionariedade, medidas de assimetria, curtose e normalidade. Ainda, busca-se avaliar a presença de *outliers* e a possível identificação para suas causas. Segundo Bayer e Souza (2007), a estatística univariada tradicional considera as variáveis de forma

independente, não particularizando suas possíveis inter-relações, mas sua utilização é de suma importância no desenvolvimento de trabalhos científicos, pois ela é capaz de mostrar como as variáveis em análise estão se comportando.

Etapa 3: Modelagem linear – ARIMA (p, d, q)

Devido às características que determinado processo produtivo possui, já é possível verificar a presença de autocorrelação nos dados. Para facilitar a identificação de autocorrelação, recorre-se à *Função de Autocorrelação (FAC)* e à *Função de Autocorrelação Parcial (FACP)* para determinar o grau de autocorrelação na variável em estudo.

Logo, estimam-se modelos concorrentes da classe geral ARIMA, servindo-se da metodologia Box e Jenkins (1970), em que se procura um modelo que represente um *ruído branco*, isto é, que tenha média zero e variância constante.

O modelo selecionado será aquele que, além de explicar o fenômeno em estudo após ser realizada a análise dos resíduos, procurando-se sempre por modelos que forneçam ruído branco, isto é, com média zero e variância constante e não autocorrelacionados, apresente melhores estatísticas nos critérios de penalidades AIC (*apud* SOUZA e LOPES, 2009). O melhor resíduo selecionado terá os seus resíduos investigados em relação à homocedasticidade, por meio do teste LM-ARCH.

Etapa 4: Análise de heterocedasticidade residual

Na presença de heterocedasticidade, o método de mínimos quadrados ordinários não produz estimativas eficientes para seus parâmetros nem mesmo uma variância mínima. A não minimização da variância conduz a erros-padrão viesados e à incorreção dos testes *t-student* e F-Snedecor, além dos intervalos de confiança. Isso justifica a necessidade de aplicar modelos não lineares da classe ARCH em séries que apresentam heterocedasticidade (SILVA, 2000).

O teste ARCH-LM, do tipo multiplicador de Lagrange, proposto por Engle (1982), testa a hipótese de que os erros resultantes da estimação do modelo linear são dependentes dos erros $a_{t-1}^2, a_{t-2}^2, \dots, a_{t-m}^2$ e baseia-se na estimação da regressão auxiliar (BEM, 1998).

Silva, Sáfiadi e Castro Júnior (2005) afirmam que, se os coeficientes de

autocorrelações estimados revelam um padrão previsível para a média condicional da série analisada e se as autocorrelações estimadas para os resíduos quadráticos também exibem sinais de previsibilidade, há uma evidência de efeitos ARCH.

O teste ARCH-LM utiliza a seguinte expressão:

$$r_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \dots + \alpha_m r_{t-m}^2 + u_t \quad (53)$$

onde $t = m+1, \dots, T$. A estatística do teste $T \times R^2$ tem distribuição assintótica χ_m^2 sob H_0 . Nesse caso, R^2 é o quadrado do coeficiente de correlação múltipla da regressão (53).

Para comprovar a presença de efeitos ARCH, testam-se as seguintes hipóteses:

H_0 : A série não possui efeitos ARCH;

H_1 : A série possui efeitos ARCH.

Regra de decisão: se *p-valor* é menor que o nível de significância adotado, rejeita-se a hipótese nula, portanto, existe heterocedasticidade condicional.

Assim, o teste ARCH-LM foi aplicado nos resíduos do modelo estimado em dois momentos. No primeiro momento, aplica-se o teste à série de resíduos quadráticos do modelo ARIMA. Se esta série possui heterocedasticidade condicional, necessita-se estimar um modelo por meio dos modelos não lineares da classe ARCH e, em um segundo momento, aplica-se o teste na série de resíduos quadráticos do modelo ARCH e nos resíduos padronizados oriundos do modelo ARCH. Desta forma, procura-se a indicação de que a série residual do modelo estimado é puramente aleatória, sem dependência na média e na variância.

Etapa 5: Modelagem não linear – ARCH (m)

O objetivo da modelagem ARCH é considerar os efeitos da volatilidade conjuntamente com os da média do processo. No passo 2, já estabelecidos os resíduos que eram ruído branco e não autocorrelacionados, passa-se, no passo 3, a investigar a dependência quadrática destes resíduos e, se estes mostrarem-se heterocedásticos, parte-se para uma modelagem conjunta ARIMA-ARCH.

Nesta etapa, os passos metodológicos são semelhantes ao passo 4, apenas

com o diferencial de que a variável em estudo passa a ser a série de resíduos do modelo ARIMA selecionado.

Após estimado o modelo conjunto ARIMA-ARCH, os resíduos oriundos do melhor modelo serão investigados em relação a sua adequação, verificando se eles seguem um ruído branco não correlacionado e não possuem heterocedasticidade residual.

O fluxograma abaixo resume as etapas metodológicas descritas anteriormente.

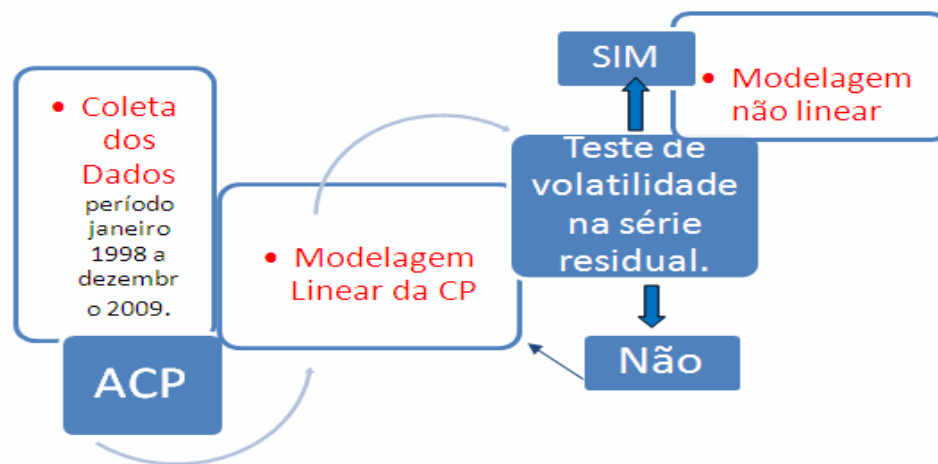


Figura II: Etapas metodológicas

Fonte: Elaboração própria.

Quantificar e qualificar as necessidades futuras de energia elétrica com qualidade e confiabilidade por meio de modelos lineares e não lineares aplicados a componentes principais é a proposta deste estudo, para que as empresas tenham mais um subsídio para as tomada de decisões e desenvolvam com eficiência a distribuição de energia para todos os setores.

Para o cumprimento de todos os passos metodológicos, faz-se necessário a utilização de *softwares* computacionais, a fim de realizar os testes, análises, estimação e modelagem dos dados. Para a análise de CP foi utilizado o *software Statistica versão 7.0* e, para as demais análises, o *EWVIES 7.0*.

4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Neste capítulo são apresentados os resultados da análise do consumo em (MWh) e do número de consumidores de energia elétrica para o Rio Grande do Sul por meio de modelos lineares e não lineares aplicados às componentes principais. Desta forma, os resultados obtidos contribuirão para revelar a aplicabilidade dos modelos de previsões nas combinações lineares, de modo a revelarem o comportamento global das variáveis, fornecendo condições para entender o comportamento conjunto das variáveis, de forma que as decisões a serem tomadas possam ser subsidiadas em bases científicas, obtendo o melhor planejamento, permitindo a alocação racional e otimizada.

Nas últimas duas décadas, tem sido observada uma tendência na reestruturação na maioria das áreas de negócio, desde o comércio de produtos até serviços públicos. O impacto da reestruturação tem sido profundo em todos os setores envolvidos e mudou significativamente as condições de competição. Nos serviços públicos, o processo iniciou no setor de telecomunicações e teve continuidade nos mercados de eletricidade e gás (DYNER e LARSEN, 2001).

O perfil do consumo de energia no País está mudando e deve demandar alterações na política de investimentos das empresas distribuidoras. A avaliação é do consultor Eduardo Bernini (http://www.correiadoestado.com.br/noticias/aumento-da-renda-mudou-perfil-de-consumo_1933/), ex-presidente da Eletropaulo. Segundo a estatal, o consumo médio por residência atingiu 165 kWh por mês, depois de quatro anos seguidos de crescimento. “Trata-se de um reflexo direto do aumento da renda e da mobilidade social verificada nos últimos anos”, comenta o especialista, que observa um novo padrão de consumo na população brasileira.

Segundo observou em entrevista recente o diretor-geral do Operador Nacional do Sistema Elétrico - ONS, Hermes Chipp (<http://www.asef.com.br/Noticias%2002%2003%202010.htm>) em vez de um único horário de pico, o sistema elétrico apresentou, no verão de 2009, um segundo pico de energia. Bernini, no entanto, demonstra preocupação com o impacto do aumento de consumo residencial sobre a qualidade da distribuição da energia nas grandes cidades.

Tendo em vista este aumento de consumo de energia, é imprescindível o estudo de projeção da demanda para se fazer desde o planejamento de geração, de transmissão, da distribuição e operação até o financeiro. Com isso, a empresa será capaz de projetar a oferta e a demanda em níveis de confiabilidade. Com o planejamento da transmissão, a empresa tem como estruturar o carregamento de linhas e transformadores, o sentido dos fluxos de intercâmbio e custos de distribuição, diminuindo os possíveis sinistros registrados como a sobrecarga, entre outros.

O aumento do PIB no estado do Rio Grande do Sul tem sido acompanhado pelo aumento dos níveis de consumo de energia elétrica nos últimos anos. Quanto ao perfil do consumo de energia elétrica, é o setor industrial que mais consome energia, com 37,72%. O setor residencial é responsável por 25,68%, o comercial por 16,84% e as áreas rurais por 12,54%. Com relação às cooperativas de eletrificação rural, apesar de 67% dos consumidores serem rurais, consomem somente 52,32% da energia elétrica distribuída, enquanto o setor industrial, com 0,64% dos consumidores, consome 23,23% da energia elétrica distribuída.

O planejamento da evolução temporal e distribuição espacial do número de consumidores em áreas urbanas e rurais vão ao encontro do atendimento com qualidade e confiabilidade, operacionalizando o despacho de unidades geradoras e das interligações, de modo a minimizar custos de transmissão e produção.

4.1 Análise de componentes principais

A ACP foi efetuada com o intuito de encontrar uma CP que fosse representativa do conjunto de dados, uma vez que a informação contida nas *p*-variáveis originais é substituída pela informação contida em *k* ($k < p$) componentes principais. A análise de componentes principais foi realizada utilizando-se o fornecimento mensal de energia por classe de consumo fornecida e o número de consumidores de energia elétrica por classe de consumo no Rio Grande do Sul em cada mês.

Estas são uma combinação linear de todas as variáveis originais, independentes entre si e escolhidas por ordem decrescente dos autovalores com o máximo de informação, em termos de variação total contida nos dados iniciais, permitindo, com isso, reduzir a dimensão do conjunto original de variáveis e permitir

construir os modelos de previsão.

A técnica de componentes principais permite reduzir o número de variáveis originais, correlacionadas em variáveis não correlacionadas em uma mesma quantidade de variáveis, transformada em componentes ortogonais, ou seja, nas componentes principais, que são uma combinação linear de todas as variáveis originais, independentes entre si e escolhidas por ordem decrescente dos autovalores com o máximo de informação, em termos de variação total contida nos dados iniciais, permitindo, com isso, reduzir a dimensão do conjunto original de variáveis e permitir construir os modelos de previsão.

A técnica de componentes principais apresenta tantas componentes principais quantos forem o número de variáveis que estão sendo analisadas, porém, com a propriedade de que, em ordem decrescente, as primeiras componentes principais retêm a maior parte da variância dos dados originais de seleção dos componentes (KAISER, 1960; KATTEL, 1960) e da variância média de 70%.

4.1.1 Análise do número de consumidores de energia elétrica por classe de consumo no Rio Grande do Sul para CEEE, AES-Sul e RGE

Neste item, faz-se a análise do número de consumidores de energia elétrica do RS. Além da análise descritiva, procede-se a análise de componentes principais com o intuito de reduzir a dimensionalidade do estudo e, no momento de realizar a modelagem estatística, pretende-se obter um comportamento global das variáveis de maior representatividade na componente analisada.

Na Figura 4, visualiza-se a seleção das componentes principais através do método gráfico *Screeplot*, sendo que a porcentagem de variação explicada pela componente está no eixo das ordenadas, e os autovalores estão representados em ordem decrescente no eixo das abscissas.

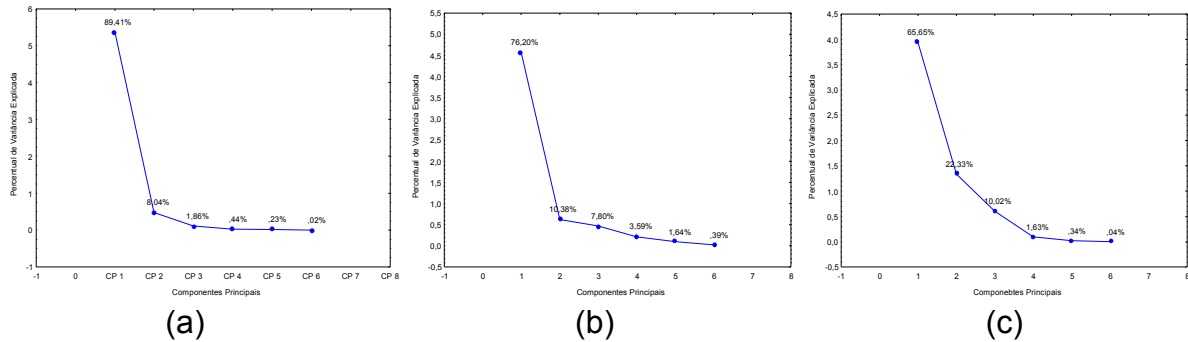


Figura 4 – Representação gráfica dos autovalores e percentual de variância explicada CEEE (a), AES-Sul (b) e RGE(c) respectivamente.

Para a distribuidora CEEE, a primeira componente explica 89,41% da variância total, sendo selecionada apenas a primeira componente. O segundo critério de seleção consiste em incluir somente aquelas componentes cujos valores próprios sejam superiores a 1. Este critério é sugerido por Kaiser (1960, *apud*, SOUZA e VICINI, 2005). Ele tende a incluir poucas componentes quando o número de variáveis originais é inferior a vinte e, em geral, utilizam-se componentes que conseguem sintetizar uma variância acumulada em torno de 70%. Segundo Lopes, Lima e Zanella (2007), o objetivo é encontrar um meio de condensar a informação contida em número de variáveis, em um conjunto menor de variáveis estatísticas.

Para a distribuidora AES-Sul, a primeira componente explica 76,20% da variância total, sendo considerada a primeira componente, pois esta componentes conseguem sintetizar uma variância acumulada superior a 70%.

Para a distribuidora RGE, as duas primeiras componentes explicam 87,98% da variância total, sendo consideradas a primeira e a segunda componente, obedecendo aos dois critérios de seleção, que consistem em incluir somente aquelas componentes cujos valores próprios sejam superiores a 1 e utilizar componentes que conseguem sintetizar uma variância acumulada em torno de 70%.

Na Tabela 2, estão demonstrados os fatores representados pelos autovalores resultantes da matriz de correlação, o percentual da variância explicado por cada fator e o percentual total de explicação da variância do conjunto de fatores.

Tabela 2 – Autovalores e variância explicada de cada componente para o número de consumidores de energia elétrica.

Extração das componentes principais					
Distribuidora	Número de Componente	Autovalores	% da variância explicada	Autovalores acumulados	% da variância explicada acumulada
CEEE	1	5,36	89,41	5,36	89,41
AES-Sul	1	4,57	76,20	4,57	76,20
RGE	1	3,94	65,65	3,94	65,65
	2	1,34	22,33	5,28	87,98

Por meio da ACP aplicada ao conjunto das variáveis, é selecionada para distribuidora CEEE uma CP ou nova variável, que represente 89,41% das variáveis originais.

A partir dessa análise, ao invés de utilizar as seis variáveis, passa a ser utilizada apenas uma CP, representada pelos seus autovalores superiores a 1, com uma perda de informação de aproximadamente 10,59% das variáveis originais.

Para a AES-Sul, constata-se que a primeira componente principal (CP_1) representa 76,20%. A partir desta análise, ao invés de utilizar as seis variáveis, passa a ser utilizada apenas uma CP, representada pelo seu autovalor superior a 1, com uma perda de informação de aproximadamente 23,8% das variáveis originais.

Para a distribuidora RGE, foram selecionadas as duas primeiras componentes principais, a (CP_1) que representa 65,65% e a (CP_2) que representa 22,33% da variabilidade das variáveis pesquisadas. Verifica-se, ainda, que os dois autovalores são superiores a uma unidade e representam uma variância explicada acumulada de 87,98%.

A partir desta análise, ao invés de utilizar as seis variáveis, passam a ser utilizados apenas dois fatores, representados pelos seus autovalores superiores a 1, com uma perda de informação de aproximadamente 12,02% das variáveis originais.

Na Tabela 3, são mostradas as influências que cada variável possui na composição de cada componente por meio das cargas fatoriais, refletindo quais componentes principais estão mais correlacionadas com as variáveis originais.

Assim, com base no poder de explicação dos autovalores, foi selecionada uma componente principal utilizada para determinação do modelo de previsão para a $CP1_{CEEE}$ e pode-se afirmar que é a mais importante para o estudo, pois ela é derivada do maior autovalor, explica 89,41% e representa os seis setores deste estudo.

A variável com maior correlação com a $CP1_{CEEE}$ é a variável nominada “Res” e representa o setor residencial, seguida das variáveis “Com” – setor comercial, Rural – setor rural, “S Publ” – setor serviços públicos, Outros – setor outros, ou seja, subdenominações das demais componentes em análise.

Para a AES-Sul, utiliza-se uma CP que representa 76,20% das variáveis explicadas. Pode-se afirmar que a $CP1_{AES-Sul}$ é derivada do maior autovalor e representa as seis variáveis em estudo. A variável com maior correlação com esta $CP1_{AES-Sul}$: Res representa o setor residencial, seguida das variáveis Com – setor comercial, Ind – setor industrial, Rural – setor rural, SPubl – setor serviços públicos, Outros – setor outros, ou seja, subdenominações das demais componentes em análise.

Tabela 3 – Cargas fatoriais por companhia.

Variáveis	CEEE					
	CP ₁	CP ₂	CP ₃	CP ₄	CP ₅	CP ₆
Res	-0,94	0,03	0,08	0,29	-0,14	-0,07
Com	-0,98	0,07	0,12	0,09	-0,02	0,13
Ind	0,76	-0,20	0,61	0,03	-0,02	0,00
Rural	-0,79	-0,59	-0,03	0,05	0,16	-0,02
S Publ	-0,84	0,44	0,25	-0,07	0,17	-0,04
Outros	-0,91	-0,16	0,10	-0,34	-0,14	-0,02
AES-Sul						
	CP ₁	CP ₂	CP ₃	CP ₄	CP ₅	CP ₆
Res	-0,44	0,03	0,11	0,62	-0,46	-0,44
Com	-0,46	0,08	0,18	0,19	-0,06	0,84
Ind	0,36	-0,25	0,89	0,07	-0,06	0,01
Rural	-0,37	-0,75	-0,04	0,10	0,53	-0,10
S Publ	-0,39	0,56	0,37	-0,16	0,54	-0,27
Outros	-0,43	-0,21	0,14	-0,73	-0,46	-0,11
RGE						

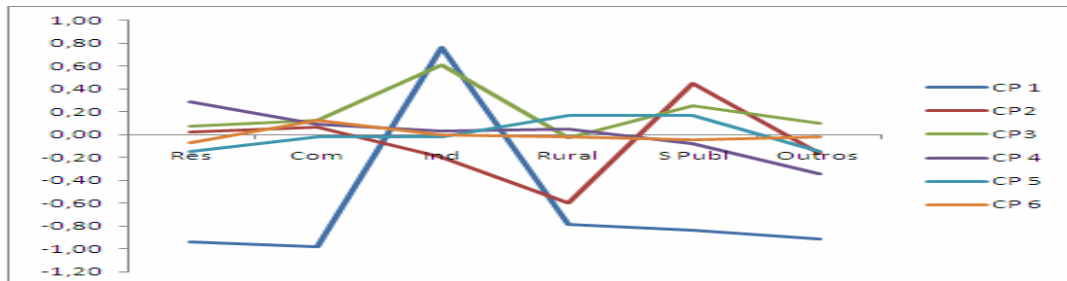
	CP ₁	CP ₂	CP ₃	CP ₄	CP ₅	CP ₆
Res	-0,97	-0,21	0,01	0,13	-0,05	0,03
Com	-0,98	-0,15	-0,01	0,08	-0,05	-0,03
Ind	-0,32	0,76	0,56	0,04	0,01	0,00
Rural	-0,96	-0,08	0,07	-0,27	-0,01	0,00
S Publ	-0,98	-0,01	-0,14	0,05	0,12	0,00
Outros	0,23	-0,83	0,51	0,01	0,03	0,00

Para a distribuidora RGE, foram selecionadas duas componentes principais utilizadas para determinação do modelo de previsão. A variável com maior correlação com esta CP_{1RGE}: Res representa o setor residencial, seguida das variáveis Com – setor comercial, Rural – setor rural, S Publ – setor serviços públicos, Outros – setor outros, ou seja, subdenominações das demais componentes em análise, e a CP_{2RGE} representa: Ind – setor industrial.

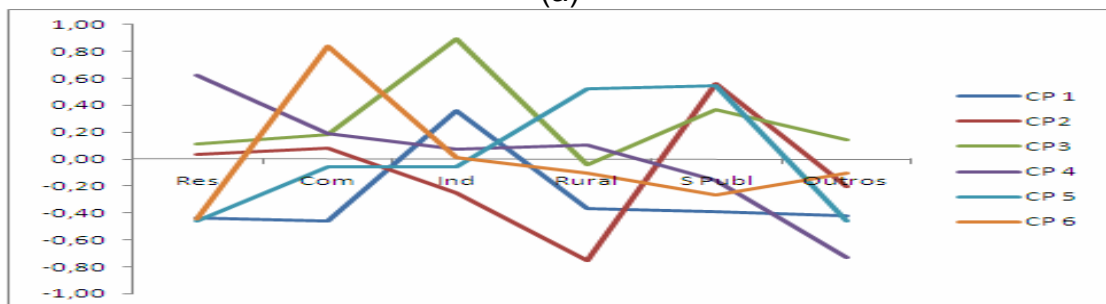
As correlações negativas de algumas variáveis em relação aos fatores mostram que as variáveis se relacionam de forma oposta, isto é, quando um setor aumenta, o outro diminui.

Os resultados da ACP são expostos na Tabela 3. Deste modo, as distribuidoras podem ser explicadas respectivamente: CEEE – por uma componente, com um poder de explicação de 89,41%; AES-Sul – por uma componente com poder de explicação de 76,20% e a distribuidora RGE - por duas componentes com poder de explicação acumulado de 87,98%. As cargas fatoriais são demonstradas na Tabela 3, em que se observa que as correlações são bastante elevadas, e isso se reflete nas semelhanças das cargas fatoriais da primeira componente principal para as distribuidoras CEEE e AES-Sul. Já, para a RGE, o ante exposto se divide em duas componentes. Segundo Alexander (2005), o deslocamento da primeira componente principal representa os movimentos paralelos de ascendência e descendência, portanto, uma mudança que ocorra em uma das variáveis que compõem a componente principal se refletirá nesta componente. Por essa razão, a primeira componente principal é chamada componente de tendência das curvas, neste caso, 89,41% da variação total da distribuidora CEEE (a), 76,20% da variação total da distribuidora AES-Sul (b) e 87,98% da variação total da distribuidora RGE (c).

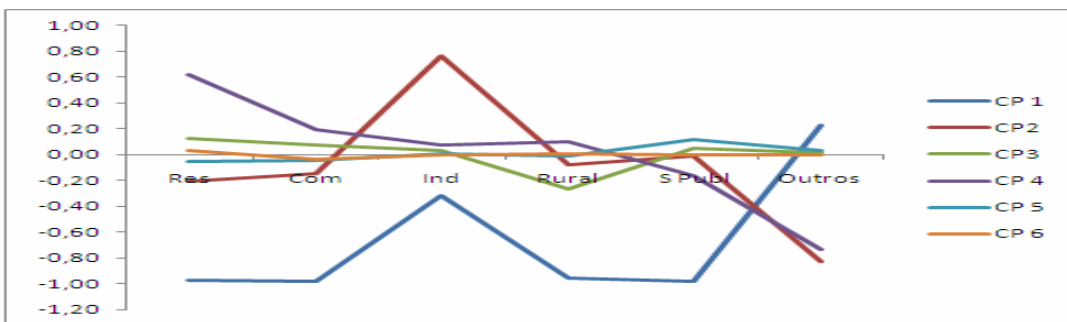
O deslocamento da componente principal provoca uma inclinação da curva das cargas fatoriais do número de consumidores de energia elétrica, conforme Figura 5.



(a)



(b)



(c)

Figura 5 – Gráfico das cargas fatoriais para as distribuidoras CEEE (a), AES-Sul (b) e RGE (c) para o número de consumidores de energia elétrica para o Rio Grande do Sul.

Pode-se perceber, analisando a Figura 5, que a ACP, além de reduzir as dimensões a serem analisadas, as componentes selecionadas são capazes de acompanhar os movimentos das mudanças de todas as curvas em análise, devido à elevada correlação que é refletida pelas cargas fatoriais da primeira componente para CEEE e AES-Sul e para a primeira e a segunda componentes para a RGE, que não somente são muito similares, mas também muito próximas de 1.

Na Figura 6, mostra-se o círculo de correlação unitário, com a nuvem de variáveis distribuídas no plano fatorial, cujas variáveis mais próximas ao círculo de

correlação são altamente representativas.

Na Figura 6a, observam-se as variáveis que estão mais correlacionadas e, assim, melhor explicam a CP_{1CEEE} e a CP_{2CEEE} . Na Figura 6b, observam-se as variáveis que estão mais correlacionadas e, assim, melhor explicam a $CP_{1AES-Sul}$ e a $CP_{2AES-Sul}$.

Na Figura 6c, observam-se as variáveis que estão mais correlacionadas e, assim, melhor explicam a CP_{1RGE} e a CP_{2RGE} .

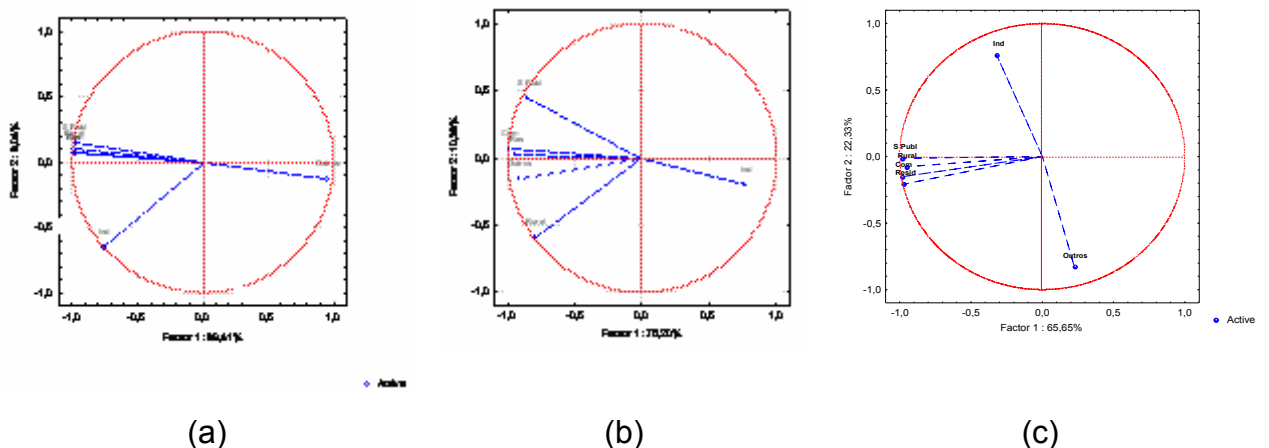


Figura 6 – Círculo de correlação unitário para as CEEE (a), AES-Sul (b) e RGE (c) respectivamente.

Pela inspeção do círculo unitário de correlação unitário, também é possível identificar as variáveis mais correlacionadas. Por exemplo, na Figura 6a, vêem-se que Serviços Públicos, Rural, Comercial e Residencial estão bem correlacionados ou apresentam o mesmo comportamento, pois o ângulo formado entre as variáveis é muito pequeno. A variável Industrial apresenta um ângulo um pouco maior com as variáveis que estão no segundo e terceiro quadrantes, mas este ângulo é menor do que 45° , mostrando que a correlação com as demais variáveis é moderada. A variável Outros apresenta-se não correlacionado, pois o ângulo formado é de aproximadamente de 180° com as variáveis Serviços Público, Rural, Comercial e Residencial. Industrial e as variáveis que incluem a variável Outros também apresentam fraca correlação entre si, pois o ângulo formado é aproximadamente 120° . É de salientar que o comportamento individual das variáveis na sua representação do sistema é muito bom, pois quase todas as variáveis estão posicionadas em cima do círculo unitário de correlação.

4.1.2 Análise do consumo em MWh de energia elétrica por classe no Rio Grande do Sul para CEEE, AES-Sul e RGE

Na Figura 7, visualiza-se a seleção das componentes principais por meio do método gráfico *Screeplot*, sendo que a porcentagem de variação explicada pela componente está no eixo das ordenadas, e os autovalores estão representados em ordem decrescente no eixo das abscissas.

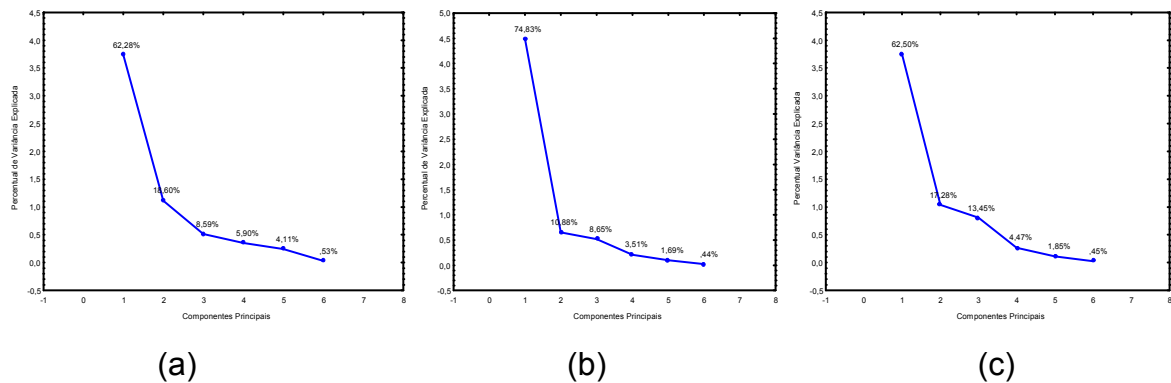


Figura 7 – Representação gráfica dos autovalores e percentual de variância explicada CEEE, AES-Sul e RGE respectivamente.

Para a distribuidora CEEE (a), as duas primeiras componentes explicam 81% da variância total, sendo consideradas as primeiras componentes, segundo critério de seleção, que consiste em incluir somente aquelas componentes cujos valores próprios sejam superiores a 1.

Para a distribuidora AES-Sul (b), a primeira componente explica 74,83% da variância total, sendo considerada a primeira componente e, para a distribuidora RGE (c), as duas primeiras componentes explicam 79,78% da variância total, sendo considerada a primeira e a segunda componentes.

Na Tabela 4, estão demonstradas as CP representadas pelos autovalores resultantes da matriz de correlação do consumo mensal de energia em MWh, o percentual da variância explicado por cada fator e o percentual total de explicação da variância do conjunto de fatores.

Tabela 4 – Autovalores e variância explicada de cada componente para o fornecimento mensal de energia por classe de consumo.

Autovalores Extração das componentes principais					
Distribuidora	Número de Componente	Autovalores	% da variância explicada	Autovalores acumulados	% da variância explicada acumulada
CEEE	1	3,74	62,30	3,74	62,30
	2	1,12	18,60	4,85	80,90
AES-Sul	1	4,49	74,83	4,49	74,83
RGE	1	3,75	62,50	3,75	62,50
	2	1,04	17,28	4,79	79,78

Por meio da ACP, foram selecionadas, para a distribuidora CEEE, duas CP, que representam 80,9% das variáveis originais. Constata-se que a primeira componente principal CP_{1CEEE} representa 62,30% das variáveis pesquisadas, e a CP_{2CEEE} representa 18,60% das variáveis. Verifica-se, ainda, que dois autovalores acima de 1 representam uma variância explicada acumulada de 80,9%.

Para a distribuidora AES-Sul, constata-se que a primeira componente principal $CP_{1AES-Sul}$ representa 74,83% da variabilidade das variáveis pesquisadas. A partir dessa análise, ao invés de utilizar as seis variáveis, passa a ser utilizado apenas um fator, representado pelos seus autovalores superiores a 1, com uma perda de informação de aproximadamente 25,17% das variáveis originais.

Para a distribuidora RGE, consideram-se as duas primeiras componentes principais CP_{1RGE} , que representa 65,50%, e a CP_{2RGE} , que representa 17,28% da variabilidade das variáveis pesquisadas. Verifica-se, ainda, que os dois autovalores acima de 1 representam uma variância explicada acumulada de 79,78%.

Na Tabela 5, estão demonstradas, por ordem de importância, as correlações existentes entre os fatores e as variáveis originais, refletindo quais são as componentes principais (os fatores) que estão mais ou menos correlacionadas com as variáveis (os setores).

Com base no poder de explicação dos autovalores, foram selecionadas as duas componentes principais utilizadas para determinação do modelo de previsão.

Nesta análise, utilizaram-se as duas CP, que representam 80,9% das

variáveis explicadas. Pode-se afirmar que a CP_{1CEEE} é a mais importante para o estudo, pois ela é derivada do maior autovalor, que explica 62,3%, e representa o maior número de variáveis (cinco das seis variáveis). A variável com maior correlação com esta CP_{1CEEE} representa: Res - setor residencial, seguida das variáveis Com – setor comercial, Rural – setor rural, S Publ – setor serviços públicos, Outros – setor outros, ou seja, subdenominações das demais componentes em análise.

Portanto, com base no poder de explicação dos autovalores, foram selecionadas duas componentes principais utilizadas para determinação do modelo de previsão.

Para AES-Sul, foi selecionada uma $CP_{1AES-Sul}$, que representa 74,83% das variáveis explicadas.

Tabela 5 – Cargas fatoriais constantes nas variáveis residencial, comercial, industrial, rural, serviços públicos e outros para as três distribuidoras em estudo.

Variáveis	CEEE					
	CP ₁	CP ₂	CP ₃	CP ₄	CP ₅	CP ₆
Res	-0,86	0,08	-0,29	0,22	0,35	0,01
Com	-0,92	0,31	-0,01	-0,08	-0,17	0,13
Ind	0,44	0,82	0,29	-0,11	0,20	0,00
Rural	-0,84	-0,20	0,03	-0,48	0,11	-0,04
S Publ	-0,84	0,47	-0,07	0,11	-0,21	-0,11
Outros	-0,73	-0,29	0,58	0,20	0,04	0,00
AES-Sul						
	CP ₁	CP ₂	CP ₃	CP ₄	CP ₅	CP ₆
Res	-0,94	0,04	0,05	0,29	-0,14	-0,07
Com	-0,98	0,09	0,10	0,09	-0,02	0,14
Ind	0,74	-0,06	0,67	0,05	-0,03	0,00
Rural	-0,77	-0,61	0,08	0,05	0,16	-0,02
S Publ	-0,82	0,50	0,19	-0,06	0,18	-0,04
Outros	-0,91	-0,13	0,15	-0,33	-0,15	-0,02

RGE						
	CP ₁	CP ₂	CP ₃	CP ₄	CP ₅	CP ₆
Res	0,24	0,26	-0,20	-1,11	-1,17	-2,77
Com	0,26	0,09	-0,16	0,11	-0,85	4,87
Ind	0,14	-0,78	0,26	0,39	-1,20	-1,18
Rural	0,23	0,27	-0,26	1,41	0,44	-2,14
S Publ	0,24	-0,32	-0,06	-0,60	2,29	0,19
Outros	0,12	0,32	1,02	0,02	0,18	-0,01

A RGE, nesta análise, faz uso de duas CP, que representam 79,78% das variáveis explicadas. A variável com maior correlação com esta CP_{1RGE} Res representa o setor residencial, seguida das variáveis Com – setor comercial, Rural – setor rural, S Publ – setor serviços públicos, Outros – setor outros, ou seja, subdenominações das demais componentes em análise, e a CP_{2RGE} representa Ind – setor industrial.

Os resultados da ACP são expostos na Tabela 4. Deste modo, as distribuidoras podem ser explicadas respectivamente: CEEE – por duas componentes com um poder de explicação acumulado de 80,30%; AES-Sul – por uma componente com poder de explicação de 74,83% e a distribuidora RGE por duas componentes com poder de explicação acumulado de 79,78%. As cargas fatoriais são demonstradas na Tabela 5, em que se observa que as correlações são bastante elevadas, e isso se reflete nas semelhanças das cargas fatoriais da primeira e segunda componente principal, para as distribuidoras CEEE e RGE. Já a AES-Sul é representada por uma componente principal. A primeira componente principal é chamada componente de tendência das curvas, neste caso, 80,9% da variação total da distribuidora CEEE (a), 74,83% da variação total da distribuidora AES-Sul (b) e 79,78% da variação total da distribuidora RGE (c).

O deslocamento da componente principal provoca uma inclinação da curva das cargas fatoriais do número de consumidores de energia elétrica, conforme Figura 8.

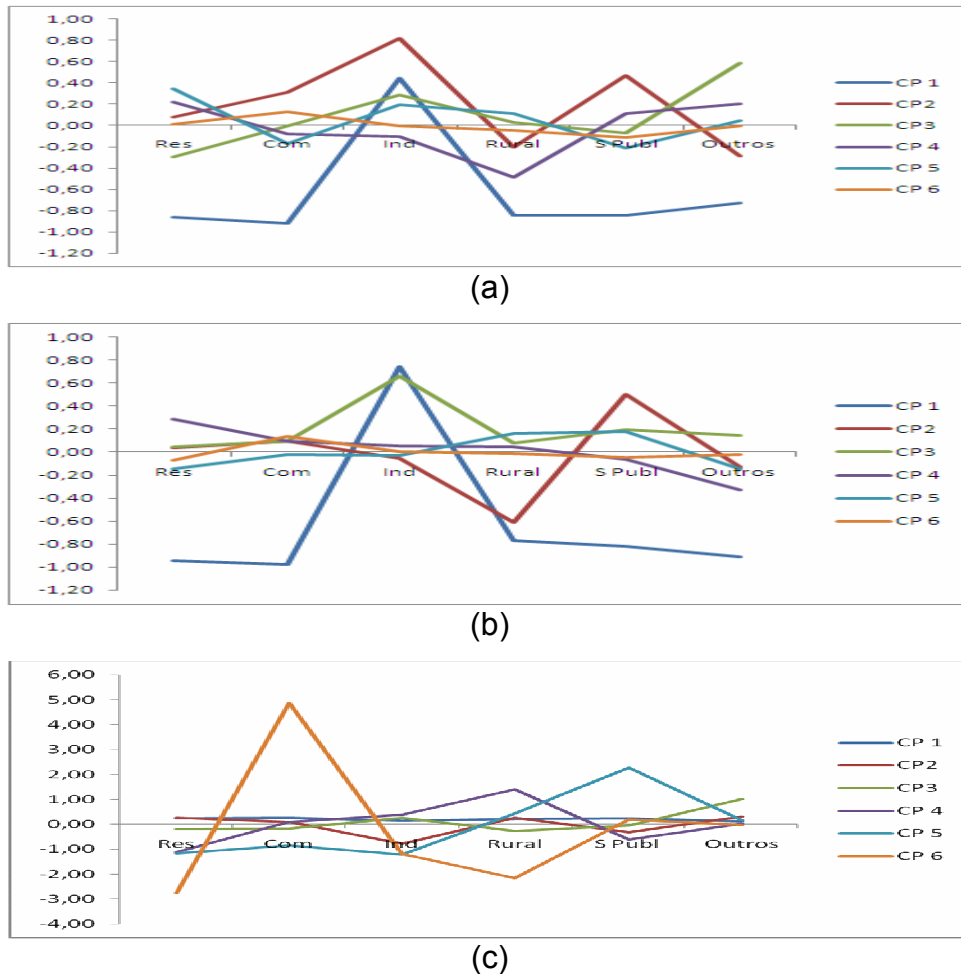


Figura 8 – Gráfico das cargas fatoriais para as distribuidoras CEEE (a), AES-Sul (b) e RGE (c) para o número de consumidores de energia elétrica para o Rio Grande do Sul.

Pode-se perceber, analisando a Figura 8, que a ACP foi capaz de reduzir as dimensões e que as componentes selecionadas acompanham o movimento das mudanças de todas as curvas em análise para a CEEE e para a AES-Sul, sendo a primeira e a segunda componente para a RGE muito similares.

Na Figura 9, mostra-se o círculo de correlação unitário, com a nuvem de variáveis distribuídas no plano fatorial, para as três distribuidoras de energia elétrica para o Rio Grande do Sul, cujas variáveis mais próximas ao círculo de correlação são altamente representativas.

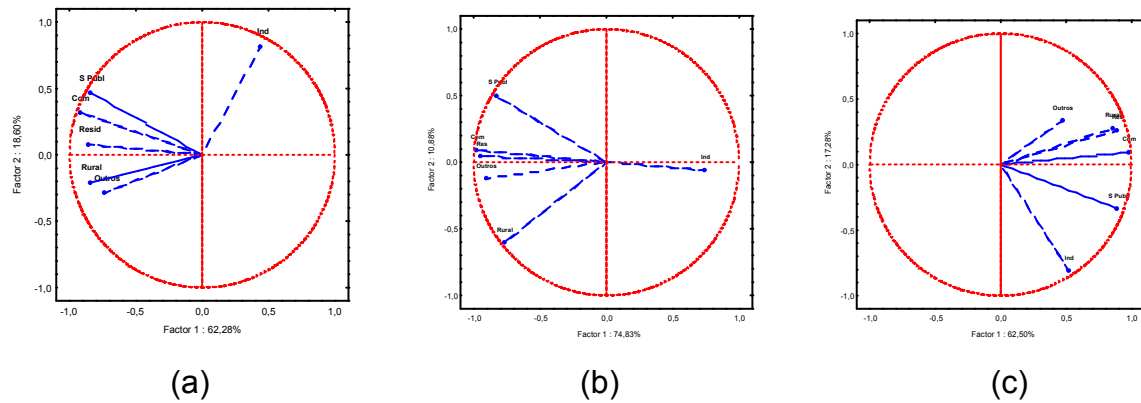


Figura 9 – Círculo de correlação unitário para as CEEE (a), AES-Sul (b) e RGE (c) respectivamente.

Na Figura 9, observam-se as variáveis que estão mais correlacionadas e assim melhor explicam a CP_{1CEEE} e a CP_{2CEEE} .

As variáveis S. Publ, Com, Resid, Rural, Outros (subdenominações das demais componentes em análise), estão bem representadas neste plano, pois se encontram próximas à circunferência e estão fortemente correlacionadas entre si, e o ângulo formado entre as variáveis é muito pequeno. Já, a variável Industrial apresenta um ângulo um pouco maior, mas menor que 45° , com as variáveis que estão no segundo e terceiro quadrantes. Pode-se observar que o comportamento individual das variáveis na sua representação do sistema é muito bom, pois quase todas as variáveis estão em cima do círculo unitário. Analogamente analisam-se os demais círculos de correlação unitária.

4.2 Estatísticas descritivas das componentes principais

Neste item serão apresentadas as estatísticas descritivas para as três companhias analisadas, como um suporte para a realização dos modelos a serem usados, com o intuito de conhecer o comportamento das variáveis no decorrer do período analisado e verificar os indícios para a presença de heterocedasticidade condicional na série. Na Tabela 6, apresentam-se algumas estatísticas descritivas das componentes principais que explicam as três distribuidoras de energia elétrica, cuja unidade de medida é a porcentagem.

Tabela 6 – Estatísticas descritivas das componentes principais para o número de consumidores e consumo de energia elétrica para as três distribuidoras.

Número de Consumidores de energia elétrica para RS					
Estatísticas	CP₁CEEE	CP₁AES-Sul	CP₁RGE	CP₂RGE	
Média	-7×10^{-15}	$-2,9 \times 10^{-15}$	$-7,3 \times 10^{-15}$	$-3,3 \times 10^{-16}$	
Assimetria	0,20	-0,43	0,52	-0,89	
Curtose	-0,38	-0,43	-0,70	-0,11	
Desvio-padrão incondicional	0,16	2,14	1,98	1,16	
Coeficiente de Variação (%)	$-2,30 \times 10^{15}$	$-7,37 \times 10^{15}$	-	$2,72 \times 10^{16}$	$-3,51 \times 10^{16}$

Consumo de energia elétrica em MWh, para RS					
Estatísticas	CP₁CEEE	CP₂CEEE	CP₁AES-Sul	CP₁RGE	CP₂RGE
Média	$1,5 \times 10^{-16}$	$-3,65 \times 10^{-15}$	$1,68 \times 10^{-15}$	$-6,75 \times 10^{-15}$	$4,04 \times 10^{-15}$
Assimetria	-0,56	-0,38	-0,38	0,20	0,79
Curtose	2,49	3,65	2,45	2,59	4,89
Desvio-padrão incondicional	1,93	1,06	2,12	1,94	1,02
Coeficiente de Variação(%)	$1,33 \times 10^{-16}$	$-2,89 \times 10^{-14}$	$1,26 \times 10^{-15}$	$-2,87 \times 10^{-14}$	$2,52 \times 10^{-14}$

Com a análise descritiva das variáveis para o número de consumidores, verificou-se, pelo coeficiente de variação, que a média da variável é considerada pouco representativa, visto que esse valor expresso em percentual está muito próximo de 50%.

Ao analisar o valor da assimetria e da curtose, observou-se que a assimetria é diferente de zero e que a curtose variou entre -0,11 e 0,52, o que é um indicativo de que algumas das componentes principais podem exibir sinais de volatilidade e de *clusters* de volatilidade, conforme Campos (2007). De acordo com Soares e Castelar (2003), se a curtose excede três unidades, a distribuição é dita leptocúrtica.

Analisando o consumo em MWh em relação à análise descritiva das variáveis, verificou-se, pelo coeficiente de variação, que a média da variável é considerada pouco representativa, visto que esse valor expresso em percentual está muito próximo de 50%.

Ao analisar o valor da assimetria e da curtose, observou-se o seguinte: como a assimetria é diferente de zero e a curtose variou entre 2,49 e 4,89, é um indicativo de que algumas das componentes principais podem exibir sinais de volatilidade.

Inicialmente, analisou-se o comportamento da variável em estudo, investigando-se em relação às medidas descritivas. A próxima etapa é estimar modelos lineares, utilizando a metodologia Box & Jenkins baseado na técnica de análise de componentes principais, verificando a presença de volatilidade e estimando-a por meio de modelos não lineares da classe geral ARCH quando necessário.

Alexander (2005) diz que ACP é um recurso muito útil, porque permite uma grande redução das dimensões e a ordenação natural do sistema e ainda fornece uma interpretação significativa das componentes principais. Além disso, esta análise facilita, em grande parte, a avaliação do movimento da volatilidade, vinculada a oscilações de demanda e ao fornecimento de energia elétrica.

Os resultados a seguir explicam as três companhias em estudo, levando em consideração todos seus setores de fornecimento de energia para o número de consumidores e para o fornecimento de energia em MWh. Nestes dois casos, a ACP possibilitou que as dimensões fossem substancialmente reduzidas, capturando todas as possíveis combinações de deslocamento das séries em estudo.

As Figuras 10 e 11 mostram o comportamento dos fatores que representam o número de consumidores para as três companhias distribuidoras de energia elétrica, CEEE, AES-Sul e RGE, para o Rio Grande do Sul. Avaliando estas figuras, verifica-se que as séries não apresentam um comportamento estacionário em relação à média.

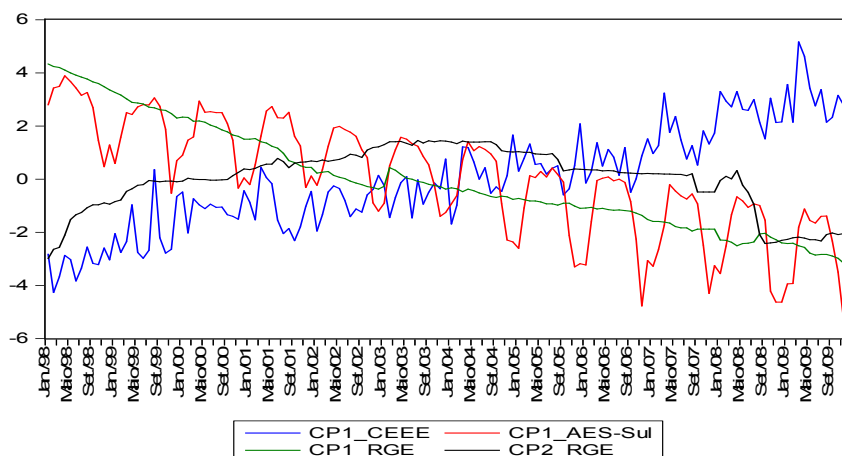


Figura 10 – Fatores que representam as três companhias para o número de consumidores de energia elétrica.

Nesta figura as companhias são representadas respectivamente pelos fatores derivados da ACP. O CP_{1CEEE} explica em 89,41% todos os setores da companhia; $CP_{1AES-Sul}$ – fator que representa todos os setores em 76,20%; CP_{1RGE} e CP_{2RGE} – representa a companhia RGE, que explica 87,98% da variabilidade dos seis setores mantidos pela empresa.

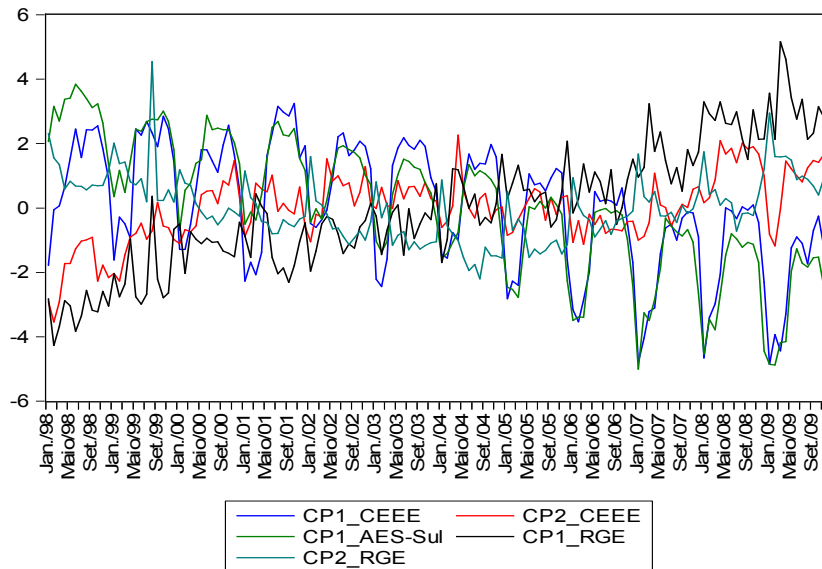


Figura 11 – Fatores que representam as três companhias para o fornecimento de energia elétrica.

Na Figura 11, as companhias são representadas respectivamente pelas componentes principais, onde a CP_{1CEEE} e CP_{2CEEE} em 80,90%; $CP_{1AES-Sul}$ – fator que representa todos os setores em 74,83%; CP_{1RGE} e CP_{2RGE} – representa a companhia RGE que explica 79,78%.

4.3 Verificação da estacionariedade das componentes principais

Como análise preliminar, é necessário investigar a estacionariedade das componentes principais das três distribuidoras de energia, de forma a classificar as séries em relação ao seu grau de integração.

A verificação da estacionariedade para as componentes principais que explicam o número de consumidores das distribuidoras de energia elétrica em estudo foi realizada por meio da construção de correlogramas, apresentados na Figura 12, 13, 14 e 15. Também conhecidos por função de autocorrelação (FAC) e

função de autocorrelação parcial (FACP), em conjunto com as estatísticas *Ljung-Box*, testes de raiz unitária e *Kwiatkowski-Phillips-Schmidt-Shin tests - KPSS*.

i) *Correlogramas*

Os gráficos da Figura 12, 13, 14, 15, apresentam as estatísticas *Ljung-Box*, que evidenciam a presença de autocorrelação para os primeiros 15 *lags*, pois os *p-valores* são menores que o nível de significância adotado. Além disso, percebe-se uma persistência nos dados, indicando uma dependência serial. Sendo assim, pode-se considerar que as séries não são estacionárias.

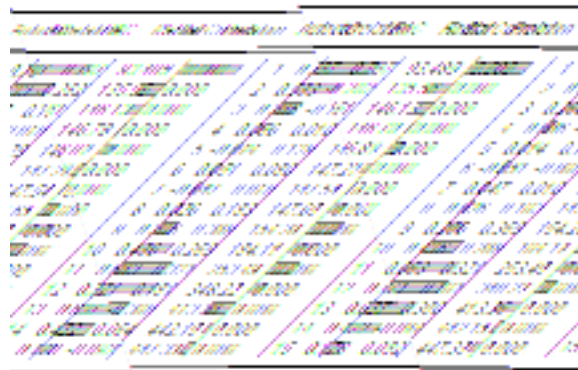


Figura 12 – Correlograma, estatística de Ljung-Box - Q(K) e *p-valor* da Componente principal para o número de consumidores da distribuidora CP1_{CEEE/RS}.

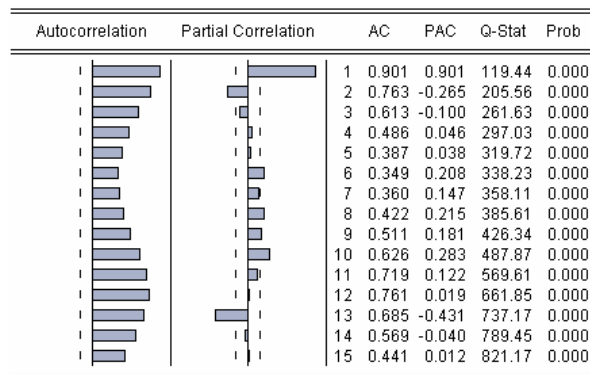


Figura 13 – Correlograma, estatística de Ljung-Box - Q(K) e *p-valor* da Componente principal para número de consumidores da distribuidora AES-Sul/RS.

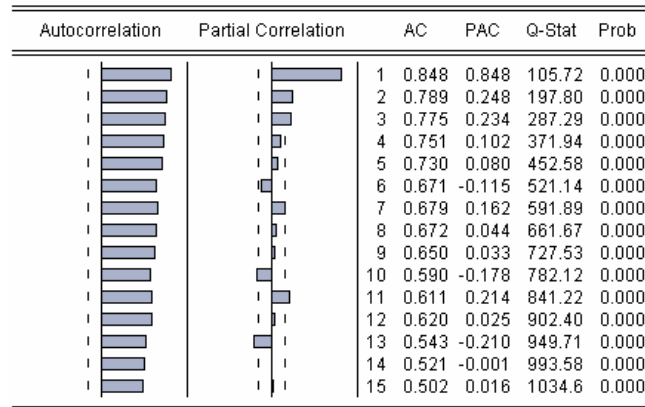


Figura 14 – Correlograma, estatística de Ljung-Box - Q(K) e p-valor da Componente principal para número de consumidores da para distribuidora CP1_{RGE}.

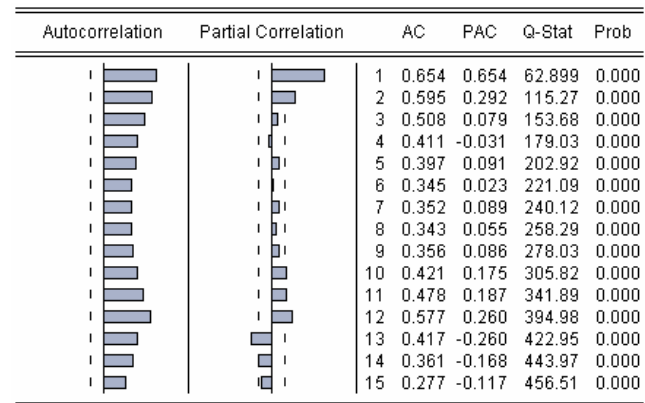


Figura 15 – Correlograma, estatística de Ljung-Box - Q(K) e p-valor da Componente principal para número de consumidores da distribuidora CP2_{RGE}.

A verificação da estacionariedade para as componentes principais que explicam o fornecimento de energia em MWh foi realizada por meio da construção de correlogramas, apresentados na Figura 16, 17, 18, 19 e 20, também conhecidos por função de autocorrelação (FAC) e função de autocorrelação parcial (FACP), em conjunto com as estatísticas *Ljung-Box*, testes de raiz unitária e KPSS.

Os gráficos das Figuras 16, 17, 18, 19 e 20 apresentam as estatísticas *Ljung-Box*, que evidenciam a presença de autocorrelação para os primeiros 15 *lags*, pois os *p-valores* são menores que o nível de significância adotado. Sendo assim, pode-se considerar que as séries não são estacionárias.

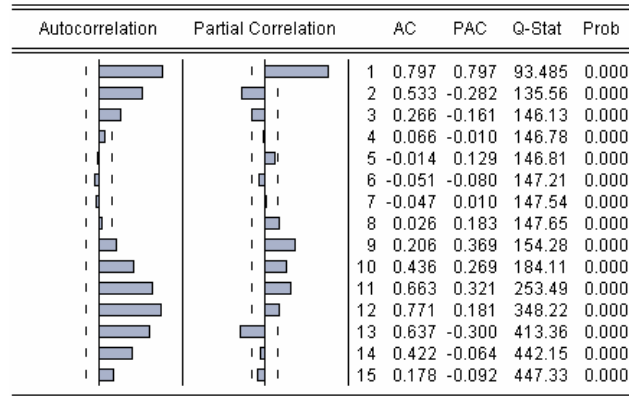


Figura 16 – Correlograma, estatística de Ljung-Box - Q(K) e p-valor da componente principal para o fornecimento de energia, em MWh, da distribuidora CP₁CEEE.

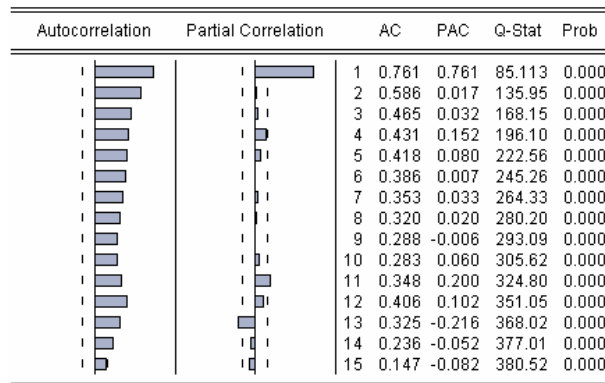


Figura 17 – Correlograma, estatística de Ljung-Box - Q(K) e p-valor da componente principal para o fornecimento de energia, em MWh, da distribuidora CP₂CEEE.

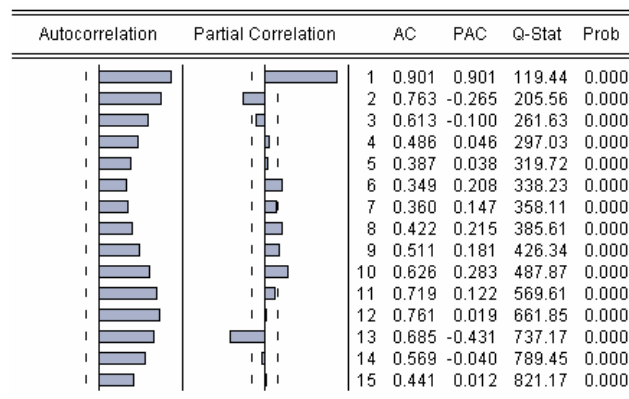


Figura 18 – Correlograma, estatística de Ljung-Box - Q(K) e p-valor da componente principal para o fornecimento de energia, em MWh, da distribuidora CP₁AES-Sul.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.848	0.848	105.72	0.000
		2	0.789	0.248	197.80	0.000
		3	0.775	0.234	287.29	0.000
		4	0.751	0.102	371.94	0.000
		5	0.730	0.080	452.58	0.000
		6	0.671	-0.115	521.14	0.000
		7	0.679	0.162	591.89	0.000
		8	0.672	0.044	661.67	0.000
		9	0.650	0.033	727.53	0.000
		10	0.590	-0.178	782.12	0.000
		11	0.611	0.214	841.22	0.000
		12	0.620	0.025	902.40	0.000
		13	0.543	-0.210	949.71	0.000
		14	0.521	-0.001	993.58	0.000
		15	0.502	0.016	1034.6	0.000

Figura 19 – Correlograma, estatística de Ljung-Box - Q(K) e p-valor da componente principal para o fornecimento de energia, em MWh, da distribuidora CP_{1RGE}.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.654	0.654	62.899	0.000
		2	0.595	0.292	115.27	0.000
		3	0.508	0.079	153.68	0.000
		4	0.411	-0.031	179.03	0.000
		5	0.397	0.091	202.92	0.000
		6	0.345	0.023	221.09	0.000
		7	0.352	0.089	240.12	0.000
		8	0.343	0.055	258.29	0.000
		9	0.356	0.086	278.03	0.000
		10	0.421	0.175	305.82	0.000
		11	0.478	0.187	341.89	0.000
		12	0.577	0.260	394.98	0.000
		13	0.417	-0.260	422.95	0.000
		14	0.361	-0.168	443.97	0.000
		15	0.277	-0.117	456.51	0.000

Figura 20 – Correlograma, estatística de Ljung-Box - Q(K) e p-valor da componente principal para o fornecimento de energia, em MWh, da distribuidora CP_{2RGE}.

Para testar estatisticamente a presença ou não de estacionariedade dos fatores, aplicam-se os testes de hipóteses ADF e KPSS conjuntamente, com a finalidade de obter um critério mais robusto na identificação do processo gerador dos fatores.

Testes de raiz unitária

Os testes de raiz unitária para as componentes principais das distribuidoras CEEE, AES-Sul e RGE para o estado do Rio Grande do Sul, quanto o número de consumidores, são mostrados na Tabela 7.

Tabela 7 – Teste de raiz unitária ADF, KPSS aplicados as componentes principais que representam respectivamente as distribuidoras CEEE, AES-Sul, e RGE, para o número de consumidores.

Distribuidora	ADF		KPSS	
CEEE (CP1)	-0,96	I(1)	1,41	I(0)
AES-Sul (CP1)	-0,09	I(1)	1,54	I(0)
RGE (CP1)	-2,36	I(1)	1,36	I(0)
RGE(CP2)	-0,83	I(1)	0,40	I(0)
Δ (CEEE– CP1)	-11,12	I(0)	0,07	I(1)
Δ (AES-Sul– CP1)	-16,03	I(0)	0,04	I(1)
Δ (RGE– CP1)	-10,51	I(0)	0,43	I(1)
Δ (RGE– CP2)	-6,34	I(0)	0,82	I(1)

Os testes de raiz unitária para as componentes principais das distribuidoras CEEE, AES-Sul e RGE para o estado do Rio Grande do Sul, quanto ao fornecimento de energia elétrica, em MWh, são mostrados na Tabela 8.

Tabela 8 – Teste de raiz unitária ADF, KPSS aplicados as componentes principais que representam respectivamente as distribuidoras CEEE, AES-Sul, e RGE, para o fornecimento de energia em MWh.

Distribuidora	ADF		KPSS	
CEEE (CP1)	-0,93	I(1)	1,27	I(0)
CEEE (CP2)	-4,19	I(1)	0,72	I(0)
AES-Sul (CP1)	-3,32	I(1)	1,49	I(0)
RGE (CP1)	-0,28	I(1)	1,46	I(0)
RGE(CP2)	-3,76	I(1)	0,38	I(0)
Δ (CEEE– CP1)	-6,56	I(0)	0,24	I(1)
Δ (CEEE– CP2)	-9,80	I(0)	0,50	I(1)
Δ (AES-Sul– CP1)	-15,47	I(0)	0,03	I(1)
Δ (RGE– CP1)	-8,77	I(0)	0,21	I(1)
Δ (RGE– CP2)	-18,65	I(0)	0,34	I(1)

4.4 Correlação linear entre as variáveis originais e as CP selecionadas

O termo correlação indica relação em dois sentidos (co + relação) e é usado em estatística para designar a força que mantém unido dois conjuntos de valores. A verificação da existência e do grau de relação entre as variáveis é o objeto de estudo da correlação.

A fim de constatar a existência de relação entre os setores e os fatores selecionados na ACP, isto é, verificar se as variáveis ou os setores estão bem representados pelos fatores selecionados.

Analisando a correlação entre as variáveis originais e as componentes principais selecionadas na análise do número de consumidores de energia elétrica para as três companhias, pode-se observar quais variáveis originais possuem maior influência na formação das componentes principais, conforme descritos na Tabela 9.

Alexander (2005) diz que uma forte correlação positiva indica que os movimentos crescentes de uma das séries de retorno tendem a ser acompanhadas por movimentos crescentes de outra e similarmemente movimentos decrescentes em ambas as séries caminham juntas. Se há uma forte correlação negativa, então os movimentos crescentes em uma das séries são associados aos movimentos decrescentes na outra.

Tabela 9 – Correlação entre número de consumidores das CP para CEEE, AES-Sul e RGE componente principal representativa desta distribuidora.

	CP ₁ CEEE	CP ₁ AES-Sul	CP ₁ RGE	CP ₂ RGE
Res	-0,99	-0,94	-0,97	-0,21
Com	-0,98	-0,98	-0,98	-0,15
Ind	-0,76	0,76	-0,32	0,76
Rural	-0,99	-0,79	-0,96	-0,08
S Publ	-0,98	-0,84	-0,98	-0,01
Outros	0,95	-0,91	0,23	-0,83

(em negrito os valores significativos)

Observa-se que todas as variáveis têm uma correlação forte significativa com a componente principal que a representa; a variável que tem menor correlação é a variável Industrial, mas mesmo assim é fortemente correlacionada.

Analisando a correlação entre as variáveis originais e as componentes principais selecionadas na análise do consumo de energia elétrica, em MWh, para

as três companhias, pode-se observar quais variáveis originais possuem maior influência na formação das componentes principais, conforme descritos na Tabela 10. Observa-se que, no geral, todas as variáveis têm uma correlação forte significativa com a componente principal que a representa. Há variáveis que têm menor correlação, mas mesmo assim apresentam elevada correlação correlacionada.

Tabela 10 – Correlação entre variáveis originais para o consumo de energia em MWh e a componente principal representativa de cada distribuidora.

Distribuidoras para o consumo em MWh de energia elétrica					
	CP ₁ CEEE	CP ₂ CEEE	CP ₁ AES-Sul	CP ₁ RGE	CP ₂ RGE
Res	-0,86	0,08	-0,94	0,89	0,26
Com	-0,92	0,31	-0,98	0,97	0,09
Ind	0,44	0,82	0,74	0,52	-0,81
Rural	-0,84	-0,20	-0,77	0,85	0,28
S Publ	-0,84	0,47	-0,82	0,89	-0,34
Outros	-0,73	-0,29	-0,91	0,47	0,33

(em negrito os valores significativos)

4.5 Modelos de Previsão das Componentes Principais Seleccionadas

Um dos principais interesses da análise de séries temporais é a previsão. No setor elétrico, a projeção constitui a base no processo de planejamento de investimentos em termos técnicos e estratégicos, sendo que a energia necessária a ser contratada pelas empresas distribuidoras advém do estudo das projeções de demanda.

4.5.1 Modelo para o número de consumidores de energia elétrica do RS

Os consumidores são o foco principal das empresas, pois a competição pela conquista de mercado faz com que estas passem a planejar o futuro de uma maneira mais complexa, envolvendo aspectos estratégicos, incertezas e, muitas vezes, limitações de informações.

Esta projeção de demanda é realizada pelas próprias empresas que realizam um estudo de mercado. O que será apresentado a seguir é a aplicação de uma metodologia que foge da tradicional, baseada em técnicas estatísticas mais complexas.

i) Análise da distribuidora de energia elétrica CEEE

A Companhia Estadual de Distribuição de Energia Elétrica – CEEE é uma empresa de economia mista pertencente ao Grupo CEEE, distribuidora dos serviços de distribuição de energia elétrica na região sul-sudeste do estado do Rio Grande do Sul.

Com área de concessão que compreende a região metropolitana de Porto Alegre, litoral e campanha gaúcha, a CEEE Distribuição atende a 72 municípios, abrangendo 73.627 km², o que corresponde aproximadamente a 32% do mercado consumidor do Rio Grande do Sul, através de seus 47.000 km de redes urbanas e rurais.

A CEEE Distribuição atendeu, em 2006, um total de 1.355 mil unidades consumidoras, o que equivale a cerca de 4 milhões de pessoas ou um terço da população gaúcha, distribuindo diretamente 6.287 GWh. No mesmo período, a empresa investiu R\$ 100 milhões em seu sistema de distribuição (Disponível em: <<http://www.ceee.com.br/pportal/ceee/Component/Controller.aspx?CC=1755>>, Acesso em: 02 dez. 2010).

No intuito de encontrar o melhor modelo que represente corretamente os diversos setores de consumo de CEEE, foi utilizado a CP1_{CEEE}, que explica 89,41% desta companhia. Dessa forma, uma análise de todos os setores foi englobada em uma única componente.

Na Tabela 11, sugere-se que o ARIMA foi o modelo de ruído branco que melhor se ajustou, cujos coeficientes são significativos no nível adotado.

Tabela 11 – Modelo ARIMA, dos critérios AIC para distribuidora CP1_{CEEE}.

CP1_{CEEE}							
	Coeficiente	D.Padrão	t-statistic	Prob.	AIC	Durbin-Watson	Jarque-Bera
ARIMA(1,1,0)	0.254690	0.081136	3.139057	0.002	-3.114	2.068	7.559

Mediante os modelos concorrentes obtidos, o melhor modelo ajustado para a componente principal que explica a distribuidora de energia elétrica CEEE é um ARIMA (1,1,0). Este modelo foi determinado com o auxílio do critério penalizador AIC. A Figura 21 ilustra o comportamento temporal dos resíduos do modelo ARIMA (1,1,0).

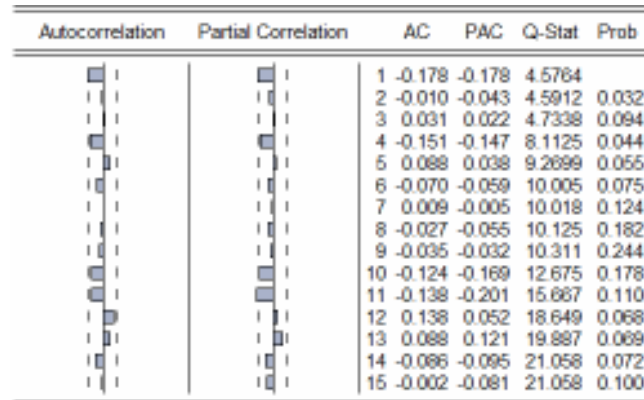


Figura 21 – Série residual do modelo ARIMA (1,1,0).

Na Figura 21, determina-se que o processo é estacionário, pois a FAC residual do modelo mostra que os *lag* apresentam as autocorrelações dentro dos limites de confiabilidade de 95%. Dessa forma, para confirmar que os resíduos não são autocorrelacionados, utilizou-se o teste de Ljung-Box Q^* , analisando os *p*-valores que foram maiores que o nível de significância adotado, assim considerando-os como estacionários.

Com o intuito de analisar a presença de heterocedasticidade na série residual, na Figura 22, apresenta-se o correlograma da série quadrática dos resíduos oriundos da CP modelada pelo ARIMA (1,1,0).

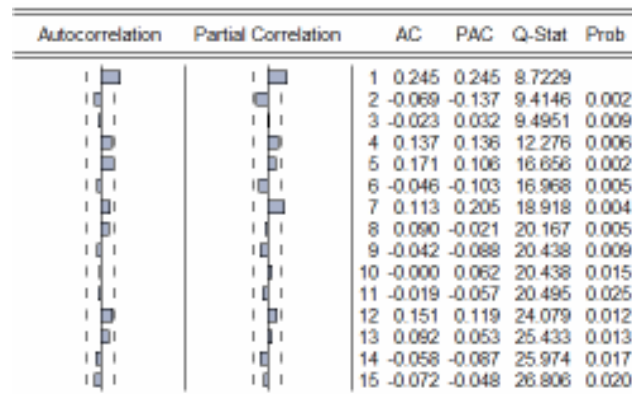


Figura 22 – Correlograma, estatística de Ljung-Box ($Q(K)$) e *p*-valor da série dos resíduos quadráticos do modelo ARIMA (1,1,0) para a componente principal que explica a concessionária $CP_{1CEE E}$.

Analisando a Figura 22, determina-se que o processo é estacionário e não autocorrelacionado, pois a FAC residual do modelo mostra que os *lag* apresentam as autocorrelações dentro dos limites de confiabilidade de 95%, e os *p-valores* são maiores que o nível de significância adotado.

Para testar estatisticamente a presença de heterocedasticidade condicional, aplicou-se o teste de hipótese ARCH-LM nos resíduos quadráticos do modelo ARIMA (1,1,0). Como foi constatada a presença desse fenômeno, há justificativa para a utilização de um modelo não linear da família ARCH para modelar a volatilidade. Encontra-se, na Tabela 12, o resultado deste teste.

Para comprovar a presença de efeitos ARCH, foram testadas as seguintes hipóteses:

H_0 : Não existem efeitos ARCH (homocedasticidade);

H_1 : Existem efeitos ARCH (heterocedasticidade).

Tabela 12 – Resultados das estatísticas *F* e ARCH-LM.

ARCH – LM			
F-statistic	10,84	Prob. F(1,139)	0,003
TR ²	18,35	$\chi^2_{(2)}$	0,004
ARCH			
F-statistic	4,47	Prob.F(1,139)	0,036
TR ²	4,40	$\chi^2_{(2)}$	0,036

A partir dos resultados encontrados, verifica-se que a estatística *F* e a estatística ARCH-LM foram significativas. Assim, rejeita-se a hipótese nula, implicando que há heterocedasticidade condicional nos resíduos quadráticos deste modelo. Isso posto, justifica-se a utilização do modelo ARCH com intuito de modelar a volatilidade do CP_{1CEEE}

Na Tabela 13, observam-se os resultados da estimação do modelo com heterocedasticidade do fator CP_{1CEEE} para o número de consumidores desta companhia de distribuição de energia elétrica.

Tabela 13 – Modelo com heterocedasticidade do fator CP1CEEE para o número de consumidores.

Método: ML – ARCH (Marquardt) – Distribuição Normal				
Equação para variância condicional				
	Coefficiente	Erro-Padrão	Estatística Z	p-valor.
Constante	0,83	0,09	-3,79	0,002
ARCH(1)	0,47	0,14	3,50	0,005

O modelo para a volatilidade do CP_{1CEEE} é um ARCH(1), que apresenta parâmetros estatisticamente significativos. Como o parâmetro do ARCH(1) é razoavelmente baixo, igual a 0,47, indica-se que, após um curto período de tempo, a variância da série deverá convergir à média histórica, ou seja, os setores que compõe a CP_{1CEEE} no período anterior não influenciaram no próximo período. Desta forma, pode-se dizer que, no período analisado, não há uma grande, flutuação na série que venha a comprometer o comportamento global das estimativas da média do processo. A região de abrangência da CEEE, que compreende as regiões metropolitanas de Porto Alegre, litoral e campanha gaúcha, deve ser a que mais sofre com a alternância do número de consumidores.

Para a validação do modelo que melhor explicou a volatilidade do CP_{1CEEE} foi preciso fazer a análise dos resíduos, por meio do teste de Ljung-Box aplicado aos resíduos e também aos resíduos quadráticos dos mesmos. A Figura 23 ilustra o comportamento da FAC e FACP dos resíduos do modelo ARCH(1).

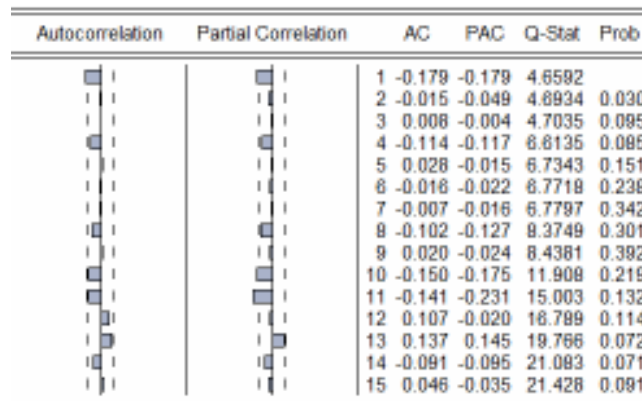


Figura 23 – Comportamento residual da FAC e FACP do modelo ARCH (1).

Analisando a Figura 23, determina-se que o processo é não autocorrelacionado e estacionário, pois a FAC residual do modelo mostra que os *lag* apresentam as autocorrelações dentro dos limites de confiabilidade de 95% e os *p-valores* são maiores que o nível de significância adotado, mostrando que os resíduos são não autocorrelacionados, indicando que a série de resíduos é um ruído branco.

Buscando verificar a existência de heterocedasticidade residual, faz-se a FAC e a FACP para os resíduos quadráticos.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.013	0.013	0.0235	
		2	-0.083	-0.083	1.0180	0.313
		3	-0.007	-0.004	1.0244	0.599
		4	0.218	0.213	8.0821	0.044
		5	-0.011	-0.018	8.0989	0.088
		6	-0.078	-0.048	9.0101	0.108
		7	0.057	0.064	9.5059	0.147
		8	0.158	0.108	13.363	0.064
		9	-0.045	-0.041	13.881	0.090
		10	0.015	0.063	13.717	0.133
		11	-0.030	-0.062	13.959	0.180
		12	0.138	0.093	16.834	0.113
		13	0.133	0.188	19.849	0.074
		14	-0.051	-0.056	20.057	0.084
		15	-0.047	-0.033	20.414	0.118

Figura 24 – Correlograma, estatística de Ljung-Box (Q(K)) e *p-valor* da série dos resíduos quadráticos do modelo ARCH (1) para a componente principal que explica a concessionária CP_{1CEEE}.

Analisando a Figura 24, determina-se que o processo é estacionário e não apresenta uma estrutura de dependência serial, pois a FAC e FACP residual do modelo mostra que os *lag* apresentam as autocorrelações dentro dos limites de confiabilidade de 95%, e os *p-valores* são maiores que o nível de significância adotado, mostrando que os resíduos são um ruído branco.

Para testar estatisticamente se o modelo conseguiu captar a presença de heterocedasticidade condicional, aplicou-se o teste de hipótese ARCH nos resíduos quadráticos do modelo ARCH(1), conforme Tabela 14.

Para comprovar a presença de efeitos ARCH, foram testadas as seguintes hipóteses:

H_0 : Não existem efeitos ARCH (homocedasticidade);

H_1 : Existem efeitos ARCH (heterocedasticidade).

Tabela 14 – Resultados das estatísticas F e ARCH-LM.

ARCH			
F-statistic	0,02	Prob.F(2,115)	0,880
TR ²	0,02	$\chi^2_{(2)}$	0,879

A partir dos resultados encontrados, verifica-se que a estatística F não foi significativa. Assim, aceita-se a hipótese nula, implicando que não há heterocedasticidade condicional nos resíduos quadráticos do modelo.

A Figura 25 mostra os valores originais, os valores estimados e os resíduos para a série $CP1_{CEEE}$.

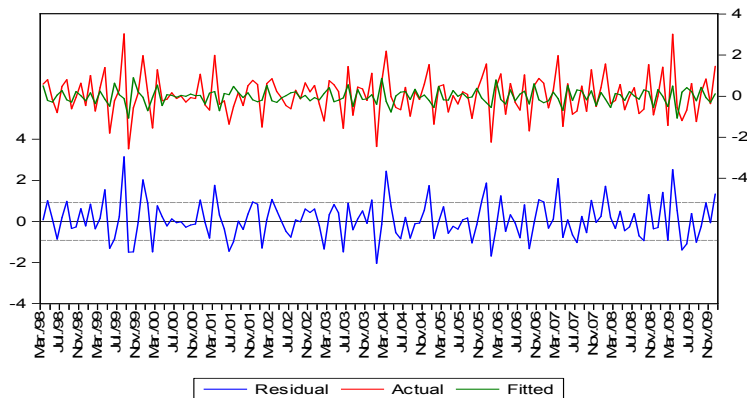


Figura 25 – Valores originais, valores estimados e os resíduos para o modelo ARCH(1).

ii) Análise da distribuidora de energia elétrica AES-Sul

Empresa do grupo americano *AES Corporation*, a AES-Sul é uma distribuidora do serviço público de energia elétrica. Com sede em Porto Alegre, atua na região centro-oeste do estado do Rio Grande do Sul. Atende a 118 municípios das regiões metropolitana e centro-oeste do estado, com aproximadamente 1.146 milhão de clientes, nas categorias residencial, industrial e rural, entre outros, sendo 4.184 milhões de pessoas beneficiadas e 50 subestações (Disponível em: <<http://www.aessul.com.br/site/empresa/Default.aspx>> Acesso em: 02 dez. 2009).

Analisando os resultados descritos anteriormente, o fator $CP_{1AES-Sul}$ apresenta um comportamento não estacionário em relação à média. Estas conclusões são demonstradas pela análise da FAC e FACP e do teste ADF e KPSS.

Do exposto anteriormente, busca-se a estimação do melhor modelo que represente a distribuidora AES-Sul para o número de consumidores.

Na Tabela 15, observa-se as estatísticas do modelo ARIMA que melhor se ajustou e formou um ruído branco no nível de significância adotado. Foram testados vários modelos concorrentes, os quais foram testados e avaliados em relação à parcimônia dos parâmetros, juntamente com a estatística AIC.

Tabela 15 – Resultados do ajuste do modelo ARIMA para distribuidora CP1_{AES-Sul}.

CP1 _{AES-Sul}							
	Coeficiente	D.Padrão	t-statistic	Prob.	AIC	Durbin-Watson	Jarque-Bera
ARI(1)	-0.424419	0.078284	-5.421546	0.0000	-0.4389	1.861	5.106
MA(1)	-0.859849	0.046902	-18.33277	0.0000			

O melhor modelo ajustado para a componente principal que explica a distribuidora de energia elétrica AES-Sul é um ARIMA (1,1,1). Na Figura 26 observa-se o comportamento temporal dos resíduos do modelo.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.063	0.063	0.4774	
		2	0.063	0.059	0.9673	
		3	-0.018	-0.025	1.0068	0.316
		4	0.127	0.127	3.0317	0.220
		5	-0.096	-0.112	4.2058	0.240
		6	-0.022	-0.023	4.2648	0.371
		7	-0.057	-0.037	4.6757	0.457
		8	-0.081	-0.097	5.5187	0.479
		9	-0.092	-0.051	6.6376	0.468
		10	-0.079	-0.071	7.4679	0.487
		11	0.024	0.046	7.5433	0.581
		12	-0.100	-0.093	8.8973	0.542
		13	-0.062	-0.060	9.4147	0.584
		14	0.001	0.021	9.4147	0.667
		15	0.017	-0.022	9.4538	0.738

Figura 26 – Série residual do modelo ARIMA (1,1,1).

Na Figura 26, determina-se que o processo é não autocorrelacionado e estacionário, pois a FAC residual do modelo mostra que os *lag* apresentam a autocorrelação dentro dos limites de confiabilidade de 95%. Dessa forma, para confirmar que os resíduos não são autocorrelacionados, utilizou-se o teste de Ljung-Box Q*, analisando os *p-valores* maiores que o nível de significância adotado. Sendo assim, pode-se afirmar que a série é estacionária e é ruído branco.

Buscando verificar a existência de heterocedasticidade residual, faz-se a FAC e a FACP para os resíduos quadráticos do modelo ARIMA (1,1,1).

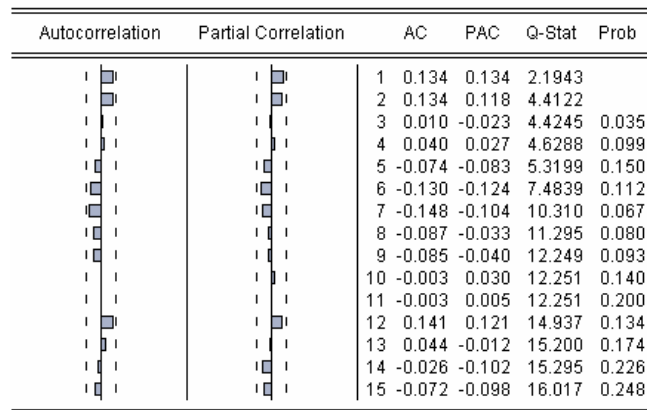


Figura 27 – Correlograma, estatística de Ljung-Box (Q(K)) e *p*-valor da série dos resíduos quadráticos do modelo ARIMA (1,1,1) para a componente principal que explica a concessionária CP_{1AES-Sul}.

Analisando a Figura 27, determina-se que o processo é estacionário, pois a FAC residual do modelo mostra que os *lag* apresentam a autocorrelação dentro dos limites de confiabilidade de 95%, e os *p*-valores são maiores que o nível de significância adotado, admitindo-se que o modelo é ruído branco.

Para testar estatisticamente a presença de heterocedasticidade condicional, aplicou-se o teste de hipótese ARCH-LM nos resíduos quadráticos do modelo ARIMA (1,1,1), o qual não sinalizou a presença de heterocedasticidade residual. Dessa forma, não há justificativa para a utilização de um modelo não linear da família ARCH para modelar a volatilidade. Encontra-se, na Tabela 16, o resultado desse teste.

Para comprovar a presença de efeitos ARCH, foram testadas as seguintes hipóteses:

H_0 : Não existem efeitos ARCH (homocedasticidade);

H_1 : Existem efeitos ARCH (heterocedasticidade).

Tabela 16 – Testes para presença de heterocedasticidade.

ARCH – LM			
F-statistic	0,80	Prob.F(2,115)	0,455
TR ²	1,59	$\chi^2_{(2)}$	0,451
ARCH			
F-statistic	2,13	Prob.F(1,116)	0,157
TR ²	2,12	$\chi^2_{(2)}$	0,144

A Figura 28 mostra os valores originais, os valores estimados e os resíduos para a série CP_{1CEEE} .

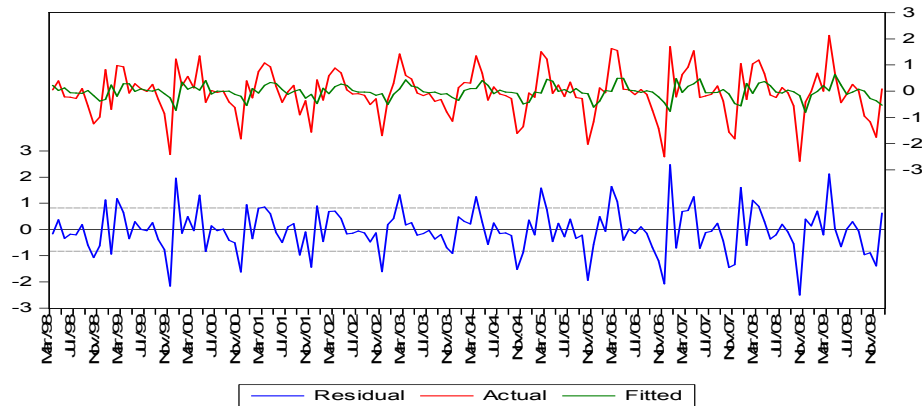


Figura 28 – Valores originais, valores estimados e os resíduos para o modelo $ARCH(1) - CP_{1AES-Sul}$.

Análise da distribuidora de energia elétrica RGE

A Rio Grande Energia (RGE) é a distribuidora de energia elétrica da região norte-nordeste do estado do Rio Grande do Sul. Privatizada em outubro de 1997, a RGE atende a 262 municípios gaúchos, o que representa 51% do total de municípios do RS.

A área de cobertura da Rio Grande Energia divide-se em duas regiões: a centro, com sede em Passo Fundo e a leste, com sede em Caxias do Sul. São 90.718km² - 34% do território do estado.

Agrupadas, essas regiões apresentam um dos melhores índices sociais e econômicos do Brasil. Também são as responsáveis pelo maior pólo agrícola, pecuário, industrial e turístico do estado. A RGE faz parte do Grupo CPFL Energia, um dos maiores grupos privados do setor elétrico brasileiro (Disponível em: <<http://www.rge-rs.com.br/ARGE/QuemSomos/tabid/116/language/pt-BR/Default.aspx>> Acesso em: 03 dez. 2010).

A distribuidora RGE passou a ser representada por duas componentes principais, desta forma têm-se duas análises em separado.

i) RGE CP_1

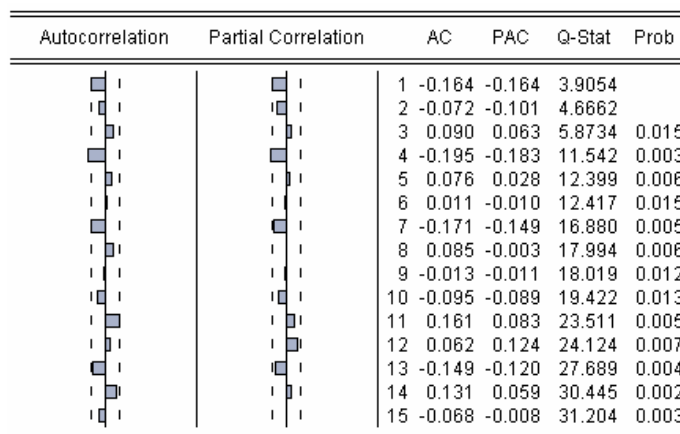
Entre os modelos concorrentes, o melhor modelo ajustado para a componente principal que explica a distribuidora de energia elétrica CP_{1RGE} , foi o SARIMA (1,1,0) (1,0,0)₁₂, conforme Tabela 17.

Tabela 17 – Modelo ARIMA, dos critérios AIC para distribuidora CP1_{RGE}.

CP1 _{RGE}							
	Coeficiente	D.Padrão	t-statistic	Prob.	AIC	Durbin-Watson	Jarque-Bera
ARI(1)	0,24	0,006	2,750	0,006	-2,98	1,99	9,48
SAR(1) ₁₂	0,19	0,091	2,075	0,039			

Mediante os modelos concorrentes obtidos, o melhor modelo ajustado para a componente principal que explica a distribuidora de energia elétrica RGE é SARIMA (1,1,0) (1,0,0)₁₂, e este modelo foi determinado com o auxílio do critério penalizador AIC.

A Figura 29 ilustra o comportamento da FAC e FACP dos resíduos do modelo SARIMA (1,1,0) (1,0,0)₁₂.

Figura 29 – Comportamento residual da FAC e FACP SARIMA (1,1,0) (1,0,0)₁₂.

Na Figura 29, determina-se que o processo é não autocorrelacionado e estacionário, pois a FAC residual do modelo mostra que os *lag* apresentam as autocorrelações dentro dos limites de confiabilidade de 95%. Dessa forma, para confirmar que os resíduos não são autocorrelacionados utilizou-se o teste de Ljung-Box Q*, analisando os *p-valores* maiores que o nível de significância adotado, mostrando que os resíduos são não autocorrelacionados.

Buscando verificar a existência de heterocedasticidade residual, faz-se a FAC e a FACP para os resíduos quadráticos, conforme a Figura 30.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.141	0.141	2.9227	
		2	-0.014	-0.034	2.9500	
		3	-0.005	0.002	2.9543	0.086
		4	-0.030	-0.030	3.0865	0.214
		5	-0.005	0.004	3.0899	0.378
		6	-0.037	-0.039	3.2961	0.510
		7	0.030	0.042	3.4329	0.634
		8	-0.027	-0.041	3.5451	0.738
		9	-0.018	-0.006	3.5935	0.825
		10	0.000	-0.001	3.5935	0.892
		11	-0.026	-0.025	3.7025	0.930
		12	-0.021	-0.018	3.7715	0.957
		13	0.075	0.084	4.6572	0.947
		14	0.009	-0.020	4.6692	0.968
		15	-0.018	-0.013	4.7240	0.981

Figura 30 – Correlograma, estatística de Ljung-Box (Q(K)) e *p-valor* da série dos resíduos quadráticos do modelo SARIMA (1,1,0) (1,0,0)₁₂ para a componente principal que explica a concessionária CP_{1RGE}.

Analisando a Figura 30, determina-se que o processo é estacionário e de dependência serial que necessite ser modelada, pois a FAC e a FACP residual do modelo mostra que os *lag* apresentam as autocorrelações dentro dos limites de confiabilidade de 95%, e os *p-valor*s são maiores que o nível de significância adotado.

Para testar estatisticamente a presença de heterocedasticidade condicional, aplicou-se o teste de hipótese ARCH-LM aos dados do modelo ARIMA (1,1,1) conforme a Tabela 18. Como não foi constatada a presença desse fenômeno, não há justificativa para a utilização de um modelo não linear da família ARCH para modelar a volatilidade.

Para comprovar a presença de efeitos ARCH, foram testadas as seguintes hipóteses:

H_0 : Não existem efeitos ARCH (homocedasticidade);

H_1 : Existem efeitos ARCH (heterocedasticidade).

Tabela 18 – Testes para presença de heterocedasticidade.

ARCH – LM			
F-statistic	4,41	Prob.F(2,115)	0,014
Obs*R-squared	0,000	$\chi^2_{(2)}$	1,000
ARCH			
F-statistic	2,86	Prob.F(1,116)	0,093
TR ²	2,84	$\chi^2_{(2)}$	0.091

A Figura 31 mostra os valores originais, os valores estimados e os resíduos para a série $CP1_{CEEE}$.

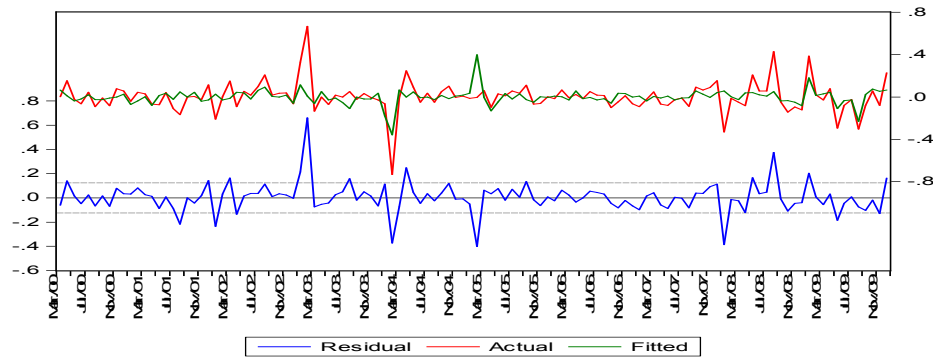


Figura 31 – Valores originais, valores estimados e os resíduos para o modelo $ARCH(1) - CP1_{RGE}$.

ii) $RGE CP_2$

Mediante os modelos concorrentes obtidos, o melhor modelo ajustado para a componente principal que explica a distribuidora de energia elétrica $CP2_{RGE}$, Tabela 19, é um $ARIMA(1,1,0)$. Esse modelo foi determinado pelos critérios penalizadores AIC.

Tabela 19 – Modelo $ARIMA$, dos critérios AIC para distribuidora $CP2_{CEEE}$.

$CP2_{RGE}$							
	Coeficiente	D.Padrão	t-statistic	Prob.	AIC	Durbin-Watson	Jarque-Bera
$ARI(1)$	0,35	0,078	4,476	0,000	-2,92	1,424	3447,162

Mediante os modelos concorrentes obtidos, o melhor modelo ajustado para a componente principal que explica a distribuidora de energia elétrica RGE é $ARIMA(1,1,0)$. A Figura 32 ilustra o comportamento temporal dos resíduos do modelo $ARIMA(1,1,0)$.

Autocorrelation		Partial Correlation		AC	PAC	Q-Stat	Prob
				1	0.086	0.086	1.0904
				2	-0.097	-0.105	2.4689
				3	0.022	0.041	2.5401
				4	-0.166	-0.186	6.6285
				5	0.027	0.074	6.7401
				6	-0.019	-0.076	6.7952
				7	-0.159	-0.130	10.659
				8	0.046	0.035	10.980
				9	-0.016	-0.046	11.021
				10	-0.059	-0.050	11.560
				11	0.166	0.131	15.903
				12	0.071	0.056	16.710
				13	-0.105	-0.117	18.468
				14	0.082	0.092	19.545
				15	-0.036	-0.023	19.753

Figura 32 – Comportamento residual da FAC e FACP ARIMA (1,1,0).

A Figura 32 mostra que o processo é estacionário e os resíduos não autocorrelacionados, pois a FAC residual do modelo mostra que os *lag* apresentam as autocorrelações dentro dos limites de confiabilidade de 95%. Dessa forma, utilizou-se o teste de Ljung-Box Q*, analisando os *p-valores* maiores que o nível de significância adotado, indicando que a série de resíduos é um ruído branco. Como forma de testar a heterocedasticidade residual da série em estudo, as FAC e a FACP podem ser observadas na Figura 33.

Autocorrelation	Partial Correlation	Autocorrelation	Partial Correlation
1 0.008 0.008 0.0049 1	1 -0.008 -0.008 0.0049 1		
2 0.004 0.004 0.0079 0.993	2 0.004 0.004 0.0079 0.993		
3 0.003 0.003 0.0184 0.991	3 0.003 0.003 0.0184 0.991		
4 -0.004 -0.004 0.1389 0.998	4 0.004 0.004 0.1389 0.998		
5 0.002 0.002 0.1389 0.998	5 -0.002 -0.002 0.1389 0.998		
6 -0.001 -0.001 0.2724 0.998	6 0.001 0.001 0.2724 0.998		
7 0.003 0.003 0.4340 0.999	7 -0.003 -0.003 0.4340 0.999		
8 -0.001 -0.001 0.5003 0.999	8 0.001 0.001 0.5003 0.999		
9 0.008 0.008 0.6149 0.000	9 -0.008 -0.008 0.6149 0.000		
10 0.007 0.007 0.6913 0.000	10 0.007 0.007 0.6913 0.000		
11 0.030 0.028 0.8050 0.000	11 -0.030 -0.028 0.8050 0.000		
12 -0.004 -0.004 0.8204 0.000	12 0.004 0.004 0.8204 0.000		
13 0.073 0.073 1.0920 0.000	13 -0.073 -0.073 1.0920 0.000		
14 -0.017 -0.017 1.0959 0.000	14 0.017 0.017 1.0959 0.000		
15 0.016 0.016 1.7360 0.000	15 -0.016 -0.016 1.7360 0.000		

Figura 33 – Correlograma, estatística de Ljung-Box (Q(K)) e *p-valor* da série dos resíduos quadráticos do modelo ARIMA (1,1,0) para a componente principal que explica a concessionária CP_{2RGE}.

Analisando a Figura 33, determina-se que o processo é não autocorrelacionado, pois a FAC residual do modelo mostra que os *lag* apresentam as autocorrelações dentro dos limites de confiabilidade de 95%, e os *p-valores* são maiores que o nível de significância adotado.

Além da inspeção visual, para testar estatisticamente a presença de heterocedasticidade condicional, aplicou-se o teste de hipótese ARCH-LM nos

resíduos quadráticos do modelo ARIMA (1,1,0). Como foi constatada a presença deste fenômeno, há justificativa para a utilização de um modelo não linear da família ARCH para modelar a volatilidade.

H_0 : Não existem efeitos ARCH (homocedasticidade);

H_1 : Existem efeitos ARCH (heterocedasticidade).

Tabela 20 – Testes para presença de heterocedasticidade.

ARCH – LM			
F-statistic	1,60	Prob.F(2,115)	0,204
TR ²	3,17	$\chi^2_{(2)}$	0,204
ARCH			
F-statistic	1,92	Prob.F(1,116)	0,168
TR ²	1,92	$\chi^2_{(2)}$	0,165

A partir dos resultados encontrados, verifica-se que a estatística F e a estatística ARCH-LM não foram significativas. Assim, aceita-se a hipótese nula, implicando que não há heterocedasticidade condicional nos resíduos quadráticos deste modelo. Isto posto, não se justifica a utilização do modelo ARCH com intuito de modelar a volatilidade do CP_{2RGE} .

A Figura 34 mostra os valores originais, os valores estimados e os resíduos para a série CP_{2RGE} .

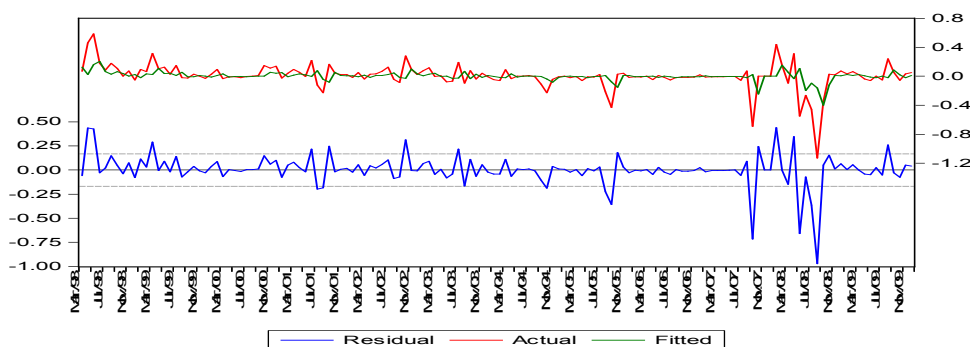


Figura 34 – Valores originais, valores estimados e os resíduos para o modelo ARIMA (1, 1, 0).

A modelagem de componentes principais por meio dos modelos ARIMA e/ou ARCH possibilita que, além da redução da dimensionalidade, ocorra a identificação

dos principais setores formadores daquela componente. De acordo com Alexander (2005), a ACP é capaz de acompanhar o movimento conjunto de diversas curvas, o que possibilita uma eficácia no acompanhamento de todas as variáveis envolvidas no sistema. Além disso, a análise possibilita uma ordenação natural das variáveis envolvidas no processo.

4.5.2 Modelo para o consumo de energia elétrica em MWh por classe de consumo no Rio Grande do Sul para CEEE, AES-Sul e RGE

A demanda de potência é medida e expressa em KW, que corresponde à potência elétrica solicitada pelo consumidor à empresa distribuidora durante o período de fornecimento mensal. O consumo efetivo é medido e expresso em KWh, o qual corresponde ao real consumo ao longo do período faturado, normalmente de 30 dias.

Desses componentes conjuntos – demanda de potência e/ou consumo e de acordo com a modalidade do fornecimento, estruturam-se os grupos de consumidores. O fornecimento se destina às unidades consumidoras, sendo estabelecidas para as seguintes classes ou setores e subclasses de consumo: Classe Residencial, Rural, Industrial, Comercial, Serviços Públicos e outros.

i) Análise do consumo em MWh da distribuidora de energia elétrica

Na busca de encontrar o melhor modelo que represente corretamente os diversos setores de consumo de CEEE, será utilizado a CP_{1CEE} , que explica 62,3% desta companhia, e a CP_{2CEE} , que explica 18,6%.

ii) Análise do consumo da distribuidora de energia elétrica CEEE

O consumo em MWh da CEEE é representado por duas componentes principais, a qual engloba um total de 80,90% de explicação. Esses componentes serão modelados nas etapas seguintes.

CEEE CP_1

A Tabela 21 sugere o modelo ARIMA que melhor se ajustou e forma um ruído branco, cujos coeficientes são significativos no nível de significância adotado, a série em estudo. Foram testados e avaliados, e o melhor modelo foi aquele que atendeu

as estatísticas de validação, conforme Tabela 21.

A Tabela 21 sugere o modelo SARIMA que melhor se ajustou e forma um ruído branco, cujos coeficientes são significativos no nível de significância adotado.

Tabela 21 – Modelo SARIMA, dos critérios AIC para distribuidora CEEE.

CP ₁ CEEE							
	Coeficiente	D.Padrão	t-statistic	Prob.	AIC	Durbin-Watson	Jarque-Bera
ARI(1)	0,36	0,113	3,163	0,002			
SAR(1) ₁₂	-0,62	0,070	-8,895	0,000	1,934	1,944	7,469
MA(1)	-0,88	0,054	-16,281	0,000			

Mediante os modelos concorrentes obtidos, o melhor modelo ajustado para a componente principal que explica a distribuidora de energia elétrica CEEE é SARIMA (1,1,1) (1,1,0)₁₂. Esse modelo foi determinado pelos critérios penalizadores AIC.

A Figura 35 ilustra o comportamento temporal dos resíduos do modelo SARIMA (1,1,1) (1,1,0)₁₂.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.020	0.020	0.0468	
		2 -0.026	-0.026	0.1274	
		3 0.044	0.045	0.3702	
		4 -0.007	-0.010	0.3762	0.540
		5 0.148	0.152	3.1327	0.209
		6 0.099	0.092	4.3622	0.225
		7 -0.139	-0.137	6.8356	0.145
		8 -0.003	-0.007	6.8366	0.233
		9 0.168	0.165	10.520	0.104
		10 -0.125	-0.152	12.558	0.084
		11 -0.045	-0.072	12.828	0.118
		12 -0.182	-0.172	17.245	0.045
		13 -0.127	-0.089	19.422	0.035
		14 0.163	0.114	23.027	0.018
		15 0.018	0.034	23.071	0.027

Figura 35 – Comportamento residual da FAC e FACP SARIMA (1,1,1) (1,1,0)₁₂.

Na Figura 35, determina-se que o processo é não autocorrelacionado, pois a FAC e FACP residual do modelo mostra que os *lag* apresentam as autocorrelações dentro dos limites de confiabilidade de 95%, sinalizando um processo ruído branco. Dessa forma, para confirmar que os resíduos não são autocorrelacionados, utilizou-

se o teste de Ljung-Box Q^* , analisando os p -valores maiores que o nível de significância adotado. Sendo assim, pode-se afirmar que os resíduos seguem um ruído branco.

Como forma de testar a heterocedasticidade residual da série em estudo, as FAC e a FACP podem ser observadas na Figura 36.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.033	-0.033	0.1333	
		2	0.023	0.022	0.2006	
		3	0.018	0.020	0.2415	
		4	-0.080	-0.079	1.0282	0.311
		5	-0.014	-0.021	1.0542	0.590
		6	-0.002	0.001	1.0546	0.788
		7	-0.150	-0.148	3.9330	0.415
		8	0.015	-0.000	3.9620	0.555
		9	-0.074	-0.071	4.6768	0.586
		10	-0.065	-0.070	5.2363	0.631
		11	0.097	0.074	6.4695	0.595
		12	0.088	0.097	7.5040	0.585
		13	-0.058	-0.070	7.9638	0.632
		14	-0.069	-0.119	8.6205	0.657
		15	0.038	0.050	8.8226	0.718

Figura 36 – Correlograma, estatística de Ljung-Box ($Q(K)$) e p -valor da série dos resíduos quadráticos do modelo SARIMA (1,1,1) (1,1,0)₁₂ para a componente principal que explica a concessionária CP_{1CEEE}.

Analisando a Figura 36, determina-se que o processo é estacionário, pois a FAC residual do modelo mostra que os *lag* apresentam as autocorrelações dentro dos limites de confiabilidade de 95%, e os p -valores são maiores que o nível de significância adotado. Assim, a hipótese nula de que as 15 primeiras autocorrelações apresentam valor zero não é rejeitada, podendo os resíduos ser considerados ruído branco.

Para testar estatisticamente a presença de heterocedasticidade condicional, aplicou-se o teste de hipótese ARCH-LM, Tabela 22, nos resíduos quadráticos do modelo SARIMA (1,1,1) (1,1,0)₁₂. Como não foi constatada a presença deste fenômeno, não há justificativa para a utilização de um modelo não linear da família ARCH para modelar a volatilidade.

Para comprovar a presença de efeitos ARCH, foram testadas as seguintes hipóteses:

H_0 : Não existem efeitos ARCH (homocedasticidade);

H_1 : Existem efeitos ARCH (heterocedasticidade).

Tabela 22 – Resultados das estatísticas F e ARCH-LM.

ARCH – LM			
F-statistic	0,27	Prob.F(2,115)	0,770
T.R ²	0,55	$\chi^2_{(2)}$	0,760
ARCH			
F-statistic	0,13	Prob.F(1,116)	0,722
TR ²	0,13	$\chi^2_{(2)}$	0,719

A partir dos resultados encontrados, verifica-se que a estatística F e a estatística ARCH-LM não foram significativas. Assim, aceita-se a hipótese nula, implicando que não há heterocedasticidade condicional nos resíduos quadráticos deste modelo.

A Figura 37 mostra os valores originais, os valores estimados e os resíduos para a série CP₂CEEE.

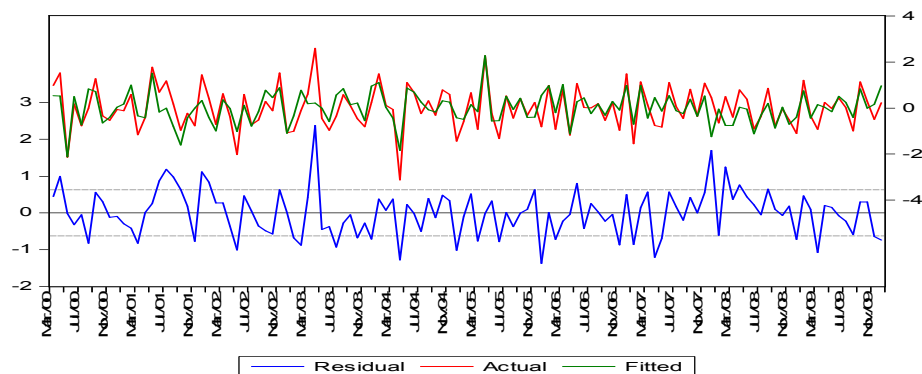


Figura 37 – Valores originais, valores estimados e os resíduos para o modelo SARIMA $(1,1,1) (1,1,0)_{12}$.

CEEE CP₂

A Tabela 23 sugere o modelo ARIMA que melhor se ajustou e forma um ruído branco, cujos coeficientes são significativos no nível de significância adotado, a série em estudo. Foram testados vários modelos concorrentes, que foram testados e avaliados segundo a quantidade de parâmetros significantes e a estatística AIC, entre todos os modelos ARIMA concorrentes.

Tabela 23 – Modelo ARIMA, dos critérios AIC para distribuidora CP_{2CEEE}.

CP _{2CEEE}							
	Coeficiente	D.Padrão	t-statistic	Prob.	AIC	Durbin-Watson	Jarque-Bera
ARI(1)	0,57	0,095	5,950	0,000	1,879	1,961	6,060
MA(1)	-0,89	0,050	-17,813	0,000			

Mediante os modelos concorrentes obtidos, o melhor modelo ajustado para a componente principal que explica a distribuidora de energia elétrica CP_{2CEEE} é ARIMA (1,1,1). A determinação desse modelo foi auxiliada pelo critério penalizador AIC.

A Figura 38 ilustra o comportamento temporal dos resíduos do modelo ARIMA (1,1,1).

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.008	-0.008	0.0102	
		2 -0.007	-0.008	0.0184	
		3 -0.120	-0.120	2.1338	0.144
		4 -0.014	-0.017	2.1634	0.339
		5 0.094	0.094	3.4957	0.321
		6 0.024	0.012	3.5836	0.465
		7 0.004	0.001	3.5856	0.610
		8 -0.035	-0.013	3.7675	0.708
		9 -0.090	-0.085	5.0067	0.659
		10 -0.125	-0.139	7.4341	0.491
		11 0.003	-0.010	7.4359	0.592
		12 0.279	0.271	19.649	0.033
		13 0.034	0.022	19.837	0.048
		14 0.049	0.062	20.217	0.063
		15 -0.099	-0.015	21.792	0.059

Figura 38 – Comportamento residual da FAC e FACP ARIMA (1,1,1).

Analisando a Figura 38, determina-se que o processo é estacionário e não autocorrelacionado, pois a FAC e a FACP residual do modelo mostram que os *lag* apresentam as autocorrelações dentro dos limites de confiabilidade de 95%, e os *p-valores* são maiores que o nível de significância adotado. Assim, a hipótese nula de que as 15 primeiras autocorrelações apresentam valor zero não é rejeitada, podendo os resíduos ser considerados ruído branco.

Como forma de testar a heterocedasticidade residual da serie em estudo, as FAC e a FACP como mostra a Figura 39.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.013	-0.013	0.0262	
		2 -0.048	-0.049	0.3672	
		3 0.019	0.017	0.4184	0.518
		4 -0.037	-0.039	0.6257	0.731
		5 -0.066	-0.066	1.2786	0.734
		6 -0.066	-0.073	1.9351	0.748
		7 -0.114	-0.124	3.9197	0.561
		8 -0.050	-0.064	4.2998	0.636
		9 0.065	0.046	4.9463	0.667
		10 -0.038	-0.050	5.1657	0.740
		11 -0.021	-0.037	5.2347	0.813
		12 0.057	0.023	5.7398	0.837
		13 -0.004	-0.026	5.7423	0.890
		14 -0.056	-0.073	6.2427	0.903
		15 0.007	-0.013	6.2516	0.937

Figura 39 – Correlograma, estatística de Ljung-Box (Q(K)) e *p*-valor da série dos resíduos quadráticos do modelo ARIMA (1,1,1) para a componente principal que explica a concessionária CP₂CEEE.

Analisando a Figura 39, determina-se que o processo é não autocorrelacionado e estacionário, pois a FAC residual do modelo mostra que os *lag* apresentam as autocorrelações dentro dos limites de confiabilidade de 95%, e os *p*-valores são maiores que o nível de significância adotado, podendo os resíduos ser considerados ruído branco.

Para testar estatisticamente a presença de heterocedasticidade condicional, aplicou-se o teste de hipótese ARCH-LM nos resíduos quadráticos do modelo ARIMA (1,1,1), conforme mostra a Tabela 24. Como foi constatada a presença deste fenômeno, há justificativa da utilização de um modelo não linear da família ARCH para modelar a volatilidade.

Para comprovar a presença de efeitos ARCH, foram testadas as seguintes hipóteses:

H_0 : **Não existem efeitos ARCH (homocedasticidade);**

H_1 : **Existem efeitos ARCH (heterocedasticidade).**

Tabela 24 – Resultados das estatísticas *F* e ARCH-LM.

ARCH – LM			
F-statistic	0,136	Prob.F(2,115)	0,872
TR ²	0,000	$\chi^2_{(2)}$	1,000
ARCH			
F-statistic	0,126	Prob.F(1,116)	0,722
TR ²	0,129	$\chi^2_{(2)}$	0,719

A partir dos resultados encontrados, verifica-se que a estatísticas F e a estatística ARCH-LM não foram significativas. Assim, aceita-se a hipótese nula, implicando que não há heterocedasticidade condicional nos resíduos quadráticos desse modelo.

A Figura 40 mostra os valores originais, os valores estimados e os resíduos para a série CP_{2RGE} .

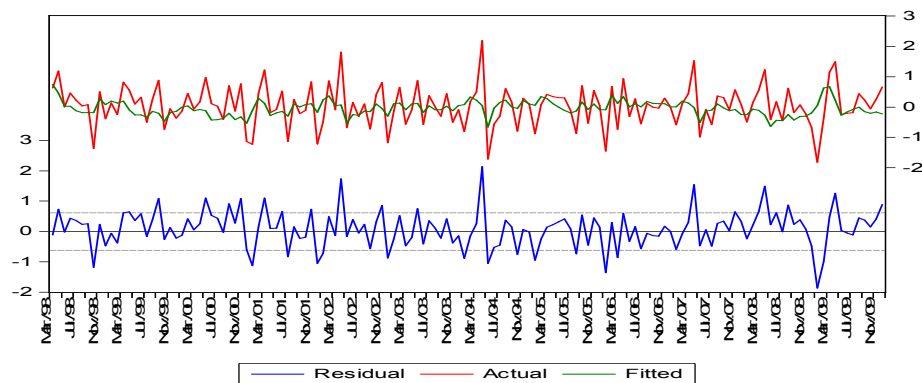


Figura 40 – Valores originais, valores estimados e os resíduos para o modelo ARIMA (1,1,1).

Análise do consumo da distribuidora de energia elétrica AES – Sul

Para encontrar o melhor modelo, que represente corretamente os diversos setores de consumo da AES-Sul, será utilizado a $CP_{1AES-Sul}$, que explica 74,83% desta companhia. A análise de todos os setores será englobada em uma única componente.

A Tabela 25 sugere o modelo SARIMA que melhor se ajustou e forma um ruído branco, cujos coeficientes são significativos no nível de significância adotado, a série em estudo. Foram testados vários modelos concorrentes, avaliados segundo a quantidade de parâmetros significantes e a estatística AIC, entre todos os modelos ARIMA concorrentes.

Tabela 25 – Modelo SARIMA, dos critérios AIC para distribuidora AES-Sul.

$CP_{1AES-Sul}$							
	Coeficiente	D.Padrão	t-statistic	Prob.	AIC	Durbin-Watson	Jarque-Bera
SARI(12)	-0,410	0,077	-5,313	0,000	1,165	1,811	7,469
MA(1)	-0,857	0,041	-20,860	0,000			

Mediante os modelos concorrentes obtidos, o melhor modelo ajustado para a componente principal que explica a distribuidora de energia elétrica AES-Sul é SARIMA (0,1,1) (1,1,0)₁₂. Esse modelo foi determinado com o auxílio dos critérios penalizadores AIC.

A Figura 41 ilustra o comportamento dos resíduos do modelo SARIMA (0,1,1) (1,1,1)₁₂.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.074	0.074	0.6631	
		2 0.099	0.094	1.8772	
		3 0.018	0.004	1.9169	0.166
		4 0.131	0.122	4.0748	0.130
		5 -0.068	-0.089	4.6517	0.199
		6 -0.005	-0.017	4.8547	0.325
		7 -0.050	-0.038	4.9748	0.419
		8 -0.057	-0.088	5.2955	0.494
		9 -0.060	-0.024	5.8633	0.556
		10 -0.092	-0.082	6.9921	0.537
		11 0.017	0.050	7.0308	0.634
		12 -0.106	-0.091	8.5315	0.577
		13 -0.027	-0.016	8.6303	0.656
		14 0.018	0.050	8.8838	0.731
		15 0.036	0.009	8.8443	0.785

Figura 41 – Comportamento residual da FAC e FACP do modelo SARIMA (0,1,1) (1,1,0)₁₂.

Analisando a Figura 41, determina-se que o processo é estacionário e não autocorrelacionado, pois a FAC e FACP residual do modelo mostra que os *lag* apresentam as autocorrelações dentro dos limites de confiabilidade de 95% e os *p-valores* são maiores que o nível de significância adotado. Assim, a hipótese nula de que as 15 primeiras autocorrelações apresentam valor zero não é rejeitada, podendo os resíduos ser considerados ruído branco.

Buscando verificar a existência de heterocedasticidade residual, faz-se a FAC e FACP para os resíduos quadráticos conforme Figura 42.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.117	0.117	1.6748	
		2 0.103	0.090	2.9743	
		3 0.002	-0.020	2.9747	0.085
		4 0.025	0.018	3.0524	0.217
		5 -0.088	-0.091	3.9996	0.263
		6 -0.106	-0.094	5.4166	0.247
		7 -0.141	-0.107	7.9833	0.158
		8 -0.089	-0.050	8.9849	0.174
		9 -0.093	-0.058	10.124	0.182
		10 -0.024	-0.002	10.201	0.251
		11 0.011	0.019	10.218	0.333
		12 0.141	0.123	12.894	0.230
		13 0.091	0.032	13.782	0.245
		14 -0.068	-0.143	14.413	0.275
		15 -0.082	-0.114	15.353	0.298

Figura 42 – Correlograma, estatística de Ljung-Box (Q(K)) e *p-valor* da série dos resíduos quadráticos do modelo SARIMA (0,1,1) (1,1,0)₁₂ para a componente principal que explica a concessionária CP₁AES-Sul.

Analisando a Figura 42, determina-se que o processo é estacionário e não apresenta estrutura de dependência serial que necessite ser modelado, pois a FAC residual do modelo mostra que os *lag* apresentam as autocorrelações dentro dos limites de confiabilidade de 95%, e os *p-valores* são maiores que o nível de significância adotado. Assim, a hipótese nula de que as 15 primeiras autocorrelações apresentam valor zero não é rejeitada, podendo os resíduos ser considerados ruído branco.

Para testar estatisticamente a presença de heterocedasticidade condicional, aplicou-se o teste de hipótese ARCH-LM, Tabela 26, nos resíduos quadráticos do modelo SARIMA (0,1,1) (1,1,0)₁₂. Como foi constatada a presença desse fenômeno, há justificativa da utilização de um modelo não linear da família ARCH para modelar a volatilidade.

Para comprovar a presença de efeitos ARCH, foram testadas as seguintes hipóteses:

H_0 : **Não existem efeitos ARCH (homocedasticidade);**

H_1 : **Existem efeitos ARCH (heterocedasticidade).**

Tabela 26 – Resultados das estatísticas *F* e ARCH-LM.

ARCH – LM			
F-statistic	1,334	Prob.F(2,115)	0,267
TR ²	2,432	$\chi^2_{(2)}$	0,296
ARCH			
F-statistic	1,645	Prob.F(1,116)	0,202
TR ²	1,650	$\chi^2_{(2)}$	0,199

A partir dos resultados encontrados, verifica-se que a estatística *F* e a estatística ARCH-LM não foram significativas. Assim, aceita-se a hipótese nula, implicando que não há heterocedasticidade condicional nos resíduos quadráticos deste modelo.

A Figura 43 mostra os valores originais, os valores estimados e os resíduos para a série CP_{AES-Sul}.

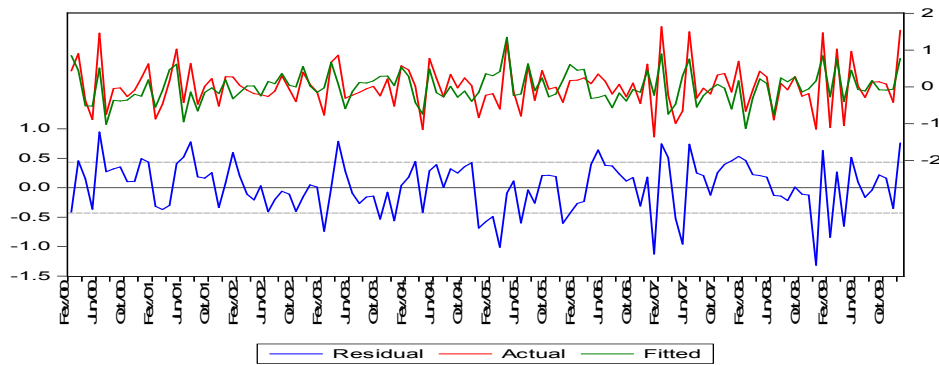


Figura 43 – Valores originais, valores estimados e os resíduos para o modelo SARIMA $(0,1,1) (1,1,0)_{12}$.

Análise do consumo da distribuidora de energia elétrica RGE

Para encontrar o melhor modelo que represente corretamente os diversos setores de consumo da RGE, será utilizado a $CP1_{REG}$, que explica 62,50% e $CP2_{RGE}$, que explica 17,28%. A análise de todos os setores será englobada em duas componentes.

Mediante os modelos concorrentes obtidos, o melhor modelo ajustado para a componente principal que explica a distribuidora de energia elétrica $CP1_{RGE}$, Tabela 27, é um SARIMA $(0,1,1) (1,1,1)_6$. Esse modelo foi determinado pelos critérios penalizadores AIC.

Tabela 27 – Modelo SARIMA, dos critérios AIC para distribuidora AES-Sul.

$CP1_{RGE}$							
	Coeficiente	D.Padrão	t-statistic	Prob.	AIC	Durbin-Watson	Jarque-Bera
SARI(6)	-0,941	0,043	-21,656	0,000	2,456	1,619	0,438
MA(1)	-1,929	0,018	-106,573	0,000			
MA(2)	0,930	0,017	53,146	0,000			
SMA(6)	0,431	0,098	4,387	0,000			

Mediante os modelos concorrentes obtidos, o melhor modelo ajustado para a componente principal $CP1_{RGE}$, que explica a distribuidora de energia elétrica RGE em 65,65% é SARIMA $(0,1,1) (1,1,1)_6$.

A Figura 44 ilustra o comportamento dos resíduos do modelo SARIMA $(0,1,1) (1,1,1)_6$.

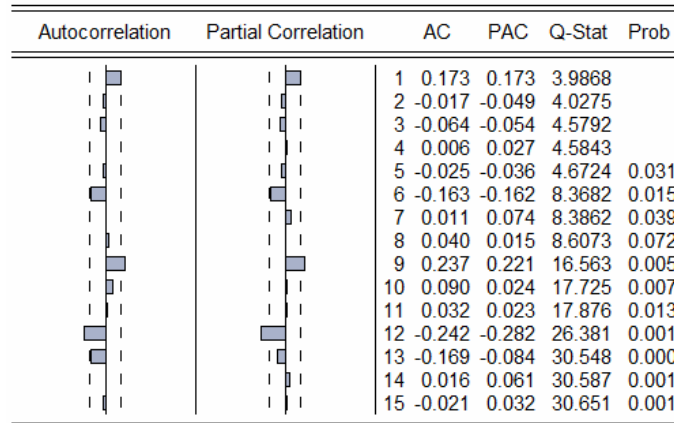


Figura 44 – Correlograma, estatística de Ljung-Box (Q(K)) e *p*-valor da série dos resíduos do modelo SARIMA (0,1,1) (1,1,1)₆ para a componente principal que explica a concessionária CP1_{RGE}.

Analisando a Figura 44, pode-se perceber que o processo é estacionário e não possui uma estrutura de dependência serial, pois a FAC e a FACP residual do modelo mostra que os *lag* apresentam as autocorrelações dentro dos limites de confiabilidade de 95%, e os *p*-valores são maiores que o nível de significância adotado. Assim, a hipótese nula de que as 15 primeiras autocorrelações apresentam valor zero não é rejeitada, podendo os resíduos ser considerados ruído branco.

Buscando verificar a existência de heterocedasticidade residual, faz-se a FAC e FACP para os resíduos quadráticos, conforme Figura 45.

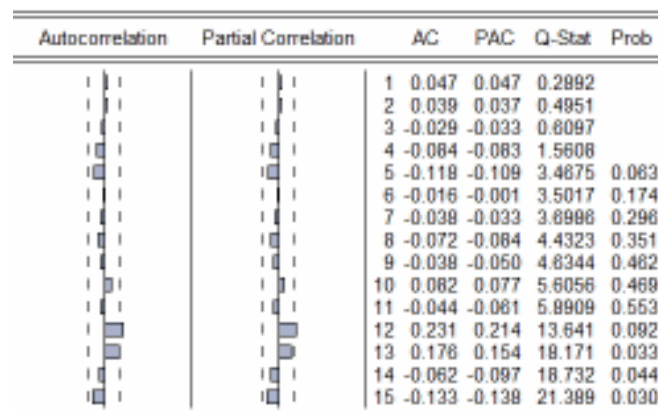


Figura 45 – Correlograma, estatística de Ljung-Box (Q(K)) e *p*-valor da série dos resíduos quadráticos do modelo SARIMA (0,1,1) (1,1,1)₆ para a componente principal que explica a concessionária CP1_{RGE}.

Para testar estatisticamente a presença de heterocedasticidade condicional, aplicou-se o teste de hipótese ARCH-LM, Tabela 28, nos resíduos quadráticos do modelo SARIMA (0,1,1)(1,1,1)₆. Como não foi constatada a presença desse

fenômeno, há justificativa da utilização de um modelo não linear da família ARCH para modelar a volatilidade.

Para comprovar a presença de efeitos ARCH, foram testadas as seguintes hipóteses:

H_0 : **Não existem efeitos ARCH (homocedasticidade);**

H_1 : **Existem efeitos ARCH (heterocedasticidade).**

Tabela 28 – Resultados das estatísticas F e ARCH-LM.

ARCH – LM			
F-statistic	2,303	Prob.F(2,115)	0,104
TR ²	4,060	$\chi^2_{(2)}$	0,131
ARCH			
F-statistic	0,286	Prob.F(1,116)	0,593
TR ²	0,290	$\chi^2_{(2)}$	0,589

A partir dos resultados encontrados, verifica-se que a estatística F e a estatística ARCH-LM não foram significativas. Assim, aceita-se a hipótese nula, implicando que não há heterocedasticidade condicional nos resíduos quadráticos do modelo SARIMA (0,1,1) (1,1,1)₆.

A Figura 46 mostra os valores originais, os valores estimados e os resíduos para a série CP1_{RGE}.

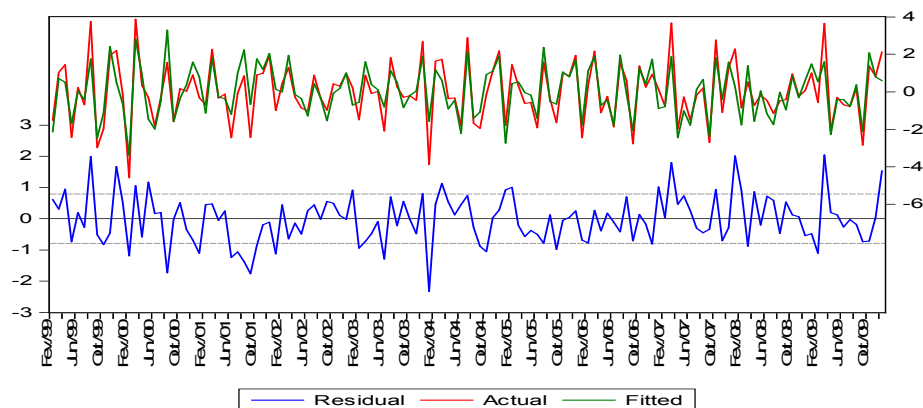


Figura 46 – Valores originais, valores estimados e os resíduos para o modelo SARIMA (0,1,1) (1,1,1)₆.

Para componente principal que explica a distribuidora CP2_{RGE} em 17,28%, o modelo que melhor representa é um modelo SARIMA (0,1,1) (1,1,1)₆. As estatísticas do modelo estão descritas na Tabela 29.

Tabela 29 – Modelo ARIMA, dos critérios AIC para distribuidora CP2_{RGE}.

CP _{1RGE}							
	Coeficiente	D.Padrão	t-statistic	Prob.	AIC	Durbin-Watson	Jarque-Bera
SARI(6)	-0,553	0,079	-6,947	0,000			
MA(1)	-0,688	0,064	-10,746	0,000	1,923	2,092	294,90
SMA(6)	-0,783	0,060	-13,007	0,000			

Mediante os modelos concorrentes obtidos, o melhor modelo ajustado para a componente principal CP2_{RGE} que explica a distribuidora de energia elétrica RGE é um modelo SARIMA (0,1,0) (1,1,1)₆.

A Figura 47 ilustra o comportamento dos resíduos do modelo SARIMA (0,1,0) (1,1,1)₆.

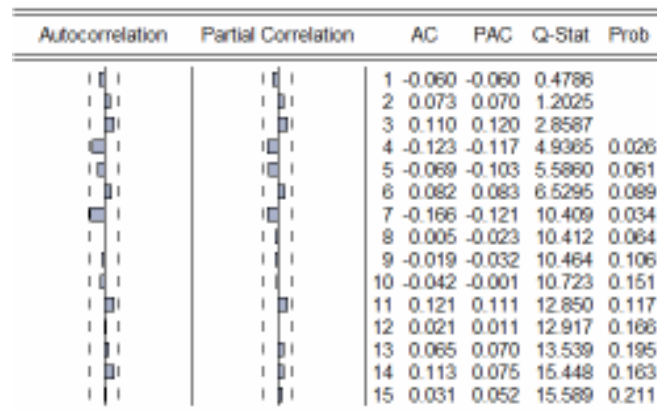


Figura 47 – Correlograma, estatística de Ljung-Box (Q(K)) e p-valor da série dos resíduos do modelo SARIMA (0,1,0) (1,1,1)₆ para a componente principal que explica a concessionária CP2_{RGE}.

Analisando a Figura 47, determina-se que o processo é estacionário e não autocorrelacionado, pois a FAC e FACP residual do modelo mostra que os *lag* apresentam as autocorrelações dentro dos limites de confiabilidade de 95%, e os *p-valores* são maiores que o nível de significância adotado. Assim, a hipótese nula de

que as 15 primeiras autocorrelações apresentam valor zero não é rejeitada, podendo os resíduos ser considerados ruído branco.

Buscando verificar a existência de heterocedasticidade residual, faz-se a FAC e FACP para os resíduos quadráticos conforme Figura 48.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.146	0.146	2.8473	
		2	0.020	-0.001	2.9010	
		3	-0.055	-0.059	3.3189	
		4	-0.025	-0.009	3.4043	0.065
		5	-0.009	-0.003	3.4155	0.181
		6	0.052	0.052	3.7916	0.285
		7	0.005	-0.012	3.7953	0.434
		8	-0.015	-0.018	3.8283	0.574
		9	-0.019	-0.009	3.8792	0.693
		10	0.026	0.033	3.9801	0.782
		11	0.036	0.027	4.1646	0.842
		12	0.242	0.234	12.739	0.175
		13	-0.009	-0.081	12.752	0.238
		14	-0.002	0.010	12.753	0.310
		15	-0.029	-0.001	12.878	0.378

Figura 48 – Correlograma, estatística de Ljung-Box (Q(K)) e *p*-valor da série dos resíduos quadráticos do modelo SARIMA (0,1,0) (1,1,1)₆ para a componente principal que explica a concessionária CP2_{RGE}.

Analisando a Figura 48, determina-se que o processo é estacionário e não apresenta uma estrutura de dependência serial, pois a FAC e a FACP residual do modelo mostra que os *lag* apresentam as autocorrelações dentro dos limites de confiabilidade de 95%, e os *p*-valores são maiores que o nível de significância adotado. Assim, a hipótese nula de que as 15 primeiras autocorrelações apresentam valor zero não é rejeitada, podendo os resíduos ser considerados ruído branco.

Para testar estatisticamente a presença de heterocedasticidade condicional, aplicou-se o teste de hipótese ARCH-LM, Tabela 30, nos resíduos quadráticos do modelo SARIMA (0,1,0) (1,1,1)₆. Como não foi constatada a presença desse fenômeno, não há justificativa da utilização de um modelo não linear da família ARCH para modelar a volatilidade.

Para comprovar a presença de efeitos ARCH, foram testadas as seguintes hipóteses:

H_0 : Não existem efeitos ARCH (homocedasticidade);

H_1 : Existem efeitos ARCH (heterocedasticidade).

Tabela 30 – Resultados das estatísticas F e ARCH-LM.

ARCH – LM			
F-statistic	0,557	Prob.F(2,115)	0,573
TR ²	0,490	$\chi^2_{(2)}$	0,782
ARCH			
F-statistic	2,779	Prob.F(1,116)	0,097
TR ²	2,762	$\chi^2_{(2)}$	0,096

A partir dos resultados encontrados, verifica-se que a estatística F e a estatística ARCH-LM não foram significativas. Assim, aceita-se a hipótese nula, implicando que não há heterocedasticidade condicional nos resíduos quadráticos do modelo SARIMA (0,1,1) (1,1,1)₆.

A Figura 49 mostra os valores originais, os valores estimados e os resíduos para a série CP_{2RGE}.

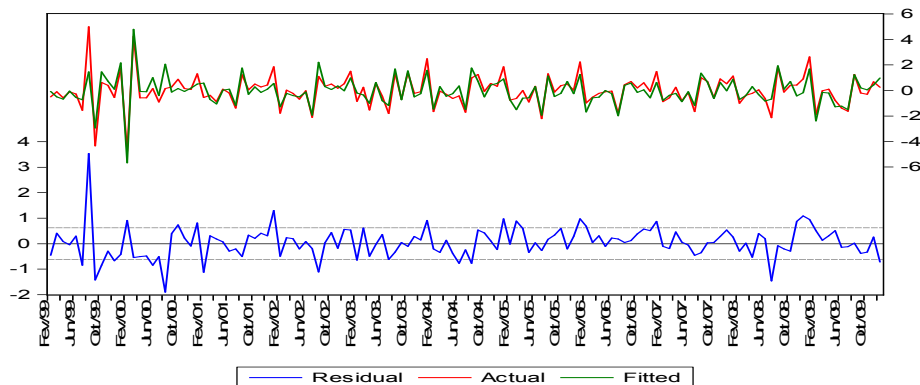


Figura 49 – Valores originais, valores estimados e os resíduos para o modelo SARIMA (0,1,1) (1,1,1)₆.

Outra aplicação importante da análise de componentes principais é a ortogonalização das novas variáveis, fazendo com que possíveis problemas de multicolinearidade que pudessem existir não afetarão o sistema.

Observa-se que, nos gráficos dos valores previstos, a série estimada dentro da amostra mantém comportamento similar com a da componente principal. Sendo assim, é possível determinar as possíveis mudanças de inclinação, bem como deslocamentos de magnitudes diferentes.

5 CONSIDERAÇÕES FINAIS

Este capítulo faz o encerramento das idéias e análises desenvolvidas durante esta pesquisa. Desta forma, serão apresentadas as conclusões, buscando relacionar o referencial teórico com as percepções relevantes levantadas durante o trabalho.

A pesquisa foi desenvolvida com o setor de energia elétrica do Rio Grande do Sul, apresentando uma metodologia alternativa às técnicas tradicionais de previsão do comportamento do número de consumidores e do consumo em MWh de energia elétrica. A inovação é determinar uma variável de referência para analisar o comportamento do consumo e dos consumidores de energia elétrica do Rio Grande do Sul por meio de modelos de previsão lineares e não lineares obtidos por componentes principais.

Cumprindo as etapas metodológicas, determina-se a variável de referência como a variável de maior representatividade, não somente as estruturas, mas também as volatilidades implícitas dos diversos setores que esta representa.

A metodologia proposta é capaz de acompanhar e determinar o movimento das séries em estudo de forma global, proporcionando medidas de gerenciamento tanto das empresas distribuidora de fornecimento quanto dos consumidores de cada empresa. A análise global mostra a tendência de oscilação das variáveis, de modo que não seja necessário se fazer uma análise individual de cada setor de consumo e distribuição.

Analisando a volatilidade das séries, foi possível observar que não existe uma oscilação muito forte na variabilidade das séries quando modeladas utilizando as componentes principais. Mas este fato é corroborado por Alexander (2005), que utiliza as componentes como um filtro para tratar a volatilidade, o que na maioria das vezes apresenta a série de componentes principais com características homocedásticas.

A distribuidora que revelou a existência de volatilidade foi a CEEE, mas com um valor abaixo de 0,5, o que mostra uma fraca volatilidade e demonstra que o nível previsto da série por meio do modelo ARIMA retornará para o patamar de desenvolvimento esperado em curto espaço de tempo. Em se tratando do número de consumidores, é possível atribuir essa volatilidade a aumentos e/ou decréscimos de consumidores, o que pode ser atribuído a planos governamentais como o

programa “Luz para Todos”, criado pelo Decreto n. 4.873, de 11 de novembro de 2003, destinado a propiciar, até o ano de 2008, o atendimento em energia elétrica à parcela da população do meio rural brasileiro que ainda não tem acesso a esse serviço público. O programa é coordenado pelo Ministério de Minas e Energia- MME e operacionalizado com a participação das Centrais Elétricas Brasileiras S.A. – Eletrobrás e das empresas que compõem o sistema Eletrobrás. Através do Decreto n. 6.442 de 25 de abril de 2008, o Governo Federal prorrogou o prazo de vigência do programa “Luz para Todos” para o final de 2010.

Conclui-se que o procedimento proposto, de primeiro aplicar a ACP nos dados para depois ajustá-los, mostrou-se satisfatório estatisticamente. Ao estudar a autocorrelação dos dados utilizando os modelos ARIMA e os modelos ARIMA-ARCH, temos uma nova perspectiva de aprendizagem sobre o processo produtivo através das informações contidas na estrutura da CP. Além disso, a investigação da estabilidade do processo se faz por meio da verificação da influência da volatilidade na detecção de clusters de volatilidade, que são potenciais para afetar o desempenho do processo produtivo.

Assim, o emprego do CP permitiu revelar o comportamento do processo, mostrando que grandes variações são capazes de influenciar na qualidade do produto, sugerindo, então, a possibilidade na extração das mais importantes fontes não autocorrelacionadas de variação de um sistema multivariado, não somente as estruturas a termo, mas também as volatilidades implícitas das diversas opções de um mesmo ativo. Dessa forma, a técnica torna-se útil para a aplicação do estudo proposto.

Além disso, a previsão do comportamento das variáveis é um recurso capaz de fornecer informações e subsídios para uma subsequente tomada de decisão.

Desta forma, sustenta-se que os modelos de previsão lineares e não lineares podem ser aplicados conjuntamente através de seus resíduos para estimar as variáveis em estudo da melhor forma possível, pois estes captam os movimentos globais das companhias em estudo e do número de consumidores visando, com isso, a um crescimento de informações para a correta tomada de decisão.

Deixa-se como sugestão de estudo a realização das previsões individuais das variáveis em estudo com o intuito de se obter valores individuais de previsões por setores, assim como a utilização de outros modelos de volatilidade como o *Generalized Autoregressive Conditional Heteroskedsticity* (GARCH) e *Integrated*

Generalized Autoregressive Conditional Heteroskedsticity (IGARCH), para verificar se as séries que não apresentaram volatilidade utilizando os modelos ARCH não apresentam estas outras características.

REFERÊNCIAS

ALEXANDER, C. **Modelos de Mercado**: um guia para a Análise de Informações Financeiras/Carol Alexander. Trad: José Carlos de Souza e Santos. São Paulo: Bolsa de Mercadorias & Futuros, 2005.

BAILLIE, R.; CHUNG, C. F.; TIESLAU, M. A. Analyzing inflation by the fractionally integrated ARFIMA-GARCH model. **Journal of Applied Econometrics**. 11: 23-40, 1996.

BARTLETT, M. S. On the theoretical specification and sampling properties of autocorrelated time-series. **Journal of the Royal Statistical Society Supplement**. 8: 27-41, 1946.

BAYER, F. M.; SOUZA, A. M. **Modelagem e previsão da taxa de ocupação hospitalar por meio de equações simultâneas: uma ferramenta de gestão**. GEPROS – Gestão de Produção, Operações e Sistemas. Ano 2, v. 2. mai./jun. 2007. p. 85-96.

BEM, A. B. **A utilização de regressão dinâmica para prever a demanda por matrícula no ensino fundamental em Santa Catarina**. 1998. 183f. (Dissertação em Engenharia de Produção) – Universidade Federal de Santa Catarina. Florianópolis, 1998.

BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. **Journal of Econometrics**. v. 31. 1986. p. 307-327.

BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis: Forecasting and Control**. San Francisco: Holden-Day, 1970.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis: forecasting and control**. 3. ed. San Francisco: Holden-Day, 1994.

BUENO, Rodrigo de Losso da Silveira. **Econometria das séries temporais**. São Paulo: Cengage Learning, 2008.

CAMPOS, K. C. Análise da volatilidade de preços de produtos agropecuários no Brasil. **Revista de Economia de Agronegócio**. v. 5, n. 3. 2007. p. 303-328.

CORRAR, J. L.; PAULO, E.; DIAS FILHO, J. M. **Análise multivariada para os cursos de administração, ciências contábeis e economia**. FIPECAF – Fundação Instituto de Pesquisas Contábeis, Atuariais e Financeiras. São Paulo: Atlas, 2007.

DYNER, I.; LARSEN, E. R. From Planning to Strategy in the Electricity Industry. **Energy Policy**. n. 29, 2001. p. 1145-1154.

ÉLIN DUXUS Consultorie e LTda. In: Manual e ajuda para o add-in Metrixus: Funções quantitativas para o mercado de capitais. Disponível em: <<http://www.duxus.com.br/produtos/tecnologia/addin/metrixus/manual/ajuda.php>>. Acesso em: 18 jun. 2009.

ENDERS, W. **Applied econometric time series**. Wiley series in probability and mathematical statistics. New York, NY: John Wiley and Sons, Inc., 1995.

ENGLE, R. F. **Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation**. *Econometrica*. 50: 987-1008, 1982.

FEE – Fundação de Economia e Estatística. Disponível em: <<http://www.fee.rs.gov.br>>. Acesso em: 10 mar. 2010.

GUJARATI, D. N. **Econometria básica**. São Paulo: Makron Books, 2000.

HAMILTON, J. **Time series analysis**. New Jersey: Princeton University Press, 1994.

HOTTELLING, H. Analysis of a complex of Statistical variables into principal components. **The Journal of Educational Psychology**. v. 24. 1933. p. 417-441/498-520.

ISSLER, J. V. Estimating and Forecasting the Volatility of Brazilian Finance Series Using ARCH Models. **The Brazilian Review of Econometrics**. v. 19, n. 1, 1999. p. 5-56.

JACKSON, G.; JONES, G. K. Spatial and temporal variation in nearshore fish and macroinvertebrate assemblage from a temperate Australian estuary over a decade. **Marine Ecology-Progress Series**. Oldendorf, 182: 253-268, 1999.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 3. ed. New Jersey: Prentice-Hall, 1992.

LI, Q.; TRICAUD, C.; SUN, R.; CHEN, Y. Q. **Great salt lake surface level forecasting using figarch modeling**. Proceedings of International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE (DETC2007-34909), Las Vegas, Nevada, USA, set. 2007. p. 4-7.

MADDALA, G. S. **Introduction to econometrics**. 2. ed. New Jersey: Prentice-Hall Inc. Englewood Cliffs, 1992.

MAKRIDAKIS, S. G.; WHEELWRIGHT, S. C.; HYNDMAN, R. J. **Forecasting: methods and applications**. 3. ed. New York: John Willey & Sons, Inc., 1998.

MORETTIN, P. A. **Econometria financeira: um curso em séries temporais financeiras**. 17º Simpósio Nacional de Probabilidade e Estatística. ABE. Caxambu, 2006.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. São Paulo: Edgard Blücher, 2004.

MORRISON, D. F. **Multivariate Statistical Methods**. New York: McGraw-Hill, 1976.

PEARSON, K. **Phil. Mag. J.** v. 6. 1901. p. 559.

PORTAL SÃO FRANCISCO. Disponível em: <<http://www.portalsaofrancisco.com.br/alfa/historia-da-eletricidade-no-brasil/historia-da-eletricidade-no-brasil-1.php>>. Acesso em: 10 mar. 2010.

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

REINSEL, G. C. **Elements of multivariate time series analysis**. New York: Springer-Verlag, 1993.

REIS, E. **Estatística multivariada aplicada**. 2. ed. Ed. Sílabo, 2001.

SÁFADI, T.; ANDRADE FILHO, M. G. **Abordagem Bayseana de modelos de séries temporais**. 12^a Escola de Séries Temporais e Econometria. ABE/SBE. 2007.

SCP-RS. Disponível em: <<http://www.scp.rs.gov.br/atlas/atlas.asp?menu=647>>. Acesso em: 10 mar. 2010.

SILVA, W. S.; SÁFADI, T.; CASTRO JÚNIOR, L. G. de. Uma análise empírica da volatilidade do retorno de commodities agrícolas utilizando modelos ARCH: os casos do café e da soja. **Revista de Economia e Sociologia Rural**. v. 43, n. 1. 2005. p. 120-134.

SILVA, W. V.; FONTANINI, C. A. C.; DEL CORSO, J. M. Garantia da qualidade do café solúvel com o uso do gráfico de controle de somas acumuladas. **Revista Produção On-Line**. v. 7, n. 2. 2007. p. 43-63.

SOARES, I. G.; CASTELAR, I. **Econometria aplicada com o uso de EViews**. Fortaleza: UFC/CAEN, 2003.

SOUZA, A. M. **Aplicação e Performance da Análise de Intervenção em Séries Macroeconômicas Brasileiras**. Monografia de Especialização (UFSM), 1991.

SOUZA, F. M.; LOPES, L. F. D. Previsão de demanda de leitos hospitalares por meio de análise de Séries Temporais. **Revista Ciência e Natura**. v. 31, n. 1. jun. 2009. p. 33-47.

VICINI, L.; SOUZA, A. M. Geração de subsídios para a tomada de decisão na cadeia produtiva da bonivicultura do Brasil. GEPROS – Gestão de Produção, Operações e Sistemas. Ano 2, v. 4. mai./jun. 2007. p. 49-64.

ZANELA, A.; LIMA, de S.; LOPES, L. F. D. **Identificação de Fatores que influenciam na satisfação dos clients de um clube recreativo por meio de análise fatorial**. GEPROS – Gestão de Produção, Operações e Sistemas. Ano 2, mai./jun. 2007. p. 167-176.