Pontifícia Universidade Católica
do Rio de Janeiro

Luiz Albino Teixeira Júnior

Combinação Geométrica de Métodos Preditivos; Aplicação à Previsão de Consumo Residencial Mensal de Energia Elétrica

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica da PUC-Rio.

Orientador: Prof. Reinaldo Castro Souza

Rio de Janeiro

Agosto de 2009

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Luiz Albino Teixeira Júnior

Combinação Geométrica de Métodos Preditivos; Aplicação à Previsão de Consumo Residencial Mensal de Energia Elétrica

Dissertação de Mestrado apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Reinaldo Castro Souza
Orientador
Departamento de Engenharia Elétrica - PUC-Rio

Profa. Marley Maria Bernardes Rebuzzi Vellasco
Departamento de Engenharia Elétrica - PUC-Rio

Prof. José Francisco Moreira Pessanha
CEPEL

Prof. José Eugenio Leal
Coordenador Setorial do Centro
Técnico Científico - PUC-Rio

Rio de Janeiro, 19 de agosto de 2009

Luiz Albino Teixeira Júnior

Graduado em Matemática pelas Faculdades Integradas de Cataguases em 2005

Ficha Catalográfica

Teixeira Júnior, Luiz Albino

Combinação geométrica de método preditivos; aplicação à previsão de consumo residencial mensal de energia elétrica / Luiz Albino Teixeira Júnior ; orientador: Reinaldo Castro Souza. – 2009.

115. ; 30 cm

Dissertação (Mestrado em Engenharia Elétrica) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2009.

Inclui bibliografia

1. Engenharia elétrica – Teses. 2. Consumo residencial mensal de energia elétrica. 3. ARIMA. 4. Redes neurais artificiais. 5. Amortecimento exponencial. 6. Combinação linear de modelos. 7. Combinação geométrica de modelos. I. Souza, Reinaldo Castro. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

Agradecimentos

Agradeço a Deus primeiramente, pois o Senhor tem me sustentando a cada dia; à minha mãe; à minha tia Irani, à minha esposa; bem como toda a minha família, por tudo.

Quero expressar também minha gratidão àqueles que colaboraram diretamente no meu curso: meu professor orientador e amigo Reinaldo Castro Souza; aos professores do departamento de engenharia elétrica; mui especialmente à professora Marley, ao professor Álvaro Veiga, ao professor Tara (Industrial); além do auxílio financeiro do CNPq.

Agradeço aos amigos (as) Márcio Leone, Álvaro Albuquerque, Iuri Steiner, Alcina, Márcia, dentre outros. Deixo os meus enormes e sinceros agradecimentos por suas participações, neste processo, tão importante na minha vida.

A Deus, primeiramente, à minha mãe, à minha tia Irani e à minha esposa, bem como toda minha família. Por tudo que são e representam para mim.

“Posso todas as coisas naquele que me fortalece”
(Filipenses 4: 17)

Resumo

Teixeira Júnior, Luiz Albino; Souza, Reinaldo Castro (Orientador). **Combinação Geométrica de Métodos Preditivos; Aplicação à Previsão de Consumo Residencial Mensal de Energia Elétrica**. Rio de Janeiro, 2009. 115p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Este trabalho propõe a aplicação de algumas metodologias de previsão de séries temporais combinadas geometricamente para a projeção pontual e intervalar de consumo residencial mensal de energia elétrica, em um horizonte de previsão de doze passos à frente. Foram utilizados três métodos: ARIMA, Amortecimento Exponencial e Redes Neurais Artificiais, bem como suas combinações lineares e geométricas. Os desempenhos dos modelos foram comparados utilizando as estatísticas residuais: erro percentual médio absoluto (MAPE), erro percentual absoluto (APE), erro médio absoluto (MAE) e Coeficiente de Explicação (R^2).

Palavras-chave

Consumo Residencial Mensal de Energia Elétrica, ARIMA, Redes Neurais Artificiais, Amortecimento Exponencial, Combinação Linear de Modelos, Combinação Geométrica de Modelos.

Abstract

Teixeira Júnior, Luiz Albino; Souza, Reinaldo Castro (Advisor). **Combinação Geométrica de Métodos Preditivos; Aplicação à Previsão de Curto Prazo de Consumo Residencial Mensal de Energia Elétrica**. Rio de Janeiro, 2009. 115p. MSc Dissertation – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

This work proposes the application of some methods of time series forecast geometrically combined to the point and interval monthly projection of the residential consumption electricity load for a twelve-step-ahead horizon. Were used three techniques: ARIMA, Exponential Smoothing and Artificial Neural Networks, as well as linear and geometric combination models. The performance of the models was compared using the residual statistics: Mean Absolute Percentage Error (MAPE), Absolute Percentage Error (APE), Mean Absolute Error (MAE) and Explanation Coefficient.

Keywords

Residential Consumption Electricity Load, ARIMA, Exponential Smoothing, Artificial Neural Networks, Linear Combination of Models, Geometric Combination of Models.

Sumário

Lista de Figuras

Lista de Gráficos

Lista de Tabelas

1
INTRODUÇÃO

1.1
Contextualização e Definição da Pesquisa

Com a implementação do novo modelo do Setor Elétrico Brasileiro, a adequação às novas regras e às legislações tornou-se necessária por partes dos agentes. Entre estes, encontram-se as concessionárias de distribuição de energia elétrica, que interagem diretamente com os consumidores finais.

A maior parte das distribuidoras atua sob forma de concessão e são administradas por empresas do setor privado. A Agência Nacional de Energia Elétrica (ANEEL) possui a missão regular e fiscalizar o mercado de energia elétrica, de forma a fornecer condições de equilíbrio, tanto aos agentes quanto à sociedade.

À concessionária de distribuição, faz-se necessário ter um planejamento adequado para expansão e reforço de seu sistema, de acordo com o crescimento da demanda de seu mercado cativo. Logo, torna-se necessário que sejam estudadas as condições de viabilidade de custo e, sobretudo, a oferta da potência futura. Assim sendo, a projeção do consumo residencial mensal de energia elétrica pressupõe a previsão dos níveis de consumo. Para tal, normalmente são utilizadas séries históricas e informações sobre os micro e macro ambientes, assim como outras variáveis importantes que possam afetá-lo.

A projeção do consumo residencial apresenta substancial relevância ao setor, desde o planejamento e controle, até a execução de várias outras ações. Portanto, estes resultados são uma parte importante da base de informações para definição de modificações, tais como: nível de investimento em infra-estrutura, adequação dos graus de necessidades de capital, gestão dos níveis do reservatório, estoques, capacidade.

Os modelos utilizam séries temporais mensais de energia faturada que devem ser desagregadas por classes de consumo. Nesta dissertação, especificamente, trabalhou-se com a classe de consumo residencial de energia

elétrica de uma concessionária distribuidora brasileira. Esta classe corresponde à maior parcela de seu mercado, em termos de número de consumidores.

Em face aos modelos preditivos vigentes na literatura aplicados ao setor, buscam-se metodologias cada vez mais acuradas no que concerne à projeção de séries de tempo. Inúmeras propostas vêm sendo utilizadas e pesquisadas a fim de alcançar esta finalidade. Dentre estas, encontram-se as principais classes de modelos estatísticos clássicos, que podem ser divididas em três categorias básicas: univariada (Amortecimento Exponencial, Box & Jenkins), causal (Função de Transferência, Regressão Dinâmica) e multivariada (Vetores Auto-Regressivos). Outra opção é a utilização dos métodos de Inteligência Artificial (Redes Neurais Artificiais, Lógica *Fuzzy*).

Neste contexto, a combinação linear de modelos preditivos, proposta inicialmente por GRANGER e BATES (1969), é uma técnica bastante difundida e uma alternativa relevante à modelagem de previsões de séries temporais. Este método, porém, implica a necessidade que sejam estimados pesos adaptativos que ponderem linearmente as previsões dos modelos que o compõe, de forma a minimizar a variância (métrica comumente utilizada) dos resíduos combinados. Em alguns casos, podem ser assumidos pesos iguais, não necessitando assim de estimá-los. Este é conhecido como modelo das médias simples, podendo ser visto como um caso especial do método linear.

Por sua vez, a combinação geométrica de métodos preditivos se apresenta como uma alternativa às usuais, inclusive à linear. Segundo MUBWANDARIKWA e FARIA (2008), a combinação de densidades da família exponencial gera uma densidade preditiva também desta família, ou seja, fortemente unimodal, de particular interesse em previsão temporal e à tomada de decisão. Em particular, no caso da distribuição *t*-student, a combinação geométrica pode ser unimodal, enquanto a linear é tipicamente multimodal.

Ainda segundo os autores, o método de combinação geométrica, aplicado aos modelos bayesianos, obteve previsões mais acuradas, em relação às projeções do modelo de combinação linear e às dos individuais, sobre a maior parte das estatísticas de acurácia (considerando a série temporal utilizada em seu estudo de caso).

Diante dos resultados de MUBWANDARIKWA e FARIA (2008), a proposta central é mostrar que a combinação geométrica dos modelos ARIMA, Amortecimento Exponencial e de Redes Neurais Artificiais é uma alternativa factível e eficiente para modelagem de séries de tempo e, em particular, à projeção de consumo residencial mensal de energia elétrica.

Outro ponto importante é a estimação dos pesos adaptativos ótimos das combinações, adotando a minimização da função MAPE (*mean absolute percentual error*), dentro da amostra, e não a minimização do MSE (*mean square error*). Esta situação foi interpretada como um problema de otimização simples, sendo resolvido com a utilização da ferramenta de otimização *solver*.

Portanto, os modelos individuais (que compõem o modelo combinado), o de médias simples e o de combinação linear, foram utilizados como métodos comparativos, em relação ao modelo de combinação geométrica.

1.2
Objetivos

Esta pesquisa objetivou desenvolver e testar empiricamente um algoritmo eficiente e viável para projetar a série de consumo residencial de energia elétrica de uma concessionária distribuidora, num horizonte de curto prazo, doze meses à frente.

Adicionado a isso, deseja-se fornecer maior flexibilidade para modelagem de séries temporais, com um método que, em inúmeras situações, pode fornecer melhores resultados.

A estimação do comportamento da série histórica foi realizada considerando seus valores passados e o presente, ou seja, foram utilizados modelos univariados.

Desse modo, previram-se valores futuros pontuais e intervalares, inferindo acerca da metodologia proposta.

1.3
Justificativa e Relevância

A proposta é inovadora, eficiente e atual, quanto ao algoritmo proposto. A utilização de modelos estatísticos clássicos e de Inteligência Artificial combinados geometricamente é inexistente na literatura.

Outro aspecto importante diz respeito à estimação dos pesos adaptativos. A abordagem utilizada é também inédita. Apesar de simples, se mostrou eficiente e pode ser usada em outras abordagens, como por exemplo, a combinação linear de métodos.

Este estudo contribuirá ao arcabouço teórico dos modelos preditivos, fornecendo uma abordagem alternativa factível de modo a proporcionar maior flexibilidade, contribuindo em inúmeras outras aplicações. Uma gama de possibilidades e abordagens podem ser utilizadas a partir do método em inúmeras aplicações.

1.4
Organização dos Capítulos

A presente dissertação está organizada em cinco capítulos:

- Capítulo 2: Descrição de alguns conceitos básicos necessários à modelagem, bem como os dois modelos estatísticos, ARIMA e Amortecimento Exponencial, e o de inteligência artificial, Redes Neurais Artificiais, utilizados à projeção de consumo residencial mensal de energia elétrica;
- Capítulo 3: São citados resultados importantes de alguns autores, no contexto de combinação de modelos. Três métodos são apresentados: Combinação Linear, Média Simples e Combinação Geométrica, sendo o último, a proposta da dissertação. Por fim, abordou-se a questão dos pesos adaptativos e constantes de ajuste e intervalos de confiança concernente às combinações, justificando sua escolha e citando outras referências e métodos de ponderação;

- Capítulo 4: Explicita os resultados dos métodos individuais e dos combinados, inferindo acerca dos mesmos;
- Capítulo 5: Conclusão sobre o método proposto (Combinação Geométrico de Modelos), em relação aos outros modelos utilizados;
- Apêndice: Contêm informações complementares acerca dos métodos utilizados.

2
MODELOS UNIVARIADOS

Neste capítulo, são apresentados alguns conceitos atinentes às séries temporais, como também aos três modelos univariados consagrados na modelagem de séries de tempo. Assim, definiram-se os métodos: Amortecimento Exponencial, Box & Jenkins e Redes Neurais Artificiais.

2.1
Conceitos Básicos

2.1.1
Conceito de Processos Estocásticos e de Séries Temporais

O conceito de série temporal está intimamente ligado ao de processo estocástico (PE). O PE é um mecanismo probabilístico gerador de dados cujo comportamento não pode ser descrito por uma função determinística, e sim estocástica. Seu comportamento futuro, portanto, é estudado somente em termos probabilísticos. Formalmente, é definido como uma função aleatória, y_t, indexada ao tempo, onde a cada instante *t*, o valor de y_t é uma variável aleatória.

Assim, o conjunto de valores $\{y_t,\ t \in T\}$ é chamado de espaço de estados; *t*, de parâmetro; *T*, de espaço paramétrico ou conjunto de índices e os valores de y_t, de estados do processo estocástico no instante *t*. Portanto, um PE pode ser classificado quanto ao espaço paramétrico e ao de estados. (MORETIN, 2006)

Quanto ao espaço paramétrico, um processo estocástico é classificado como contínuo, se o conjunto *T* assume valores não contáveis (contínuos). Por outro lado, se os mesmos assumem valores contáveis (discretos), tem-se um PE discreto. (MORETIN, 2006)

O espaço de estados de um PE, por sua vez, pode ser contínuo ou discreto, caso as variáveis aleatórias sejam, respectivamente, contínuas ou discretas. Dessa maneira, os modelos para os processos estocásticos são específicos, de acordo

com a combinação entre os dois tipos de espaços: paramétrico e estados. (MORETIN, 2006)

A estrutura probabilística de um processo estocástico pode ser definida através da especificação da distribuição de probabilidade conjunta. No entanto, é difícil especificá-la. Em termos práticos, caracteriza-se um PE por um modelo, a partir do qual se torna possível obter a evolução de seus momentos principais (média, variância, covariância), possibilitando, por exemplo, a realização de previsões pontuais e intervalares de determinada série de tempo.

Uma série temporal é formalmente definida como uma realização de um processo estocástico.

De acordo com SOUZA & CAMARGO (2004), as séries de tempo podem classificadas como discretas, contínuas, determinísticas, estocásticas, multivariadas e / ou multidimensionais. Assim, um processo estocástico pode ser interpretado como uma família de trajetórias (ou realizações) de uma variável aleatória. O conjunto com todas as trajetórias é denominado *Ensemble*. Na prática, é observada uma única série temporal e, por seguinte, estimado um possível processo estocástico (modelo) que a gerou, obtendo-se estimativas que o caracterizam.

No estudo de séries temporais, três etapas são básicas:

- Análise,
- Modelagem e
- Previsão.

Salienta-se que a aplicação de técnicas e modelos, em qualquer destes três estágios, permite obter informações estruturadas à tomada de decisão.

2.1.2
Análise de Séries Temporais

A análise de séries temporais consiste na utilização de ferramentas estatísticas a fim de revelar suas características relevantes, tais como: dependência, estacionariedade, periodicidade, sazonalidade, volatilidade,

distribuição incondicional. Assim, é possível obter subsídios para a especificação da(s) classe(s) de modelos adequada(s) à sua modelagem.

De acordo com SOUZA & CAMARGO (1996), dois domínios podem ser especificados em sua modelagem:

- Domínio no tempo; e
- Domínio na frequência.

No primeiro, considera-se a evolução no tempo do processo, isto é, o interesse se concentra na dimensão de um determinado evento no instante *t*. Para tal, são utilizadas as funções de autocovariância e autocorrelação.

A função que descreve a dependência linear entre os valores de uma série temporal é denotada como função de autocorrelação (ACF).

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\text{cov}(y_t, y_{t-k})}{\sigma_{y_t}\sigma_{y_{t-k}}} \tag{2.1}$$

Assim definida, a função autocorrelação amostral é dada por (2.2).

$$r_k = \frac{\sum_{t=k+1}^{n} (y_t - \overline{y})\ (y_{t-k} - \overline{y})}{\sum_{t=1}^{n} (y_t - \overline{y})^2} \tag{2.2}$$

Pode estender-se o conceito de autocorrelação. Na mensuração da autocorrelação entre duas observações seriais y_t e y_{t-k}, a dependência dos termos intermediários, y_{t-1}, y_{t-2}, ..., y_{t+k-1}, pode ser eliminada. A estatística que faz este cálculo é conhecida como função de autocorrelação parcial (PACF), representada por: $cor\,(y_t, y_{t-k} / y_T, ..., y_1)$. (SOUZA & CAMARGO, 1996)

Segundo SOUZA & CAMARGO (2004), a função de autocorrelação parcial amostral é definida pela equação (2.3).

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^{k} \hat{\phi}_{kj}\ \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^{k} \hat{\phi}_{kj}\ \hat{\rho}_j} \tag{2.3}$$

Sendo: $\hat{\phi}_{k+1,j} = \hat{\phi}_{kj} - \hat{\phi}_{k+1,k+1}\ \hat{\phi}_{k,k+1-j}$, onde $j = 1, ..., k$.

Na segunda abordagem, a análise é realizada considerando a frequência com que os eventos ocorrem, dado que os componentes harmônicos da série temporal supracitada sejam relevantes. A densidade espectral é utilizada para a caracterização das propriedades de frequência do processo estocástico gerador.

De acordo com SOUZA & CAMARGO (2004), considerando um processo estocástico estacionário y_t, com função de autocovariância finita, espectro é definido na equação (2.4).

$$f(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \rho(k) \exp(-iwk), \text{ onde: } -\pi \leq w \leq \pi \text{ ; e} \quad (2.4)$$

$$\rho(k) = \sum_{k=-\infty}^{\infty} \exp(iwk) f(w), \text{ onde: } k = 0, \pm 1, \pm 2, ... \quad (2.5)$$

Nota-se que a função de densidade espectral normalizada é a transformada de Fourier da função de autocorrelação e representa a contribuição das componentes de frequência presentes no processo para com a variância (potência total).

Na prática, o espectro tem de ser estimado, onde estimador mais utilizado é o periodograma com janelas espectrais.

2.1.3
Modelagem de Séries Temporais

A modelagem de séries temporais consiste em encontrar um modelo que melhor descreva sua estrutura de dependência observada. De acordo com a abordagem e as características dos dados disponíveis, são selecionados os modelos e / ou as classes que podem ser usados.

Os modelos estatísticos aplicados às séries temporais podem ser classificados em três categorias:

- Univariados: utilizam somente os valores passados da própria série para explicar os futuros. Podem citar-se: Métodos de Amortecimento Exponencial, Box & Jenkins;
- Causais: conhecidos também como modelos de função de transferência usam os valores defasados da série temporal endógena, das exógenas e do

erro presente. Podem citar-se: Modelos de Regressão, Regressão Dinâmica e Função de Transferência; e

- Multivariados: nestes, tem-se um vetor endógeno, e não um escalar, sendo possível projetar várias séries temporais, simultaneamente. Podem citar-se: Vetores Auto-Regressivos, GARCH Multivariados, Vetores de Correção de Erros.

Não obstante, os modelos de inteligência artificial têm o objetivo de desenvolver algoritmos que sejam capazes de reproduzir tarefas simples realizadas por seres humanos e que necessitam de cognição como raciocínio, aprendizagem e auto-aperfeiçoamento (HAYKIN, 2001). Estes métodos possuem a capacidade de realização de tarefas realizadas por qualquer categoria de modelo estatístico utilizada. Assim, citam-se: as Redes Neurais Artificiais, a Lógica Fuzzy, os Algoritmos Genéticos.

2.1.4
Previsão de Séries Temporais

Uma vez construído o modelo, previsões probabilísticas podem ser efetuadas. Com base em informações passadas e atuais, são estimadas previsões pontuais e intervalares.

A notação de previsão normalmente é dada por $\hat{y}_T(h)$, definida, na Estatística, como a esperança condicional da variável aleatória y_{T+h}, para (*h=1, 2, 3...*). Assim: $E[y_{T+h} / y_0,...,y_T] = \hat{y}_T(h)$, onde $(y_0,...,y_T)$ é a amostra observada, e $h \in \mathbb{N}^*$.

Na inteligência artificial, representa-se apenas um valor como previsão. Uma previsão quantitativa, portanto, pode ser caracterizada através de um número pontual projetado *k* passos à frente por um modelo, associado uma medida de incerteza.

2.1.5
Estacionariedade de um Processo Estocástico

Um processo estocástico $\{ y_t \}$ é classificado como fracamente estacionário (ou de 2º ordem, ou amplo, ou em sentido lato), se alguns de seus momentos incondicionais (média, variância e autocovariância) permanecem inalterados ao longo do tempo.

Consoante ALEXANDER (2004), tal é dito de covariância estacionária quando as propriedades (2.6), (2.7) e (2.8).

- $E[Z_t] = E[Z_{t+k}] = \mu$, para $k \in \mathbb{Z}$; (2.6)
- $E[(Z_t - \mu)^2] = E[(Z_{t+k} - \mu)^2] = \gamma(0) = \sigma_{z_t}^2$, para $k = 0$ e $\forall\ t$; e (2.7)
- $E[(Z_t - \mu)\ (Z_{t+k} - \mu)] = E[(Z_{t+m} - \mu)\ (Z_{t+k+m} - \mu)] = \gamma(k)$, para $k \in \mathbb{Z}^*$, $m \in \mathbb{Z}$ e $\forall\ t$. (2.8)

Onde a média e a variância (incondicionais) do processo são constantes no tempo e a estrutura de dependência linear $\gamma(k)$ depende apenas da distância entre os períodos, diminuindo à medida que cresce.

Outro tipo de processo estocástico é o denominado fortemente estacionário (ou estrito), a característica que permanece invariante é a distribuição de probabilidade conjunta, dependendo apenas da defasagem. Isto é, sua estrutura probabilística permanece invariante no tempo:

$$F(y_{t_1},...,y_{t_k}) = F(y_{t_1+m},...,y_{t_k+m}),\ \forall\ t_1,...,t_k, m \in \mathbb{N}. \quad (2.9)$$

Em particular, se um processo estocástico gaussiano é fracamente estacionário, o mesmo é também estritamente estacionário, visto que a distribuição normal é determinada unicamente pelos termos do primeiro e segundo momentos.

Na prática, as distribuições são desconhecidas, de modo que a se adota a estacionariedade de 2º ordem, por ser mais branda. Isso é relevante, visto que, por definição, as propriedades estatísticas como, média, variância e autocorrelação, observadas no passado, são as mesmas no futuro, o que possibilita a estimação de características importantes do processo, de forma, muitas vezes, bastante simples - o que não ocorre no caso de séries não estacionárias.

Alguns testes estatísticos são utilizados para verificação de estacionariedade: Dikey-Fuller, Dikey-Fuller Aumentado, Phillips-Perron.

Mais detalhes sobre os testes de raiz unitária de Dikey-Fuller e Dikey-Fuller Aumentado (usado na pesquisa) podem ser verificados no Apêndice B.

2.1.6
Processo Estocástico Ruído Branco

Seja uma seqüência $\{a_t, t \in \mathbb{N}\}$ de variáveis aleatórias independentes e identicamente distribuídas, com média igual a zero e variância constante.

Tais variáveis são chamadas de choques aleatórios e a seqüência, processo estocástico ruído branco. De acordo com SOUZA & CAMARGO (2004), têm-se:

- $a_t \sim distr.(0,\ \sigma_a^2), \forall t$; e (2.10)
- $E[a_t,\ a_{t-k}] = E[a_{t+m},\ a_{t-k+m}],\ k \in \mathbb{Z}^*$ e $m \in \mathbb{Z}^*$,
- Autocorrelação: $\rho_k = 1$, para $k = 0$, e $\rho_k = 0$, para $k \neq 0$;
- Autocorrelação parcial: $\phi_{kk} = 1$, para $k = 0$, e $\phi_{kk} = 0$, para $k \neq 0$;
- Densidade espectral: $f(w) = \frac{1}{2\pi}$; e
- Densidade espectral (não normalizada): $h(w) = \frac{\sigma_a^{\ 2}}{2\pi}$.

2.1.7
Processos Estocásticos não-Estacionários Homogêneos

Embora a teoria mostre conveniência prática no uso de séries temporais estacionárias, poucas possuem esta propriedade. Porém, em muitos casos, existe a possibilidade de torná-la, aplicando sucessivas diferenças. Quando isso é possível, tem-se um processo não estacionário homogêneo ou integrado.

O procedimento, porém, exclui alguns processos de comportamento explosivos e / ou altamente não lineares.

O número de vezes que a série original deve ser diferenciada para tornar-se estacionária é chamado de ordem de homogeniedade *n* (ou integrada de ordem *n* ou, simplesmente, *I(n)*).

Assim, se y_t é integrada de primeira ordem, significa que $W_t = y_t - y_{t-1} = \nabla y_t$ é estacionário.

Onde:

$\nabla = (1 - B)$, é o operador de atraso (SOUZA & CAMARGO, 2004).

No apêndice A, encontram-se mais detalhes sobre outros tipos operadores e definições.

2.2
Modelos de Amortecimento Exponencial

Inicialmente foi desenvolvida por Robert G. Brown, no período em que trabalhava para a marinha norte-americana, durante a II guerra Mundial, a metodologia de amortecimento exponencial (MAE) ganhou destaque realmente em 1970.

Os modelos de amortecimento exponencial são classificados como sendo automáticos e de validade local. Baseiam-se na premissa de que as observações mais recentes contêm mais informações que as mais antigas e, por isso, o peso decresce exponencialmente, à medida que a observação torna-se mais antiga - diferentemente do modelo de médias móveis, em que todas as observações têm o mesmo nível de importância, limitando-o. (SOUZA, 1983).

Para um melhor entendimento do MAE, faz-se necessário descrever, de forma sucinta, a evolução de dois métodos que o precedem, considerando uma série temporal observada $(y_0, ..., y_T)$.

- Método Ingênuo (NAIVE):

  Também conhecido como *random walk* (passeio aleatório), é o mais simples para realização de previsão de uma variável, onde o previsto é

igual ao último observado (independente do número de passos à frente). (MONTGOMERY, 1990)

Portanto, o método possui a seguinte equação estocástica:

$$y_t = y_{t-1} + \varepsilon_t \text{, onde } \varepsilon_t \sim N(0, \sigma^2),\ \forall\ t \in \mathbb{N} \tag{2.11}$$

Denotando a previsão *h* passos à frente, tem-se $\hat{y}_T(h)$. Pode-se escrever a equação de previsão conforme (2.12).

$$E[y_{T+h} + \varepsilon_{t+h} / y_0, ..., y_T] = \hat{y}_T(h) = y_T \text{, onde } h = 1, 2, ... \tag{2.12}$$

O método ingênuo, por vezes, é utilizado como critério de comparação, através da estatística GMRAE (*Geometric Mean of the Relative Absolute Error*).

- Média Móvel:

No método de média móvel, cada nova observação é calculada pela média de *n* (tamanho da janela) pontos anteriores. (MONTGOMERY, 1990)

$$y_t = \frac{1}{n}(y_{t-1} + ... + y_{t-n}) \tag{2.13}$$

O tamanho da janela é o parâmetro a ser ajustado, sendo necessário buscar um equilíbrio, pois quanto maior o tamanho da janela, mais suave o comportamento da média e, conseqüentemente, mais imune a ruídos e movimentos curtos. Caso contrário, reage de maneira muito lenta às mudanças significativas.

Assim, pode-se ter um modelo de média móvel simples, no qual a série é caracterizada localmente por seu nível, sendo as mudanças ocorridas de caráter muito lento.

Existe a média móvel dupla, onde a série possui tendência aditiva e que a média sofre uma alteração linear ao longo do tempo.

A média móvel tripla é utilizada em séries com tendência quadrática, sendo seu efeito multiplicativo.

De acordo com MONTGOMERY (1990), têm-se:

- O modelo de média móvel simples.

$$y_t = y_{t-1} + \varepsilon_t \text{, onde } \varepsilon_t \sim N(0,\sigma^2),\ \forall\ t \in \mathbb{N} \quad (2.14)$$

A equação de previsão *h* passos à frente, feita no instante *T* dada por:

$$E[y_{T+\tau} + \varepsilon_{t+h} / y_0,...,y_T] = \hat{y}_T(h) = \hat{a}(T) \quad (2.15)$$

Considerando $\hat{a}(T)$, tem-se:

$$\hat{a}(T) = M_T = \frac{1}{N}(y_{t-1} + ... + y_{t-N+1}) \quad (2.16)$$

- O modelo para a média móvel dupla.

$$y_t = a_1 + a_2\, t + \varepsilon_t \text{, onde } \varepsilon_t \sim N(0,\sigma^2),\ \forall\ t \in \mathbb{N} \quad (2.17)$$

Sendo a previsão, para *h* passos à frente, definida por:

$$\hat{y}_T(h) = \hat{a}_1(T) + \hat{a}_2(T)\ [T+h] \text{, onde } h = 1,2,... \quad (2.18)$$

As estimativas são calculadas por:

$$\hat{a}_1(T) = 2M_T - M_T^{[2]} - \hat{a}_1(T)T \quad (2.19)$$

$$\hat{a}_2(T) = \frac{2}{N-1}[M_T - M_T^{[2]}] \quad (2.20)$$

Sendo:

$M_T = \frac{1}{N}(y_{t-1} + ... + y_{t-N+1})$, onde M_T é a média móvel simples de tamanho *N*.

$M_T^{[2]} = \frac{1}{N}(M_{t-1} + ... + M_{t-N+1})$, onde M_T é a média móvel dupla de tamanho *N*.

- O modelo estocástico para média móvel tripla.

$$y_t = a_1 + a_2\, t + a_3\, t^2 + \varepsilon_t \text{, onde } \varepsilon_t \sim N(0,\sigma^2),\ \forall\ t \in \mathbb{N} \quad (2.21)$$

A previsão, para *h* passos à frente, é dada por:

$$\hat{y}_T(h) = \hat{a}_1(T) + \hat{a}_2(T)\,[T+h] + \hat{a}_3(T)[T+h]^2, \text{ onde } h = 1,2,... \qquad (2.22)$$

Sendo, $\hat{a}_1(T)$, $\hat{a}_2(T)$ e $\hat{a}_3(T)$ os estimadores dos parâmetros no instante *T*, calculados em função das médias móveis de tamanho *N*: simples (M_T), dupla ($M_T^{[2]}$) e tripla ($M_T^{[3]}$).

O modelo de previsão para dados com sazonalidade, utilizando médias móveis pode ser escrito por:

$$\hat{y}_T(h) = \frac{1}{N} E[y_{t+h-S} + y_{t+h-2S} + ... + y_{t+h-nS}] \qquad (2.23)$$

Mais detalhes acerca estimadores descritos acima, bem como os modelos com componente sazonais de médias móveis, podem verificados em MONTGOMERY (1990) e MORETIN (2006). O fluxograma da figura 2.1 descreve a evolução dos modelos.

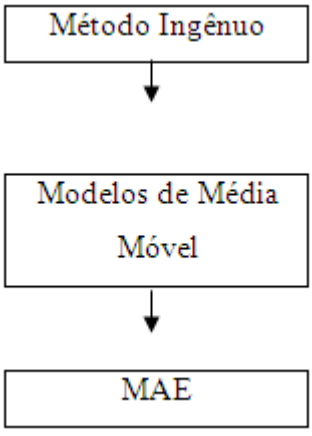

Figura 2.1 - Fluxograma da Evolução dos Métodos

2.2.1
Modelos de Amortecimento Exponencial de Brown

Os modelos automáticos não sazonais de Brown podem ser divididos em três categorias: simples, linear e quadrático. Em relação às médias móveis, a

diferença consiste em atribuir diferentes relevâncias aos dados. Portanto, pode-se dizer que trabalha com médias móveis ponderadas.

Segundo MORETIM (2006), definem-se:

- Modelos simples: $y_t = a_1 + \varepsilon_t$, onde $\varepsilon_t \sim N(0,\sigma^2),\ \forall\ t \in \mathbb{N}$; (2.24)

$$\hat{a}(T) = \alpha\ y_T + (1-\alpha)\ y_{T-1} = \alpha\ y_T + (1-\alpha)\ \hat{a}(T-1)$$

- Modelo Linear: $y_t = a_1 + a_2\ t + \varepsilon_t$, onde $\varepsilon_t \sim N(0,\sigma^2),\ \forall\ t \in \mathbb{N}$; e (2.25)

$$\hat{a}_1(T) = 2M_T - M_T^{[2]}$$

$$\hat{a}_2(T) = \frac{\alpha}{1-\alpha}\ [M_T - M_T^{[2]}]$$

$$M_T = \alpha\ y_T + (1-\alpha)\ M_{T-1}$$

$$M_T^{[2]} = \alpha\ M_T + (1-\alpha)\ M_{T-1}^{[2]}$$

- Modelo quadrático: $y_t = a_1 + a_2\ t + a_3\ t^2 + \varepsilon_t$, onde $\varepsilon_t \sim N(0,\sigma^2),\ \forall t \in \mathbb{N}$ (2.26)

$$\hat{a}_1(T) = 3M_T - 3M_T^{[2]} + M_T^{[3]}$$

$$\hat{a}_1(T) = 3M_T - 3M_T^{[2]} + M_T^{[3]}$$

$$\hat{a}_2(T) = \frac{\alpha}{1-\alpha}\ [M_T - 2M_T^{[2]} + M_T^{[3]}]$$

$$\hat{a}_3(T) = \frac{\alpha}{2(1-\alpha)^2}\ [(6-\alpha)\ M_T - 2(5-\alpha)\ M_T^{[2]} + (4-3\alpha)\ M_T^{[3]}]$$

$$M_T = \alpha\ y_T + (1-\alpha)\ M_{T-1}$$

$$M_T^{[2]} = \alpha\ M_T + (1-\alpha)\ M_{T-1}^{[2]}$$

$$M_T^{[3]} = \alpha\ M_T^{[2]} + (1-\alpha)\ M_{T-1}^{[3]}$$

A constante de amortecimento ($0 < \alpha < 1$) é tal que minimiza a soma ordinária dos resíduos. Ou seja, tem-se a estimação via OLS (*ordinary least square*). (Para mais detalhes, MORETIM (2006)).

2.2.2
Modelos de Amortecimento Exponencial de Holt

Possui a mesma formulação que o anterior, porém o diferencial está na atualização dos parâmetros.

Enquanto o de Brown usa o mesmo hiperparâmetro (α) para atualização das estimativas, este utiliza dois hiperparâmetro, um para nível e outro para tendência. O modelo é usado em séries com tendência linear, portanto.

Segundo MORETIM (2006), tem-se:

$$y_t = a_1 + a_1\, t + \varepsilon_t \tag{2.27}$$

A equação de previsão, considerando a origem deslocada para *T*, é dada por:

$$\hat{y}_T(h) = \hat{a}_1(T) + \hat{a}_2(T)\; h \tag{2.28}$$

No caso da equação (2.28), os estimadores $\hat{a}_1(T)$ e $\hat{a}_2(T)$ são calculados em função das constantes amortecidas α e β, respectivamente.

Isso é dado pela equação (2.29):

$$\hat{a}_1(T) = \alpha\; y_T + (1-\alpha)\; [\hat{a}_1(T-1) + \hat{a}_2(T-1)] \tag{2.29}$$

$$\hat{a}_2(T) = \beta\; [\hat{a}_1(T) + \hat{a}_1(T-1)] + (1-\beta)\; [\hat{a}_2(T-1)]$$

Uma extensão deste método é conhecida como *damped trend*. Assim, mais um hiperparâmetro é considerado, φ. A equação de previsão, portanto, é dada por (2.30):

$$\hat{y}_T(h) = \hat{a}_1(T) + [\sum_{j=1}^{h} \varphi^{j-1}]\; \hat{a}_2(T) \tag{2.30}$$

Assim, se φ pertencer ao intervalo $]0,1[$, a tendência será amortecida (*damped*), consequentemente a previsão se aproximará da assíntota: $\hat{a}_1(T) + \frac{\varphi}{1-\varphi}\; \hat{a}_2(T)$. Caso contrário, se φ>1, a tendência teria um comportamento explosivo.

Mais detalhes acerca desta abordagem podem ser encontrados em TAYLOR (2003b). Sobre o método de Holt, em MORETIM (2006).

2.2.3
Modelos de Amortecimento de Holt-Winters

Diferentemente dos métodos de Brown e Holt, a metodologia de Holt-Winters incorpora uma componente de sazonalidade em sua estrutura. São consideradas, então, três componentes e, consequentemente, três constantes de amortecimento.

O modelo aditivo é recomendado às séries temporais homocedásticas; o multiplicativo, às heterocedásticas.

2.2.3.1
Modelo Aditivo de Amortecimento de Holt-Winters

Supondo que a série $\{y_t\}$ possua somente um padrão sazonal e que esse padrão não aumente juntamente com o nível da série. Formalmente:

$$y_t = \mu(t) + \rho_t + \varepsilon_t \tag{2.31}$$

Onde $\mu(t) = a_1 + a_2 t$; sendo a_1, a componente do nível e a_2, a da tendência. O ρ_t é o fator sazonal aditivo e ε_t, o componente aleatória.

Segundo MORETIM (2006), os fatores sazonais aditivos são definidos de forma que sua soma seja igual a zero, tendo como restrição:

$$\sum_{j=1}^{k} c_j(t) = 0. \tag{2.32}$$

A equação de previsão é dada por:

$$\hat{y}_T(h) = \hat{a}_1(T) + \hat{a}_2(T)(T+h) + \hat{\rho}_{T+h}(T) \tag{2.33}$$

Assumi-se que as estimativas da componente de tendência e dos fatores sazonais do modelo aditivo, apresentado pelas equações (3.34), sejam denotadas por: $\hat{a}_1(T)$, $\hat{a}_2(T)$ e $\hat{\rho}_T$, respectivamente.

Nível: $\hat{a}_1(T) = \alpha\,[y_T - \hat{\rho}_{m(T)}(T-1)] + (1-\alpha)\,[\hat{a}_1(T-1) + \hat{a}_2(T-1)]$; (2.34)

Tendência: $\hat{a}_2\ (T)\ =\ \beta\ [\hat{a}_1\ (T)-\hat{a}_1\ (T-1)]+(1-\beta)\ [\hat{a}_2\ (T-1)]$; e

Sazonalidade: $\hat{\rho}_{m(T)}\ (T)=\ \gamma\ [y_T-\hat{a}_1\ (T)]+(1-\gamma)\ [\hat{\rho}_{m(T)}(T-1)]$.

Onde α, β e γ são constantes amortecidas pertencentes ao intervalo $[0,1]$. O modelo, porém, necessita que os parâmetros iniciais sejam estimados, a saber: $\hat{a}_1(0)$, $\hat{a}_2(0)$ e $\rho_i(0)$ $i=1,...,L$.

A estimativa desses parâmetros iniciais pode ser obtida através dos dados históricos. (Para mais detalhes, MONTGOMERY, 1990)

2.2.3.2
Modelo Multiplicativo de Amortecimento de Holt-Winters

A tendência continua possuindo uma formulação aditiva. O modelo é capaz de incorporar tanto a tendência linear quanto o efeito sazonal.

$$y_t=\ (a_1+a_2\ t)\ _x\ \rho_t+\varepsilon_t. \tag{2.35}$$

A definição dos parâmetros é a mesma do modelo aditivo e L, o tamanho do ciclo sazonal.

De acordo com MORETIM (2006), os fatores sazonais do modelo multiplicativo possuem restrição

$$\sum_{j=1}^{L}\rho_j\ (t)\ =\ L \tag{2.36}$$

A equação de previsão é dada por:

$$\hat{y}_T\ (h)\ =\ [\hat{a}_1\ (T)+a_2\ (T)\ (T+h)]\ _x\ \hat{\rho}_{m(T+h)}\ (T)+\varepsilon_t \tag{2.37}$$

O procedimento de atualização é dado pelas equações (2.38):

$$\text{Nível: } \hat{a}_1(T)\ =\ \alpha\ (\frac{y_T}{\hat{\rho}_{m(T)}(T-1)})+(1-\alpha)\ [a_1\ (T-1)+a_2\ (T-1)] \tag{2.38}$$

Tendência: $\hat{a}_2(T)\ =\ \beta\ (a_1\ (T)-a_1\ (T-1))+(1-\beta)\ [a_2\ (T-1)]$.

Sazonalidade: $\hat{\rho}_{m(T)}\ (T)=\ \gamma\ [\frac{y_T}{\hat{a}_1(T)}]+(1-\gamma)\ \hat{\rho}_{m(T)}(T-1)\cdot$

Para os demais fatores: $\hat{\rho}_j\ (T)=\ \hat{\rho}_j\ (T),\ \ j\ =\ 1,2,...,L\ -\ \forall j\ \neq\ m(T)$.

Onde α, β e λ são constantes de amortecimento. O índice sazonal local $\hat{\rho}_{m(T)}(T)$ é estimado pela razão do valor observado e o nível local $a_1(T)$. A estimação dos valores iniciais: $\hat{a}_1(0)$, $\hat{a}_2(0)$ e $\rho_i(0)$, *i=1,..., L* é obtida através da amostra. (Para mais detalhes, MONTGOMERY, 1990).

2.3
Modelos de BOX & JENKINS

Os modelos ARIMA foram inicialmente formulados por BOX & JENKINS na década de 1970.

Com argumentos de estacionariedade e ergodicidade do processo estocástico subjacente, procura-se detectar o sistema probabilístico gerador da série temporal, através das informações nela contidas.

Adicionalmente, baseia-se na também premissa de que uma série temporal não estacionária pode ser modelada a partir de *d* diferenciações e da inclusão de um componente auto-regressivo integrado de médias móveis. O mesmo raciocínio vale à componente sazonal.

Segundo HILL, GRIFFITHS & JUDGE (1999), quanto ao método ARIMA:

> *Relacionam-se os valores correntes de uma variável somente com os seus valores passados e erros passados e correntes.*

A modelagem BOX & JENKINS assume dois princípios: parcimônia, que estabelece um modelo com o menor número de parâmetros, e, ciclo iterativo, estratégia de seleção de modelos. (MORETIM, 2006)

A metodologia tem como base a Teoria Geral de Sistemas Lineares, a qual supõe que a passagem de um ruído branco por um filtro de memória infinita gera um processo estacionário de segunda ordem. (SOUZA & CAMARGO, 2004)

Assim sendo, tem-se:

$$y_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots, \; k = 0, 1, \ldots \tag{2.39}$$

Em que:

$$\psi(B) = (1 + \psi_1 B^1 + \psi_2 B^2 + \ldots). \tag{2.40}$$

Assim sendo, tem-se um filtro linear com entrada a_t (ruído branco); saída, y_t (PE linear discreto) e uma função de transferência, $\psi_k(B)$.

Definindo $\tilde{y}_t = y_t - \mu$, tem-se que:

$$\tilde{y}_t = \psi(B)\, a_t \tag{2.41}$$

Conforme destacado, caso a sequência de pesos $\{\psi_k, \; k \geq 1\}$ for finita ou infinita e convergente, o filtro será estável, implicando estacionariedade do processo y_t. Neste caso, μ é a média do processo. Caso contrário, y_t é não-estacionário e μ não tem significado específico. A não ser como um ponto de referência para o nível num determinado instante. (SOUZA & CAMARGO, 2004)

De (2.41) se tem:

$E[y_t] = \mu + E\left(a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j}\right)$, e como $E[a_t] = 0$, $\forall\, t$, tem-se que $E[y_t] = \mu$, se a série $\sum_{j=1}^{\infty} \psi_j$ convergir.

Pode-se escrever $\tilde{y}_t$ de forma alternativa, como a soma ponderada de valores passados $\tilde{y}_{t-1}$, $\tilde{y}_{t-2}$, ..., mais um ruído branco:

$$\tilde{y}_t = \pi_1 \tilde{y}_{t-1} + \pi_2 \tilde{y}_{t-2} + \ldots + a_t. \tag{2.42}$$

Segue-se que:

$$\left(1 - \sum_{j=1}^{\infty} \pi_j B^j\right) \tilde{y}_t = a_t. \tag{2.43}$$

Ou: $\pi(B)\, \tilde{y}_t = a_t$. (2.44)

Onde:

$$\pi(B) = (1 - \pi_1 B^1 - \pi_2 B^2 - ...) . \quad (2.45)$$

Por seguinte (2.49) e (2.50), obtém-se:

$$\pi(B)\ \psi(B)\ a_t = a_t \quad (2.46)$$

De modo que:

$$\pi(B) = \psi^{-1}(B) \quad (2.47)$$

Esta relação pode ser usada para obter os pesos π_j em função dos pesos ψ_j e vice-versa. Um processo linear é estacionário se a série $\psi(B)$ convergir para $|B| \leq 1$ e invertível, se $|B| \leq 1$. (BOX, JENKINS & REINSEL (1994) provam este fato).

No concerne à função de transferência, $\psi(B)$, tem-se:

$$\psi_k(B) = \theta(B) / \phi(B) \quad (2.48)$$

Pois um polinômio infinito equivale a uma divisão de polinômios.

$$\phi(B)\ y_t = \theta(B)\ a_t ;$$

$$y_t = \phi(B)^{-1}\ \theta(B)\ a_t . \quad (2.49)$$

Onde:

$\phi(B) = (1 - \phi_1 B^1 - ... - \phi_p B^p)$: polinômio característico da parte auto-regressiva; e

$\theta(B) = (1 - \theta_1 B^1 - ... - \theta_q B^q)$: polinômio característico da parte média móvel.

Sendo $\phi(.)$ e $\theta(.)$ polinômios cujos graus são, respectivamente, os índices *p* e *q*. Consoante a notação de BOX & JENKINS, o modelo (2.49) é denominado ARMA (*p*,*q*), um caso particular do método.

Figura 2.2 - Esquema Ilustrativo do Modelo ARMA (*p,q*) da Equação (2.49)

Quando não satisfeitos os pressupostos de normalidade e homocedasticidade da série temporal, é feita sua transformação com o intuito de obter tais propriedades nos dados transformados. BOX & JENKINS sugerem a classe de transformações BOX & COX.

SOUZA & CAMARGO (2004) destacam que nem sempre é possível estabilizar a variância e obter normalidade, simultaneamente.

De acordo com SOUZA (1981), *apud in* SOUZA & CAMARGO (2004):

> *A normalidade é o objetivo principal, podendo levar à variância constante. O inverso também é possível, isto é, estabilizar a variância, sendo que a normalidade pode ser obtida aproximadamente.*

2.3.1
Modelos Auto-Regressivos

Se em (2.50) $\pi_j = 0,\ j > p$, obtém-se uma modelo auto-regressivo de ordem *p* denotado por AR (*p*) ou ARMA (*p,0*).

AR (*p*): $\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \phi_2 \tilde{y}_{t-2} + ... + \varphi_p \tilde{y}_{t-p} + a_t$ (2.50)

Renomeando os pesos π_j para φ_j, conforme a notação usual de BOX & JENKINS. Assim, o polinômio característico é dado por:

$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 ... - \phi_p B^p)$: operador não sazonal auto-regressivo (AR). Ou:

$$\varphi(B)\ \tilde{y}_t\ =\ a_t \qquad (2.51)$$

A representação (2.51) é equivalente a uma média móvel infinita de ruídos brancos - MA (∞), visto que é trivialmente inversível. Porém, para condição de estacionariedade, requer-se $|\phi| < 1$ ou, de forma equivalente, que o polinômio $\varphi(B) = [\psi(B)]^{-1}$ tenha suas raízes reais e / ou complexas (a norma, no caso) fora do círculo unitário. Ou seja, que a série $\psi(B)$ convirja para $|B| \leq 1$. (SOUZA & CAMARGO, 2004)

2.3.2
Modelos Médias Móveis

Considere o processo linear (2.52) e suponha que $\psi_j = 0$, $j > q$; obtém-se um processo de médias móveis (*moving averege*) de ordem *q*, que é denotado por MA (*q*) ou ARMA (*0,q*).

$$y_t\ =\ \mu + a_t - \theta_1\ a_{t-1} - ... - \theta_q\ a_{t-q} \qquad (2.52)$$

Sendo, $\tilde{y}_t\ =\ y_t - \mu$, tem-se:

Modelo MA (*q*): $\tilde{y}_t\ =\ \theta(B)\ a_t \qquad (2.53)$

Onde:

$\theta\ (B)\ =\ (1 - \theta_1\ B - ... - \theta_q\ B^q)$: operador não sazonal de médias móveis.

A condição de inversibilidade do processo MA (*q*) é similar àquela de estacionariedade do processo AR (*p*), portanto $|\theta| < 1$, ou seja, a série $\pi(B) = [\theta(B)]^{-1}$ converge para $|B| \leq 1$. Isso é equivalente afirmar que zeros do polinômio $\theta(B)$ estão fora do círculo unitário. Assim, o processo MA (*q*) pode ser escrito como AR (∞). Estes modelos possuem uma representação trivialmente estacionária. (SOUZA & CAMARGO, 2004)

2.3.3
Modelos Auto-Regressivos e de Médias Móveis

O modelo ARMA (*p*,*q*) é escrito como na equação 2.49, onde ϕ (.) e θ (.) são polinômios de ordem *p* e *q*, respectivamente. É composto duas partes: uma auto-regressiva e de outra de médias móveis.

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \phi_2 \tilde{y}_{t-2} + ... + \phi_p \tilde{y}_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - ... - \theta_q a_{t-q} \quad (2.54)$$

Sendo $\varphi(B)$ e $\theta(B)$ os polinômios característicos auto-regressivos e de médias móveis, respectivamente, podem ser escritos de forma compacta:

$$\phi(B) \tilde{y}_t = \theta(B) a_t \quad . \quad (2.55)$$

Onde:

$\theta(B) = (1 - \theta_1 B^1 - ... - \theta_q B^q)$: operador não sazonal de médias móveis; e

$\phi(B) = (1 - \phi_1 B^1 - ... - \phi_p B^p)$: operador não sazonal auto-regressivo.

Para a verificação da invertibilidade e estacionariedade, analisam-se os polinômios $\theta(B)$ e $\phi(B)$, respectivamente.

Caso o correlograma da ACF e PACF apresente decaimento senoidal amortecido e / ou o exponencial, simultaneamente, uma sugestão plausível é o ARMA (*1*,*1*). (SARTORIS, 2003)

2.3.4
Modelos ARIMA

O modelo ARIMA (*p*,*d*,*q*) é uma extensão do ARMA (*p*,*q*). O índice *d* representa o número de vezes que a série original foi diferenciada. O resto é similar.

Portanto, o modelo é dado por ARIMA (*p*,*d*,*q*):

$$\phi(B) \nabla^d \tilde{y}_t = \theta(B) a_t \quad (2.56)$$

Onde:

$\nabla^d = (1-B)^d$: operador diferença não sazonal de ordem *d*.

Se $w_t = \nabla^d z_t$ for estacionária, é possível representar por um modelo ARMA, ou seja, $\varphi(B)\ w_t = \theta(B)\ a_t$. Logo, w_t é uma diferença de z_t, então esta última é uma integral de w_t. Por isso é dito que z_t segue um modelo auto-regressivo, integrado, de médias móveis.

Portanto, o modelo supõe que a *d*-ésima diferença da série z_t pode ser representada por um modelo ARMA, estacionário e invertível. O ARIMA, na verdade, é um caso especial de um processo integrado. (SOUZA & CAMARGO, 2004).

2.3.5
Modelos ARIMA Multiplicativos

Os modelos SARIMA (p,d,q) x $(P,D,Q)_S$, também conhecidos como *ARIMA multiplicativo*, possuem duas partes: a simples e a sazonal. De modo que a sazonalidade da série temporal passa a ser considerada.

O modelo SARIMA (*p*,*d*,*q*) * (*P*,*D*,*Q*) é dado por:

$$\Phi(B^S)\ \phi(B)\ \nabla^d\ \nabla^D{}_S\ \tilde{y}_t = \Theta(B^S)\ \theta(B)\ a_t \qquad (2.57)$$

Onde:

$\nabla^d = (1-B)^d$: operador diferença não sazonal de ordem *d*;

$\nabla^D{}_S = (1-B^S)^D$: operador diferença sazonal de ordem *D*;

$\theta(B) = (1-\theta_1 B^1 - ... - \theta_q B^q)$: operador não sazonal de médias móveis;

$\phi(B) = (1-\phi_1 B^1 - ... - \phi_q B^q)$: operador não sazonal auto-regressivo;

$\Phi(B^S) = (1-\Phi_1 B^{1S} - ... - \Phi_P B^{PS})$: operador sazonal auto-regressivo; e

$\Theta(B^S) = (1-\Theta_1 B^{1S} - ... - \Theta_P B^{PS})$: operador sazonal de médias móveis.

Uma vez definido o tamanho do período sazonal, analisam-se os correlogramas da ACF e PACF a fim de estimar os índices *p, d* e *q*, como também *P*, *D* e *Q*, atinentes à parte sazonal. O raciocínio para estimação do modelo é similar à parte simples, mas é feita sob os períodos *S*, *2S*, *3S*,... (SOUZA & CAMARGO, 2004). Para mais detalhes sobre a metodologia ARIMA, veja: HILL, GRIFFITHS & JUDGE (1999), SOUZA & CAMARGO (2004) e MORETTIM (2006).

2.3.6
Identificação de modelos

É possível realizar a identificação, *a priori*, de um modelo plausível através de análises de dependência linear nos correlogramas da ACF e PACF.

Segundo SARTORIS (2003), algumas características destes gráficos caracterizam determinados processos estocásticos:

- MA (*q*) – O correlograma da ACF tem picos do *lag 1* até o *q*, caindo de forma brusca a zero no *lag q+1*. O gráfico da PACF apresenta um decaimento exponencial ou senoidal amortecida Por exemplo, assuma um MA (*1*), o pico será no *lag 1*, declinando abruptamente a zero a partir do *lag 2*. Caso $\theta_1 < 0$, o pico será positivo e se $\theta_1 > 0$, negativo. O correlograma da PACF mostra um decaimento exponencial negativo para $\theta_1 > 0$ e senoidal amortecida para $\theta_1 < 0$;

- AR (*p*) – O correlograma da ACF tem um decaimento exponencial ou senoidal amortecido. O gráfico da PACF tem pico do *lag 1* até o *p*, caindo bruscamente a zero no *lag p+1*. Por exemplo, assuma um AR (*1*), o pico será no *lag 1*, declinando abruptamente a zero a partir do *lag 2*. Caso $\phi_1 > 0$, o pico será positivo e se $\phi_1 < 0$, negativo. O correlograma da ACF mostra um decaimento exponencial positivo para $\phi_1 > 0$ e senoidal amortecida para $\phi_1 < 0$;

- SAR (*P*) e SMA (*Q*) – de igual modo às partes simples ar (p) e ma (q), a análise é realizada, respectivamente, considerando, porém, os ciclos sazonais. Assim, para *S = 12*, suponha que haja uma queda brusca no lag 36 da FACP e que o correlograma da ACF apresente um decaimento exponencial ou senoidal amortecido nos lag's: 12, 24, 36,... A parte sazonal poderia ser representada pelo polinômio SAR (*1*).

2.3.7
Análise de Resíduos

A análise de resíduos consiste em aplicar testes estatísticos a fim de verificar se os pressupostos (normalidade, estacionariedade fraca e independência) são respeitados. O teste de normalidade que pode ser utilizado é o Jarque-Bera.

$$JB = \left(\frac{\hat{S}-0}{\sqrt{6/n}}\right)^2 + \left(\frac{\hat{K}-3}{\sqrt{24/n}}\right)^3 = \frac{n}{6}\hat{S}^2 + \frac{n}{24}(\hat{k}-3) \qquad (2.58)$$

Onde $\hat{S}$ é o coeficiente de assimetria estimado; $\hat{K}$, o de curtose e *n*, o tamanho da amostra. Supondo que S e K sejam normais, a estatística *JB* converge assintoticamente à distribuição $\chi^2_{(2)}$. A hipótese nula é de normalidade. Outros testes também podem ser usados, como: Anderson & Darling, QQ-plot.

Quanto à estacionariedade, podem ser usados os testes de Dikey-Fuller, Dikey-Fuller aumentado (*ADF(m)*) e Philips-Perron. A evidência de independência, por sua vez, pode ser verificada através do teste de Ljung-Box.

$$LB(m) = n(n+2)\sum_{k=1}^{m} (n-k)^{-1}\, r_k^2(\hat{\varepsilon}) \qquad (2.59)$$

Sob a hipótese de nula de não existência de auto-dependência linear na série de resíduos até o lag *m* de um modelo ARIMA (*p,d,q*). O teste possui distribuição assintótica $\chi^2_{(m-p-q)}$. Estes pressupostos são importantes, pois propriedades estatísticas desejáveis e consistência das estatísticas de teste.

2.3.8
Teste de Sobrefixação

De acordo com SOUZA & CAMARGO (2004), o teste de sobrefixação consiste, basicamente, na elaboração de um modelo com um número superior de parâmetros ao fixado, como uma tentativa de cobrir supostas direções de discrepâncias. Submete-se, então, à análise estatística para averiguação de necessidade de parâmetros adicionais.

2.3.9
Estimação

Muitos métodos de estimação dos parâmetros dos modelos ARIMA foram desenvolvidos (veja BOX, JENKINS & REINSEL (1994), BROCKWELL & DAVIS (1977)). A estimação pode ser feita através de OLS e máxima verossimilhança. Um método, por vezes utilizado, é o OLS interativo.

2.3.10
Previsões

A previsão h passos à frente é calculada através do valor esperado condicional à amostra. Formalmente:

$$E[y_{T+h} / y_T, ..., y_1] = \hat{y}_T(h), \ \forall \ h \in \mathbb{N}^* . \qquad (2.60)$$

Em síntese, segundo HILL, GRIFFITHS & JUDGE (1999), são necessárias seis etapas à modelagem BOX & JENKINS:

1. Identificação dos valores sugeridos para p, d, q, P, D, Q, a partir das análises dos correlogramas;
2. Estimação dos parâmetros do modelo (OLS, Máxima verossimilhança, OLS interativo);
3. Estatísticas de aderência (significância das estimativas, análise residual e análise das estatísticas de desempenho);

4. Se satisfatório, realiza-se o procedimento 5. Caso contrário, indica-se que outros valores para *p, d, q, P, D, Q* (procedimento 2);
5. Teste de sobrefixação; e
6. Previsão

2.4
Modelos de Redes Neurais Artificiais

As redes neurais artificiais (RNA) são sistemas paralelos compostos por unidades de processamentos simples, conhecidas como neurônios ou processadores, que representam funções lineares e não-lineares. Os referidos são dispostos em uma ou mais camadas, sendo interligados por um grande número de conexões (sinapses) que comumente estão associadas a pesos, responsáveis por ponderar os sinais (dados) de entrada recebidos por respectivo neurônio.

O funcionamento de uma RNA é inspirado nos neurônios biológicos e em sua estrutura paralela de processamento. Portanto, possui a capacidade de adquirir, armazenar e utilizar conhecimento experimental, podendo ser utilizada em problemas de reconhecimento de padrões, clusterização e previsão. (HAYKIN, 2001)

O neurônio biológico, apresentado pela figura 2.3, é, de forma simplista, a unidade básica da estrutura do cérebro. Sua membrana exterior é chamada de dentritos, que são dispositivos de entrada, cujo objetivo é conduzir os sinais das extremidades para o corpo celular. Por outro lado, o axônio é o dispositivo de saída que transmite o sinal do corpo celular para suas extremidades. As extremidades do axônio são conectadas com os dentritos de outros neurônios através das sinapses que, por sua vez, possuem neurotransmissores responsáveis por permitir a propagação dos pulsos nervosos, podendo ser excitatórias ou inibitórias.

Em outras palavras, o neurônio biológico é uma célula que, ao ser estimulada em uma entrada, gera uma saída que é enviada a outros neurônios. Esta saída depende da força de cada uma das entradas e magnitude da conexão associadas a cada entrada (sinapse). Uma sinapse pode excitar um neurônio, aumentando sua saída, ou inibi-lo, diminuindo-a.

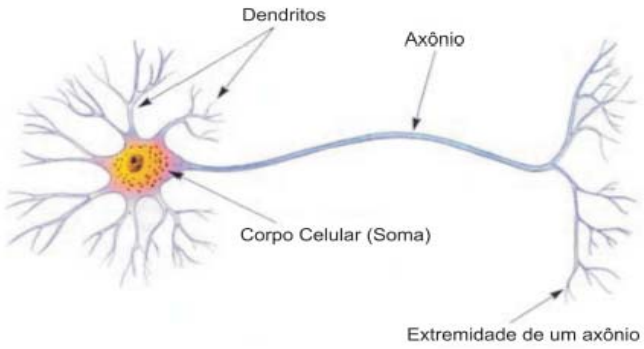

Figura 2.3 - Neurônio Biológico

Em 1943, Warren McCulloch e Walter Pitts publicaram um trabalho pioneiro no qual introduziram a conceito de redes neurais como máquinas computacionais (neuro-computação). O neurônio foi proposto com pesos fixos, ou seja, não ajustáveis.

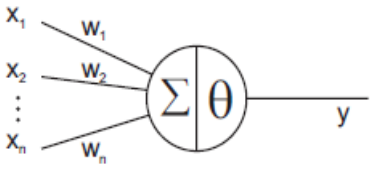

Figura 2.4 - Neurônio de McCulloch-Pitts

Cada x_i ($i=1,...,n$), representa uma entrada e w_i ($i=1,...,n$), um peso associado à esta. O somatório representa a soma ponderada das entradas e o teta, a função de ativação. Após várias pesquisas, chegou-se a uma rede neural constituída de vários neurônios artificiais, altamente conectados. Na década de 80, diversas redes neurais artificiais capazes de trabalhar com funções não linearmente separáveis foram propostas. Dentre tais, destacam-se as famosas Redes Neurais Artificiais *Feed-forward Multilayer Perceptron* (MLP), as quais são amplamente utilizadas em aplicações envolvendo previsão, classificação, identificação de padrões.

Na figura 2.5, tem-se a estruturação básica de um neurônio artificial, composto por dois módulos de processamento:

- Regra de propagação: executa uma soma ponderada das entradas multiplicadas pelos pesos sinápticos associados a cada entrada do neurônio; e

- Função de ativação: é uma função que é aplicada ao resultado da regra de propagação. O resultado da função ativação é a saída do neurônio artificial.

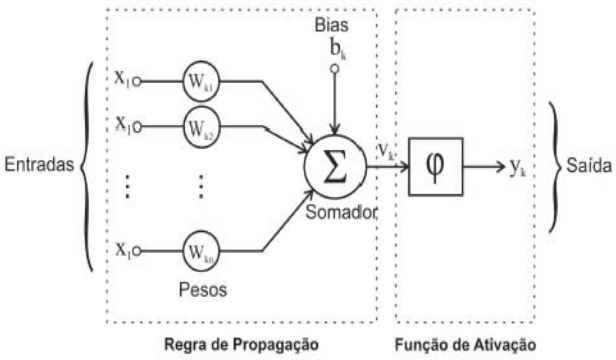

Figura 2.5 - Arquitetura de um Neurônio Artificial

2.4.1
Normalização dos Sinais de Entrada

Antes de os sinais de entrada serem inseridos à rede, é necessário realizar seu pré-processamento. Uma opção é a normalização através do método pela faixa (usado nesta dissertação), que tem o objetivo de suprimir o problema de paralisia neural (HAYKIN, 2001).

A equação (2.61) mostra cálculo desta transformação.

$$y_n = 2 \times \frac{y - \min y}{\max y - \min y} - 1 \qquad (2.61)$$

Onde:

- y : padrão original;
- y_n : padrão normalizado;
- $\min y$: valor mínimo do vetor de entrada; e
- $\max y$: valor máximo do vetor de entrada.

Assim, os dados de entrada são mapeados no intervalo [-1, 1]. Um aspecto importante é que a escolha da transformação deve considerar o domínio da função

de ativação utilizada. Por outro lado, o cálculo da desnormalização é necessário para que os sinais de saída da rede retornem à escala original. Isolando-se o valor de *y*, na equação (2.61), obtém-se este cálculo.

2.4.2
Estrutura

Não existe nenhum procedimento determinístico para se estruturar uma rede neural artificial. Portanto, sua estruturação é feita de forma heurística, tendo como parâmetros básicos principais:

- Representação dos dados (*I/O*);
- Tamanho das amostras de treino, validação e teste;
- Número de camadas (*layers*);
- Número de neurônios por camada;
- Topologia;
- Função de ativação e
- Algoritmo de aprendizado.

2.4.3
Funções de Ativação

Comumente, as funções de ativação são não-lineares, o que transformam as redes neurais em sistemas computacionais capazes de resolver problemas complexos. Assim, destacam-se:

- Função logística sigmoidal: mapeia os sinais de entrada dos neurônios no intervalo [0,1]. É a função geralmente adotada, por ser contínua, monotônica, não-linear e facilmente diferenciável em qualquer ponto;
- Função tangente hiperbólica: mapeia os sinais de entrada dos neurônios no intervalo [-1, +1]. Possui as mesmas características e emprego da função logística sigmoidal, possibilitando que as saídas sejam simétricas; e

- Função linear: os neurônios com esta função de ativação podem utilizados como aproximadores lineares.

Tabela 2.1 - Funções de Ativação

Função	Equação
Linear	$y = x + b$
Logística Sigmoidal	$y = \frac{1}{1+e^{-(n+b)}}$
Tangente Sigmoidal	$y = \frac{e^{(x+b)} - e^{-(x+b)}}{e^{(x+b)} + e^{-(x+b)}}$

2.4.4 Topologia

A topologia de uma rede neural diz respeito ao número de camadas e ao modo como estas estão interconectadas. O aumento do número de camadas implica maior complexidade e tempo de processamento ao modelo neuronal. Um número maior de neurônios por camada acarreta o crescimento dos graus de liberdade das funções de ativação. Todavia, quanto maior os graus de liberdade, menor a capacidade de generalização da rede neural. (HAYKIN, 1999)

Os dois principais tipos de estrutura topológica que compõem os modelos RNA são a unidirecional (*feed-forward*) e a recorrente (*feed-back*). (HAYKIN, 1999)

- *Feedfoward*: são estruturas de redes neurais cujas conexões entre neurônios de diferentes camadas (inter-camadas) obedecem à direção

"entrada $\Rightarrow$ saída", não havendo, assim, conexões entre neurônios de mesma camada e nem tampouco para as anteriores; e

- Recorrente: são estruturas com realimentação, onde um neurônio pode ser retroalimentado pela sua saída, de forma direta ou indireta. Cada camada pode conter conexões entre os neurônios da mesma camada (estímulos laterais) e de camadas anteriores e posteriores. Não existe, portanto, um sentido único para o fluxo dos dados entre neurônios ou entre camadas. (ZURATA, 1992)

A figura 2.6 ilustra a topologia de uma rede neural artificial *feed-forward* com três camadas.

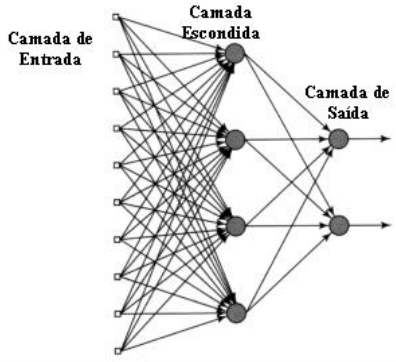

Figura 2.6 - Arquitetura Neural *Feedfoward* com Três Camadas.

2.4.5 Aprendizado

O aprendizado de uma rede neural artificial se constitui no ajuste do conjunto de pesos de modo a executar uma tarefa específica, acontecendo basicamente de duas formas distintas: aprendizado supervisionado e não-supervisionado (para mais detalhes do não-supervisionado, HAYKIN (1999)).

- Supervisionado: utiliza um conjunto de pares (entrada - saída), em que, para cada padrão de entrada, é especificado um padrão de saída desejado (resposta desejada). Ou seja, tem-se a saída fornecida pela rede neural e

um sinal desejado. Subtraindo-os, obtém-se o respectivo valor de erro. (HAYKIN, 1999)

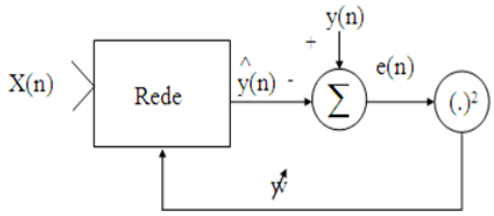

Figura 2.7 - Modelo de Treinamento Supervisionado

O processo básico do aprendizado supervisionado, ilustrado na figura 2.7, pode ser sumarizado pelos seguintes passos:

i. Escolha inicial dos pesos. Em geral, estes valores são pequenos e escolhidos aleatoriamente;
ii. Apresentação de uma entrada cuja saída correspondente é conhecida;
iii. Cálculo da saída a ser gerada pela rede;
iv. Cálculo do erro na camada de saída (sinal desejado subtraído do gerado);
v. Ajuste dos pesos utilizando algum algoritmo de aprendizado; e
vi. Verificação do erro. Se satisfatório, fim; caso contrário, retorna-se ao ii.

2.4.6
Algoritmo de Aprendizado Gradiente Decrescente

O algoritmo gradiente decrescente é baseado no de retropropagação do erro e criado a partir da generalização da Regra de Delta, proposta por WINDROW & HOFF (1960), utilizado em redes do tipo *feedfoward perceptron* (uma camada). O ajuste dos pesos das conexões objetiva à minimização do erro quadrático médio.

$$E_{SSE} = \frac{1}{2}\sum_p \sum_j (t_{pj} - s_{pj})^2$$

Onde: t_{pj}, valor desejado de saída p do processador j da camada de saída e s_{pj}, estado de ativação de saída p do processador j da camada de saída.

O processo de minimização do erro quadrático médio ocorre com a utilização do método gradiente decrescente.

$$\Delta W_{i,j} = -\eta \frac{\partial E}{\partial W_{i,j}}$$

Sendo: η, a taxa de aprendizado e $\frac{\partial E}{\partial W_{i,j}}$, a derivada parcial do erro em relação ao peso da respectiva conexão.

A principal restrição na minimização do erro, quanto ao gradiente decrescente, é que a função de ativação do respectivo neurônio tem de ser monotônica e diferenciável em qualquer ponto.

O algoritmo BP atualiza os pesos através da retropropagação do erro da saída da rede. A atualização dos pesos é realizada, de forma contínua, no sentido contrário ao fluxo dos sinais de entrada. A topologia da rede que utiliza esta regra é formada, em geral, por neurônios não-lineares, na camada oculta, e na de saída, por lineares.

2.4.7
Algoritmo de Aprendizado Levenberg-Marquardt

O algoritmo de Levemberg-Marquardt (LM) é considerado o método mais rápido para treinamento de redes *feed-forward backpropagation*, desde que a rede possua uma quantidade moderada de pesos sinápticos.

O algoritmo LM utiliza uma aproximação do método de Newton que é obtida a partir da modificação do método de Gauss-Newton, introduzindo o parâmetro μ, conforme mostra a equação (2.62).

$$\Delta w = (J^T J + \mu I)^{-1} J^T \varepsilon \qquad (2.62)$$

Onde:

- Δw: diferença entre os pesos inicial e final;
- μ: escalar que controla a derivação dos erros, permitindo que o termo ($J^T J$) possa ser invertido;
- J: jacobiano dos erros da camada de saída. Cada elemento da matriz J representa uma derivada parcial de um elemento da matriz de erros com o seu correspondente peso;

- I: matriz identidade multiplicada pela constante μ ; e
- ε : vetor de erros da rede neural calculados.

2.4.8
Redes Perceptron de Múltiplas Camadas

As redes Perceptrons de Múltiplas Camadas (*Multilayer Perceptron* - MLP) são constituídas da(s) camada(s): de entrada, oculta(s) e de saída. Em outras palavras, contêm uma ou mais camadas de neurônios ocultos, que não são a de entrada nem tampouco a de saída da rede. Os neurônios ocultos capacitam a rede a aprender tarefas complexas, extraindo progressivamente as características mais significativas dos padrões de entrada.

Para resolução de inúmeros problemas, as redes MLP têm sido utilizadas, através de seu treinamento de forma supervisionada, com o algoritmo de retropropagação de erro (*Error Back-Propagation*), que é baseado na regra de aprendizagem por correção de erro.

Existem atualmente vários algoritmos de treinamento para as redes neurais MLP, tais são geralmente do tipo supervisionado. De acordo com os parâmetros a serem atualizados, os algoritmos de treinamento podem ser classificados como:

- Estáticos: não alteram a estrutura da rede, variando apenas os pesos sinápticos; e

- Dinâmicos: podem modificar o tamanho da rede neural, como o número de camadas, de neurônios nas camadas escondidas e / ou de conexões.

Assim, quando é utilizado o aprendizado estático, a mesma regra de aprendizado é empregada às redes, com diferentes formatos e tamanhos. O *backpropagation* é um algoritmo supervisionado que utiliza pares, entrada e saída desejada, a fim de que, por meio de um mecanismo de correção de erros, sejam ajustados os pesos da rede. O treinamento ocorre em duas fases: *forward* e *backward*, tendo dois sentidos distintos.

- *Forward*: utilizada para definir a saída da rede neural, dado determinado padrão de entrada, onde o vetor de entrada é aplicado aos nodos da rede e propaga seu efeito camada por camada, sendo, nesta etapa, todos os pesos sinápticos fixos; e

- *Backward*: utiliza a saída desejada e a fornecida pela rede neural para atualização dos pesos de suas conexões, onde a resposta real é subtraída da desejada para produção de um sinal de erro. Este, por sua vez, é propagado contra a direção das conexões sinápticas, ajustando os pesos sinápticos, de modo que a resposta real da rede se aproxime da desejada. (HAYKIN, 2001)

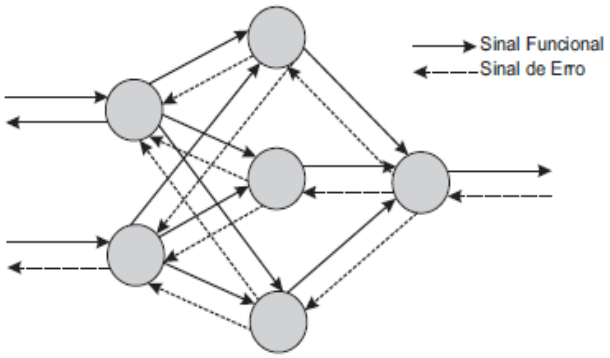

Figura 2.8 – Esquema de Propagação dos Sinais da MLP

Mais detalhes acerca dos métodos das redes neurais artificiais podem ser verificados em HAYKIN (1999); RUSSEL & MARK (1999); BRAGA, CARVALHO & LUDEMIR (2000).

3
MÉTODOS DE COMBINAÇÃO DE MODELOS

Considere uma situação onde existe um número k ($k > 2$) de modelos preditivos plausíveis para um processo de séries temporais $\{ y_t \}$ - para $t = 1, 2, ...$ -, no entanto há o interesse em determinar apenas um modelo único a ser usado para sua projeção temporal.

Para FARIA e MUBWANDARIKWA (2008), existem três possibilidades de obtê-lo:

1. Escolher um modelo individual de um conjunto de modelos plausíveis $M = \{M_1, ..., M_k\}$, baseado em algum critério de seleção, e usá-lo para previsão;

2. Combinar as previsões oriundas dos modelos do conjunto M e utilizar a previsão combinada; e

3. Combinar as densidades preditivas oriundas dos modelos do conjunto M e utilizar a preditiva combinada para previsão.

Acerca do item *1*, existe diferentes estatísticas de aderência que, em consonância com a abordagem, podem ser adotadas na escolha do melhor modelo. Estas medidas de desempenho, portanto, são escolhidas, de forma a obter o modelo que melhor se ajusta à dinâmica temporal da série (MAPE, MAE, R^2). Os itens *2* e *3* tratam de abordagens distintas, porém com o mesmo fim, projetar séries temporais. As medidas de aderência utilizadas para os individuais podem também ser utilizadas para os de combinação.

Em se tratando de combinações de previsões ou de densidades preditivas, ambas as abordagens podem ser tratadas como combinação de modelos (ou métodos), dependendo, então, da abordagem considerada.

No contexto de combinação de modelos, alguns resultados, obtidos por alguns autores, merecem atenção e a sugerem como uma alternativa factível e eficiente.

Para HOLLAUER, ISSLER e NOTINI (2008), algumas razões sugerem que a combinação de previsões pode construir previsões de desempenho superior:

> *Primeiro, a pura diversificação de previsões leva à diminuição do erro diversificável. A segunda razão diz respeito à robustez, na medida em que, não sendo particular a especificação, não padece também de suas deficiências, além de incluir mais variáveis e informações de especificação de outros modelos, uma vez que se pesam informações de modelos que incluem outras variáveis e seus lag's. Ademais, a experiência mostra que modelos parcimoniosos combinados possuem desempenho superior.*

MAHMOUD (1989) destaca três motivos que encorajam o uso de combinação de previsões:

i. A combinação naturalmente melhora a acurácia e diminui a variância do erro da previsão combinada;
ii. Os usuários que possuem pouca experiência na área podem utilizar métodos simples de combinação, pois podem acarretar melhoras à acurácia; e
iii. A combinação pode ser feita com pouco ou nenhum custo.

Desde os estudos iniciais sobre combinação de previsões, resultados empíricos têm mostrado que a previsão combinada é mais acurada, em inúmeras situações, que qualquer previsão individual formadora da combinação (RAUSSER & OLIVEIRA, 1976; CLEMEN, 1989; MAKRIDAKIS *et al*., 1998 e ARMSTRONG, 2001a). Além disso, a demonstração teórica de que a combinação linear de modelos é superior a qualquer método individual, para o método de

variância mínima, pode ser vista em DICKINSON (1975) , *apud in* WERNER & RIBEIRO (2006),

Segundo GUPTA & WILSON (1987), se os modelos são 'conhecidos', *a priori*, esta informação deve ser incorporada à previsão combinada. Para ARMSTRONG (2001a), quando existem evidências de que determinada técnica possui maior acurácia, deve-se incorporá-la à combinação com maior peso associado.

HANKE, REITSH & WICHERN (2001), citando ARMSTRONG (1989), MAHAMOUD (1989) e CLEMEN (1989), defendem que os ganhos são razoáveis.

ZOU e YANG (2004) afirmam que:

> *Ao se combinar métodos, é possível reduzir a variabilidade das previsões em relação ao valor verdadeiro.*

MAKRIDAKIS & WINKLER (1983) destacam que a acurácia da combinação de previsões depende do número de técnicas distintas que são combinadas. A variabilidade nas medidas de aderência está associada à escolha das técnicas e pode ser reduzida com a inclusão de outras. Ainda conforme os autores, a acurácia cresce à medida que se aumenta o número de modelos combinados, embora o grau de saturação encontrado tenha sido de quatro ou cinco técnicas. LIBBY & BLASHFIELD (1978) registram que a maioria das melhorias ocorreu quando utilizadas duas ou três.

Estudos empíricos têm mostrado as combinações lineares dependentes da estimativa das correlações têm apresentado, na prática, um desempenho muito fraco (NEWBOLD & GRANGER (1974), WINKLER & MAKRIDAKIS (1983) e CLEMEN, (1989)).

De acordo com WINKLER (1989), fatores complicadores, tais como: dependência entre as previsões, incerteza acerca de suas características e instabilidade no processo preditivo sugerem que seja proveitoso obter combinações de técnicas relativamente robustas. Além do mais, pode ser mais

proveitoso, em algumas situações, considerar a simples média aritmética, sem considerar os refinamentos do procedimento de combinação.

ARMSTRONG (2004) afirma que a combinação pode ser feita também através da média simples das previsões apenas.

Uma vez que o desempenho da combinação seja, geralmente, medido pela acurácia, combinar previsões, para ARMSTRONG (2001a), aumenta a aderência aos dados, na medida em que as componentes da combinação contenham informações úteis e mais independentes.

Ainda segundo ARMSTRONG (2001a):

> *As previsões, muitas vezes, apresentam correlação positiva forte. Existem duas maneiras de gerar previsões independentes: a primeira é a utilização de previsões que empregam diferentes informações para compor as previsões combinadas, com a mesma técnica. A outra consiste na utilização de técnicas diferentes, pois analisam os dados de diferentes formas. Além disso, a previsão combinada deve ser usada quando se tem duas ou mais técnicas razoáveis disponíveis e na existência de incerteza na seleção da melhor metodologia.*

BATCHELOR & DUA (1995) afirmam que a combinação é mais efetiva quando os métodos são consideravelmente diferentes.

MAKRIDAKIS & WINKLER (1983) destacam que a utilização de técnicas relativamente simples pode ser viável e menos arriscada que o uso de métodos complexos. Para CHAMBERS *et al.* (1971), os custos aumentam à medida que os métodos individuais tornam-se mais sofisticados. Com efeito, diminuem à proporção que se utilizam técnicas mais simples.

Segundo CLEMEN (1989), a combinação é uma abordagem atraente para realizar previsões, visto que, ao invés de tentar escolher a melhor técnica, formula-se o problema questionando que técnicas poderiam ajudar na melhoria da

acurácia. Como as previsões podem ser afetadas por diversos fatores, cada método pode contribuir capturando algum tipo de informação que influencia tais fatores.

FARIA e MUBWANDARIKWA (2008) colocam que o modelo combinado é capaz de produzir projeções mais acuradas, visto que é um agregador de informações. Ainda de acordo com os autores, a principal vantagem acontece em face à incerteza de qual modelo individual utilizar.

Para realização de combinações de previsões, FLORES & WHITE (1988) destacam que duas dimensões podem ser consideradas:

- Seleção das componentes do modelo de combinação; e
- Seleção do método de combinação.

Conforme FLORES & WHITE (1988), as componentes dos modelos de combinação são denotadas como previsões-base e podem ser classificadas, em três categorias: objetivas, subjetivas ou através da utilização de ambas (objetivas e subjetivas). A categoria objetiva engloba os modelos de Amortecimento Exponencial, ARIMA, Redes Neurais Artificiais, assim como outros procedimentos com base matemática. Por outro lado, a subjetiva inclui todas as abordagens que envolvem o julgamento humano, tal como grupo focado ou opinião de especialistas.

A segunda dimensão concerne à maneira com a qual as técnicas devem ser combinadas. Tal é alvo de estudo há muito tempo. De acordo com CLEMEN (1989), alguns métodos têm sido desenvolvidos para encontrar a melhor combinação. Apesar disso, o resultado tem sido unânime: combinar previsões conduz ao aumento de acurácia da previsão combinada, em relação às previsões-base. A dimensão dos métodos de combinação envolve duas abordagens: objetiva ou subjetiva. A objetiva se refere aos métodos que fazem uso da matemática, de forma que os resultados possam ser repetidos. Em contrapartida, a subjetiva inclui esforços intuitivos para combinar subjetivamente previsões-base, empregando conhecimento e opinião individual ou de grupo.

Nesta dissertação, os esforços concentraram-se especificamente em combinações e previsões objetivas A figura 3.1 ilustra esta ideia, considerando *j* modelos *h* passos à frente, a figura 3.1 ilustra os modelos de combinação objetiva

de métodos objetivos. A partir de uma amostra, estima-se *j* modelos e posteriormente e os combinam, gerando a previsão combinada.

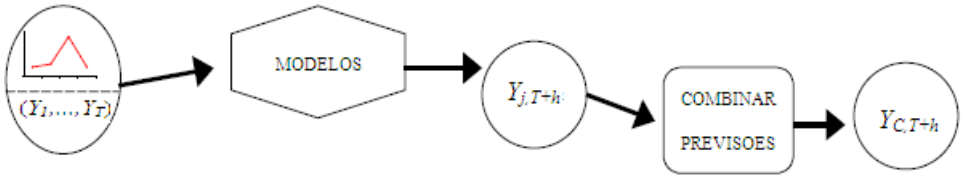

Figura 3.1 – Esquema Ilustrativo dos Métodos de Combinação (Objetiva) de Previsões Objetivas

Assim o sendo, três métodos de combinações (linear, média simples e geométrico) são definidos, a seguir, e, posteriormente, no capítulo 4, aplicados e contrapostos, segundo determinadas medidas de acurácia.

3.1
Combinação Linear Modelos de Previsão

O estudo de GRANGER & BATES (1969) é considerado o artigo seminal em modelos combinação de previsões.

Neste, os autores propuseram que o método de combinação deveria constar de uma combinação linear de duas previsões objetivas não-viciadas.

Assim, tem-se o modelo (amostral) é dado por:

$$\hat{y}_{CL,T+h} = \hat{\omega}\ \hat{y}_{1,T+h} + (1-\hat{\omega})\ \hat{y}_{2,T+h} \qquad (3.1)$$

Onde:

- $\hat{y}_{1,T+h}$, previsão do modelo 1;
- $\hat{y}_{2,T+h}$, previsão do modelo 2;
- $\hat{y}_{CL,T+h}$, previsão do modelo de combinação linear; e
- $\hat{\omega}$, estimador adaptativo do modelo de combinação linear.

Os pesos a serem atribuídos a cada previsão poderiam ser iguais. Contudo, os autores preferiram dar maior relevância à previsão mais acurada. Assim, visando a estimar um valor de peso ω, propuseram cinco métodos de estimação de

pesos que objetivam à minimização da variância dos resíduos da previsão combinada. Mais detalhes acerca da variância e dos pesos GRANGER & BATES (1969) podem ser verificados no apêndice D.

NEWBOLD & GRANGER (1974) ampliaram o número de previsões base, mantendo todas as suposições de GRANGER & BATES (1969), quanto aos pesos e aos cinco procedimentos de estimação, a serem combinadas. Através da combinação de três previsões obtida dos métodos ARIMA, Holt-Winters e de auto-regressão *Stepwise*, concluíram que a combinação trouxe melhoras à acurácia de sua previsão. Assim, tem-se que:

$$\hat{y}_{CL,T+h} = \sum_{j=1}^{3} \hat{\omega}_j \, \hat{y}_{j,T+h} \tag{3.2}$$

WINKLER & MAKRIDAKIS (1983), *apud in* WERNER & RIBEIRO (2006), também analisaram combinações ponderadas entre dez métodos de previsão, e os resultados obtidos confirmaram as conclusões de NEWBOLD & GRANGER (1974). Estes resultados consistiram na comparação das estatísticas MAPE de mil e uma séries temporais, o que permitiu concluir ser melhor ignorar os efeitos da correlação no cálculo de combinações ponderadas.

GRANGER & RAMANATHAN (1984), *apud in* WERNER & RIBEIRO (2006), chamaram a atenção para o fato de que os métodos convencionais de combinação linear de previsões poderiam ser vistos como uma forma estruturada de regressão. Argumentaram ainda que o método é equivalente ao método de mínimos quadrados ordinários (MQO), considerando a previsão combinada como variável resposta e as individuais base como explicativas.

Genericamente, as previsões pontuais são combinadas linearmente utilizando algum mecanismo ponderador de forma à minimização da variância dos resíduos combinados ou outra função residual. Os pesos podem ser fixos ou variáveis (não necessariamente positivos) e / ou somar ou não uma unidade. Destaca-se ainda que as médias, simples, ponderada ou harmônica, também podem ser utilizadas na combinação de previsões.

Na abordagem clássica de GRANGER e NEWBOLD (1986), destaca-se que as previsões pontuais são combinadas linearmente usando algum mecanismo

ponderador, visando à minimização da variância dos resíduos. Portanto, o modelo de combinação linear, *h* passos à frente, pode ser descrito pela forma geral (3.3):

$$\hat{y}_{CL,\ T+h} = \sum_{j=1}^{k} \hat{\omega}_{jt}\, \hat{y}_{j,\ T+h} \tag{3.3}$$

Sendo $\hat{y}_{CL,T+h}$, a previsão combinada para o instante $T+h$; $\hat{\omega}_{jt}$, o peso estimado (não necessariamente positivo ou normalizado) e $\hat{y}_{j,\ T+h}$, a previsão modelo *j*, para o instante $T+h$ - considerando uma amostra $(y_{1,\ldots,}y_T)$.

EVANS (2003) coloca ainda a possibilidade do uso de um modelo de regressão para a realização das combinações com uma constante aditiva.

Outra abordagem é a combinação linear de densidades preditivas, na qual algumas propriedades merecem destaque.

CLEMEN e WINKLER (1999) mostram que, se todos os modelos de probabilidades são similares, então o modelo probabilístico combinado é igual a estes, *apud in* MUBWANDARIKWA (2007).

Consoante TITTERINGTON, SMITH e MARKOV (1985), *apud in* MUBWANDARIKWA (2007), a combinação linear de densidades Gaussianas $\rho(y|\ \mu_j, \sigma_j^{\ 2})$, com média μ_j e variância $\sigma_j^{\ 2}$ (*j = 1, 2*), são unimodais sob certas condições restritivas. Conhecendo-se, por exemplo, as médias (diferentes) e as variâncias (iguais) de duas normais, a unimodalidade ocorre se $(\mu_1 - \mu_2)^2 < 4\sigma^2$. De outro modo, tem-se uma densidade combinada bimodal.

Uma condição suficiente para unimodalidade de densidades normais combinadas é mostrada por EINSENBERGER (1964), *apud in* MUBWANDARIKWA (2007):

$$(\mu_1 - \mu_2)^2 < \frac{27}{4} \frac{\sigma_1^{\ 2}\sigma_2^{\ 2}}{(\sigma_1^{\ 2} + \sigma_2^{\ 2})} \tag{3.4}$$

Onde $(\mu_1, \sigma_1^{\ 2})$ e $(\mu_2, \sigma_2^{\ 2})$ são, respectivamente, os parâmetros, média e variância, de duas densidades Gaussianas. Este resultado pode ser estendido para mais densidades com o princípio de que, em ordem, a relação acima deve ser satisfeita.

Particularmente, as condições para unimodalidade de distribuições v*t*-student são mais restritivas que a normal e, portanto, mais difíceis de serem obtidas. Além disso, estas combinações são tipicamente multimodais. Em MUBWANDARIKWA (2007), encontram-se detalhes e refinamentos acerca das combinações lineares da distribuição *t*.

Existem distribuições em que a unimodalidade é difícil de ser encontrada. Assim, uma possível solução é a adoção de aproximações, como simulações Markov Chain Monte Carlo (MCMC), que estimam a preditiva combinada. (MUBWANDARIKWA, 2007)

De acordo com MUBWANDARIKWA & FARIA (2008), a formulação geral desta abordagem pode ser descrita por:

$$\rho_{CL(\rho_1,\ldots,\rho_k)}\,(y_{CL,T+h}\,|\;D_T) \;=\; \sum_{j=1}^{k} \omega_j\;\rho_j(y_{j,T+h}\,|\;D_T) \qquad (3.5)$$

Sendo:

- $\rho_{CL(\rho_1,\ldots,\rho_k)}\,(y_{CL,T+h}\,|\;D_T)$, a densidade preditiva resultante da combinação linear de densidades, dado as informações D_T ;
- $\rho_j(y_{j,T+h}\,|\;D_T)$, a densidade preditiva do modelo *j* (*j = 1*, ..., *k*) para o instante *T+h,,* dado D_T ; e
- ω_j , o peso adaptativo associado à densidade do modelo *j*.

3.2
Combinação Geométrica de Modelos de Previsão

Consoante MUBWANDARIKWA & FARIA (2008), a combinação geométrica de modelos bayesianos, em seu estudo de caso, forneceu maior acurácia, em relação aos modelos individuais base e sua respectiva combinação linear - na maioria das estatísticas de aderência consideradas.

Ainda de acordo com os autores, a combinação geométrica de densidades preserva a forma distribucional de muitas distribuições. Em particular, a combinação geométrica de distribuições da família exponencial é também membro desta família e apresenta, por isso, unimodalidade. Diferentemente do modelo de combinação linear que, em geral, ocorre somente sob certas condições.

No caso da distribuição *t*-student, enquanto a linear é tipicamente multimodal, nem sempre a geométrica o será.

Em MUBWANDARIKWA (2007), encontra-se a prova de que a densidade resultante da combinação geométrica de distribuições da família exponencial também é uma densidade exponencial. Podem ser verificadas ainda as condições necessárias para garantir a unimodalidade da combinação geométrica de distribuições *t*-student.

Segundo MUBWANDARIKWA & FARIA (2008), tem-se que:

> *Assumindo que as densidades* $\rho_1(y_{j,T+h} \mid D_T),...,\rho_k(y_{j,T+h} \mid D_T)$ *sejam densidades fortemente unimodais da família exponencial. Logo a densidade da combinação geométrica,* $\rho_{CG(\rho_1,...,\rho_k)}(y_{CG,T+h} \mid D_T)$, *é também fortemente unimodal da família exponencial.*

Consoante a notação de MUBWANDARIKWA & FARIA (2008), o modelo de combinação geométrica de distribuições pode ser formalmente escrito pela equação (3.6).

$$\rho_{CG(\rho_1,...,\rho_k)}(y_{CG,T+h} \mid D_T) = \prod_j^k c\, \rho^{\omega_j}{}_j(y_{j,T+h} \mid D_T) \qquad (3.6)$$

Onde:

- $c^{-1} = \int \prod_{j=1}^{k} \rho_j^{\omega_j}(y_{T+h} \mid D_T)\, \partial y_{T+h}$;
- ω_j, ($j = 1,...,k$), peso associado à distribuição $\rho_j(y_{T+h} \mid D_T)$ do modelo *j*;
- $\rho_j(y_{j,T+h} \mid D_T)$, distribuição do modelo *j*; e
- $\rho_{CG(\rho_1,...,\rho_k)}(y_{CL,T+h} \mid D_T)$, distribuição resultante da combinação geométrica de distribuições individuais.

Em particular, seja $\rho_j(y_t \mid \mu_j, \tau_j)$ uma densidade normal associada ao modelo *j*, com média e precisão ($\tau = \sigma^{-2}$) denotada por μ_j e τ_j ($j = 1,...,k$), respectivamente. Como estas distribuições pertencem à família exponencial, a combinação geométrica destas, $\rho_{CG}(y_t \mid \tilde{\mu}, \tilde{\tau})$, é também gaussiana.

MUBWANDARIKWA (2007) define as medidas de média e precisão, respectivamente, por: $\tilde{\mu} = \dfrac{\sum_{j=1}^{k} \tau_j \, \mu_j}{\sum_{j=1}^{k} \mu_j}$ e $\tilde{\tau} = \sum_{j=1}^{k} \omega_j \, \tau_j$

Diante dos resultados de MUBWANDARIKWA & FARIA (2008), a proposta desta dissertação é testar e inferir acerca da combinação geométrica de previsões pontuais, utilizando como modelos base dois estatísticos e um de inteligência artificial. Posteriormente, gerar cenários para os modelos univariados base de forma a construir intervalos confiança para os métodos de combinação.

A abordagem de combinação geométrica de previsões pode ser descrita como na equação (3.7).

$$\hat{y}_{CG,T+h} = \prod_{j}^{k} \hat{c} \; y^{\hat{\omega}_j}{}_{j,T+h} \tag{3.7}$$

Sendo:

- $\hat{y}_{CG,T+h}$, a previsão do modelo de combinação geométrica para o instante *T+h*, (*h = 1,2,...*);
- $\hat{c}$, a constante estimada multiplicativa de ajuste;
- $\hat{y}_{j,\,T+h}$, a previsão modelo *j*, para o instante *T+h* e
- $\hat{\omega}_j$, para ($j = 1,...,k$) e $\sum_{j=1}^{k} \omega_{jt} = 1$, o peso adaptativo associado à projeção $\hat{y}_{j,T+h}$ associada ao modelo *j* - considerando uma amostra $(y_{1,...,}y_T)$.

Evidentemente, as restrições aos pesos podem não ser assumidas, caso acarrete ganho de desempenho à modelagem. Segundo WERNER & RIBEIRO

(2006), dentre muitas conclusões obtidas pelos pesquisadores do assunto, a principal é que os modelos combinados cujos pesos são não negativos e a média, quase sempre, apresentam métricas superiores.

Assim, assumindo tais restrições, seja um vetor de pesos adaptativos $W_t = (\omega_t, \ ..., \ \omega_{k,t})'$ de um modelo de combinação geométrica. Dado que algum peso e a constante de ajuste sejam igual a *1*, obviamente todos os outros tem de ser iguais a zero, caracterizando um caso particular do método. Isso indica que o modelo individual é igual ao geométrico, sendo indiferente combiná-los, *a priori*.

O modelo geométrico possui vantagens consideráveis ante o linear. Primeiro, a melhor combinação pode não ser a linear. Não obstante, efeitos lineares também podem ser reproduzidos de forma aproximada pelo modelo geométrico (quando feita a correta ponderação), implicando que, em situações que o linear seja melhor, pode representar equivalente.

De outra maneira, o oposto não é possível, uma vez que os modelos de combinação linear somente são capazes de produzir efeitos lineares à previsão combinada.

3.3
Pesos Adaptativos e Constante Ajuste

Inúmeras abordagens têm sido propostas, na literatura, para obtenção de pesos adaptativos. Algumas os consideram fixos no tempo; outras, porém, variáveis. Assim, tem-se que:

- A primeira abordagem, proposta por WINKLER (1968), *apud in* WERNER & RIBEIRO (2006), assumia que os pesos eram iguais a $\frac{1}{k}$, sendo *k* o número de modelos;
- A metodologia de *quasi-Bayes*, por SMITH e MARKOV (1978), , *apud in* WERNER & RIBEIRO (2006), estima os pesos como um problema de classificação, utilizando a distribuição *Dirichlet*;

- MUBWANDARIKWA (2007) faz uma abordagem bayesiana cujos pesos são variáveis no tempo utilizando um método denotado por *Informative Prior Weights*;
- Outros autores podem ser citados ainda, como: BATES & GRANGER (1969), *Optimal Combination*; TURKHEIMER (2003), *Weights Akaike*; FARIA & SOUZA (1995), *quasi-Bayes* revisado.

Nesta dissertação, a estimação dos parâmetros dos métodos de combinação foi interpretada como um problema não linear de otimização e, por isso, a ferramenta *solver* (Excel) foi utilizada.

Os pesos e a constante de ajuste consideradas fixas no tempo. Com efeito, para o caso linear, evitou-se o problema de baixo desempenho preditivo, ocorridos quando as estimativas dos pesos dependem da correlação dos resíduos. (Mais detalhes, veja: NEWBOLD & GRANGER (1974), WINKLER & MAKRIDAKIS (1983) e CLEMEN, (1989)).

Em geral, a função a ser minimizada é o MSE, dentro da amostra. Considerando a abordagem adotada, os resultados utilizando esta estatística, porém, foram inferiores (fora da amostra, principalmente), em termos de acurácia, em relação à função MAPE (função custo proposta) - inclusive em relação ao próprio MSE. Foi testada também a eliminação das restrições impostas aos pesos adaptativos nas duas combinações. Embora, dentro da amostra, as métricas de aderência tenham melhorado, observou-se desempenho inferior fora da amostra.

É importante salientar que a questão como instabilidade temporal dos valores de APE (*absolute percentual error*) também foi considerada. Em outras palavras, analisaram-se os valores de MAPE (fora da amostra), bem como a flutuação temporal e valores de máximo e de mínimo do APE.

Verificou-se ainda que a constante *c* (multiplicativa) do modelo geométrico melhorou o desempenho, fora da amostra, quando assumiu valor igual a um. Para o caso linear, a constante (aditiva) ótima assumiu valor igual a zero. Portanto, às combinações, neste caso, as constantes de ajuste (aditiva e multiplicativa) foram consideradas elementos neutros. Em resumo, para os modelos de combinação linear e geométrica, consideraram-se as seguintes restrições:

- Estimativas normalizadas e positivas, cuja soma é igual a uma unidade, para os pesos adaptativos;
- Constantes de ajuste como elementos neutros; e.
- Função utilidade MAPE (*in sample*) a ser otimizada.

Especificamente à projeção de consumo residencial mensal de energia elétrica, a combinação linear de previsões é descrita conforme o modelo da equação (3.8).

$$\hat{y}_{CL,T+h} = (BJ_{T+h}) * \hat{\omega}_1 + (HW_{T+h}) * \hat{\omega}_2 + (RNA_{T+h}) * \hat{\omega}_3 \quad (3.8)$$

Ainda, considerou-se o modelo de WINCKLER (1968), com pesos iguais, que pode ser interpretado como um caso particular do modelo de combinação linear:

$$\hat{y}_{CL,T+h} = [(BJ_{T+h}) + (HW_{T+h}) + (RNA_{T+h})] * \frac{1}{3} \quad (3.9)$$

O modelo de combinação geométrica de previsões é dado por:

$$\hat{y}_{CG,T+h} = \hat{c} * (BJ_{T+h})^{\hat{\omega}_1} * (HW_{T+h})^{\hat{\omega}_2} * (RNA_{T+h})^{\hat{\omega}_3} \quad (3.10)$$

Onde, para os três métodos combinados, têm-se as definições:

- BJ_{T+h}, previsão do modelo ARIMA para o instante $T+h$;
- HW_{T+h}, previsão do MAE para o instante $T+h$;
- RNA_{T+h}, previsão do modelo de RNA para o instante $T+h$; e
- $\hat{\omega}_1$, $\hat{\omega}_2$, $\hat{\omega}_3$, pesos associados linearmente às respectivas previsões.

3.4
Intervalos de Confiança das Combinações

Uma vez escolhidos e estimados os modelos individuais, é possível gerar densidades preditivas através da utilização do método de Quase-Monte Carlo.

O procedimento de simulação utilizado para os modelos estatísticos, nesta dissertação, inicia-se com a geração de uma sequência de números quase-aleatórios independentes pertencentes à distribuição *U [0,1]*. Posteriormente, estes

são inseridos em um algoritmo de inversão (*Inversão de Moro*) que os interpreta como probabilidades acumuladas, de forma a fornecer amostras independentes pertencente à distribuição normal-padrão. Por seguinte, as amostras normais padrão são filtradas por Cholesky, gerando resíduos na escala da série temporal considerada (no caso, a de consumo residencial). Assim, para cada instante, realiza-se este procedimento *n* vezes. Mais detalhes sobre o algoritmo de Moro, assim como as sequências de Quase-Monte-Carlo, podem ser verificados no Apêndice C. O procedimento utilizado para as redes neurais são explanados mais adiante.

Segundo ALBUQUERQUE (2008), no caso multivariado, após cada sorteio, o vetor *u* (normal-padrão), no instante *t*, é filtrado pela matriz triangular inferior *Z*, advinda da decomposição de Cholesky da matriz de covariância amostral de seus resíduos ($\sum_{\hat{\varepsilon}}$). A populacional também pode ser usada, se conhecida. Para o caso univariado, a matriz *Z*, na verdade, é o desvio-padrão (escalar). O mesmo é multiplicado pelas amostras normais-padrões independentes, gerando resíduos na escala da série temporal, normalmente distribuídos.

Segundo HAMILTON (1994), toda matriz real definida positiva pode ser decomposta em uma matriz triangular inferior *Z* e uma matriz diagonal *D*, com números positivos em sua diagonal principal. Após a decomposição de Cholesky, têm-se as seguintes equações:

$$\sum_{\varepsilon_t} = ZDZ^T \quad \text{(caso multivariado)} \tag{3.11}$$

$$\hat{\sigma}_{\varepsilon_t}^{\,2} = \hat{\sigma}_{\varepsilon_t} \; 1 \; \hat{\sigma}_{\varepsilon_t} \quad \text{(caso univariado)} \tag{3.12}$$

A matriz *Z* (desvio-padrão) multiplicada pelo vetor de erros constrói-se um vetor normal-padrão $u_t (n \text{ x } 1)$ para o caso univariado, a cada instante *t*. ALBUQUERQUE (2008)

$$u_t \equiv Z^{-1}\varepsilon_t \tag{3.13}$$

De forma equivalente a (3.13), a multiplicação dos elementos do vetor u_t pelo escalar *Z* resulta no vetor ε_t, para cada *t*.

$$Zu_t = \varepsilon_t \tag{3.14}$$

A média de ε_t ainda é zero, pois os elementos de u_t foram sorteados de uma distribuição de normal-padrão e *Z* é uma constante. Desse modo, com a

decomposição da variância fora da amostra ($\hat{\sigma}_{\hat{\varepsilon}_{T+i}}{}^2$), foi possível transformar um vetor de choques normais padrão independentes u_t em um de choques ε_t na escala da série temporal supracitada no horizonte de previsão considerado. A equação vetorial (3.15) explicita, em termos matemáticos, o salientado.

$$\begin{bmatrix} y_{T+i,\ cenário\ 1} \\ . \\ . \\ . \\ y_{T+i,\ cenário\ n} \end{bmatrix} = \begin{bmatrix} \hat{y}_{T+i} \\ . \\ . \\ . \\ \hat{y}_{T+i} \end{bmatrix} + \hat{\sigma}_{\hat{\varepsilon}_{T+i}} \begin{bmatrix} u_1 \\ . \\ . \\ . \\ u_n \end{bmatrix}, \ i = 1,2...,h \qquad (3.15)$$

A cada sorteio, os resíduos na escala da série são substituídos na equação (3.15), obtendo, ao fim de *n* sorteios, a respectiva densidade preditiva, para *T+h*. Alguns testes foram realizados, na presente pesquisa, a fim de verificar se houve convergência: histograma, QQ-plot, PP-plot, teste de normalidade (software @Risk).

Especificamente para os dois modelos estatísticos adotados, para cada instante fora da amostra, o software estatístico forneceu o desvio-padrão estimado o qual foi multiplicado pela sequência de normal padrão, gerando os respectivos resíduos (no caso, geraram-se 1.000). Em seguida, estes foram somados a cada previsão, gerando os cenários (*out of sample*).

Para as redes neurais artificiais, adotou-se outro procedimento, quanto à geração dos cenários. O problema deveu-se ao fato de as RNA não possuírem um modelo explícito que possibilite a estimação do desvio padrão amostral fora da amostra, em função dos parâmetros, conforme os modelos estatísticos. Para o último elemento da validação, o procedimento foi similar aos modelos estatísticos, ou seja, calculo-se o desvio-padrão amostra dos resíduos até o instante relativo à última observação da validação e, então, aplicou-se a equação (3.15).

Cada cenário foi inserido na janela da rede neural, gerando 1.000 cenários para o período posterior (no caso, para o primeiro elemento da amostra de teste). Como a janela possui tamanho 5, para o segundo elemento da amostra de teste, utilizou-se o os quatro últimos pontos da validação e o primeiro do teste. Isso possibilitou gerar o cenário seguinte. Tanto o primeiro ponto de teste e quanto o

último da validação são variáveis, ou seja, cada respectivo cenário de ambos os pontos foi inserido na janela da rede neural de forma a gerar o cenário para o instante seguinte, onde os outros três são fixos. E assim sucessivamente até o décimo segundo passo. Como foi observado um crescimento da variância dos cenários (o que naturalmente ocorre), à medida que o horizonte ficava maior, adotou-se este procedimento para os fins da pesquisa. Além disso, foram feitos teste estatísticos para verificação de normalidade e todos (a 5% de significância) não rejeitaram a hipótese de normalidade, para todos os passos à frente projetados fora da amostra.

Tendo em vista a geração de cenários para os três métodos individuais, foi possível a combinação dos cenários, o que possibilitou a geração de intervalos de confiança dos métodos combinados, com 95% de credibilidade. Seguiram-se, assim, os procedimentos descritos, a seguir, para os modelos de combinação:

1. Linear: cada cenário individual foi linearmente combinado, gerando o combinado. Ao final, calculou-se a variância da densidade preditiva combinada e, então, procedeu-se à construção do intervalo de confiança, dada a previsão no respectivo instante;
2. Médias Simples: o mesmo procedimento, adotado para o caso linear, foi utilizado para a construção do intervalo de confiança do método de média simples; e
3. Geométrica: os cenários individuais foram combinados segundo a equação de combinação geométrica de previsões.

4
APLICAÇÕES À SÉRIE TEMPORAL DE CONSUMO RESIDENCIAL MENSAL DE ENERGIA ELÉTRICA

No capítulo 4, são aplicados os métodos individuais e os combinados à projeção de curto prazo da série de consumo residencial mensal de energia elétrica. A amostra contém cento e cinquenta e nove observações, sendo doze usadas para análise fora da amostra. Seguiram-se, à modelagem, três etapas, respectivamente: análise (quando necessária), modelagem e previsão. Os softwares utilizados foram: E-Views, FPW, @Risk, SPSS, Excel (*solver*, programação em VBA) e MATLAB (programação).

Após a exposição dos gráficos e tabelas, seus resultados foram comentados. Salienta-se ainda que alguns resultados foram explicitados somente na seção 5.5 (Comparação de Métodos).

Seguem-se, nesta ordem, as metodologias abordadas: ARIMA, MAE, RNA e combinações (linear, média e geométrica). O fluxograma da figura 4.1 descreve as etapas da metodologia.

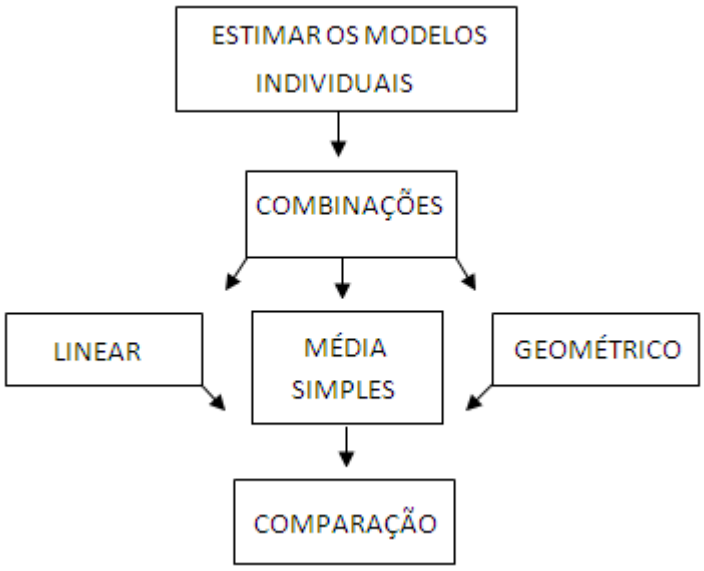

Figura 4.1- Fluxograma da Metodologia

4.1
Aplicação do Método ARIMA

A metodologia BOX & JENKINS impõe fortes restrições à série subjacente: estacionariedade de 2º ordem, normalidade e série de resíduos ruído branco. É fundamental que as mesmas sejam observadas para que propriedades estatísticas importantes e desejáveis do modelo não sejam perdidas.

4.1.1
Análise dos Dados

A análise dos dados de consumo, anteriormente à construção do modelo ARIMA, compõe-se, portanto, de dois diagnósticos:

- Normalidade dos dados: Análise de QQ-plot e estatísticas de teste de Kolmogorov-Smirnov; e
- Estacionariedade de 2º ordem: teste de raiz unitária (Dickey Fuller Aumentado).

4.1.1.1
Testes de Normalidade

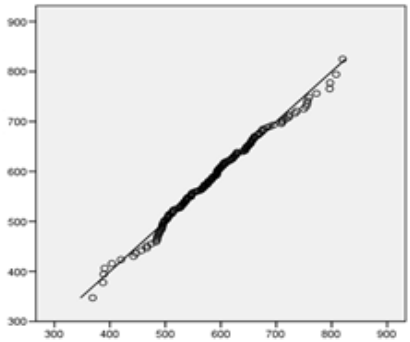

Gráfico 4.1 - QQ-plot da Série Consumo Residencial Mensal de Energia Elétrica

Nota-se que a maioria dos pontos se encontra sobrepostos à reta, dando indícios de normalidade dos dados.

Tabela 4.1 - Teste de Normalidade Kolmogorov-Smirnov

	Kolmogorov-Smirnov		
	Estatística *t*	Graus de liberdade	*P*-valor
Consumo	0,051	148	0,2

Como o p-valor é igual a 20% (sendo o nível de significância 5%), não se rejeita a hipótese normalidade.

4.1.1.2
Teste de Raiz Unitária

Na tabela 4.2, encontram-se os valores da estatística de teste DFA, sob hipótese nula de existência de raiz unitária.

Tabela 4.2 - Teste de Dickey-Fuller Aumentado

		Estatística *t*
Estatística de teste ADF		-5,847944
Valores críticos	1% de significância	-3,476143
	5% de significância	-2,881541
	10% de significância	-2,577514

Para os níveis de significância: 1%, 5% e 10%, a hipótese de tendência estocástica da série foi rejeitada. Portanto, as hipóteses de normalidade e estacionariedade de 2º ordem foram não foram rejeitadas, a 5% de significância.

4.1.2
Modelagem

O modelo ajustado, após terem sido testadas inúmeras ordens, foi o SARIMA (*1,0,0*) * (*1,0,3*). Destaca-se que a FAC e a FACP da série foi inicialmente analisada, porém visualmente não se caracterizou nenhum PE conhecido. Por isso, foram omitidas.

Tabela 4.3 - Parâmetros e Estatísticas do Modelo Ajustado

Termo	**Coeficiente**	**Erro padrão**	**Estatística *t***	**P-valor**
ar [1]	0,8184	0,0498	16,414	0,0000
sar [12]	0,3681	0,0882	4,1716	0,0001
sar [24]	0,6320	0,0882	7,1600	0,0000
sma [12]	-0,2720	0,1116	-2,4374	0,0163
sma [24]	-0,8959	0,0243	-36,8597	0,0000
sma [36]	0,2581	0,1032	2,5010	0,0138

Todas as estimativas dos parâmetros apresentam significância estatística, a 5% de significância, sendo estatisticamente diferentes de zero.

Tabela 4.4 - Principais Estatísticas de Aderência

R^2	81,03%	R^2 ajustado	80,02%
DW	2,1395	Ljung-Box (p-valor)	76,81%
MAPE	5,82%	BIC	4,43E+004

Analisando a tabela 4.4, verifica-se a não existência de dependência linear dos resíduos, dado as estatísticas de Durbin-Watson e Ljung-Box. Quanto à capacidade explicativa, obteve-se um nível de 81,05%, com o menor BIC e MAPE, dentre os possíveis modelos. Os correlogramas da FAC e FACP, no gráfico 4.2, confirmam as estatísticas de teste, visto que as autocorrelação encontram-se dentro do intervalo de confiança, sendo, portanto, estatisticamente não diferentes de zero.

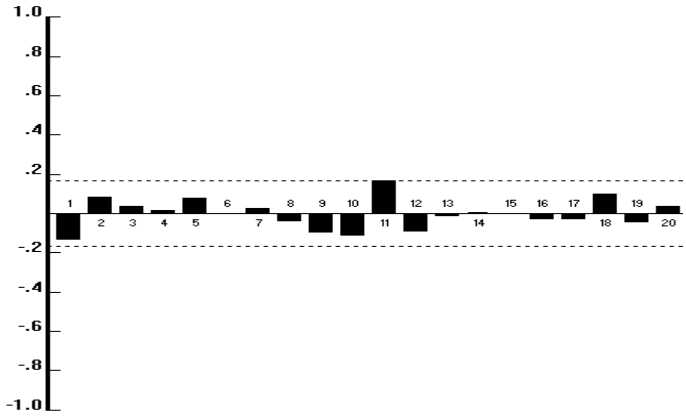

Gráfico 4.2 - Correlograma da FAC dos Resíduos

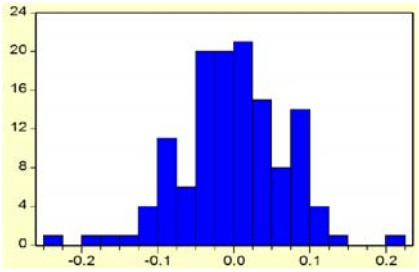

Gráfico 4.3 - Histograma dos Resíduos

Tabela 4.5 - Teste de Normalidade dos Resíduos

Jarque-Bera	P-valor	Curtose	Assimetria
4,713351	9,4735%	3,763440	-0,271137

O teste de Jarque-Bera e o gráfico 4.3 mostram que os resíduos são gaussianos (a 5% de significância). A curtose e a assimetria apontam para o mesmo resultado, pois assumiram valores próximos aos teóricos (*3* e *0*, respectivamente), caracterizando uma densidade Gaussiana. Por fim, foi feito o teste de DFA na série de resíduos para verificação de estacionariedade.

Tabela 4.6 - Teste de Dickey-Fuller Aumentado dos Resíduos

		Estatística *t*
Estatística de teste ADF		-11,77509
Valores críticos	1% de significância	-3,485115
	5% de significância	-2,885450
	10% de significância	-2,579598

A 1%, 5% e 10% de significância, os resíduos são fracamente estacionários. Portanto, a variância é homocedástica. O *arch test* foi feito a fim de verificar evidência de *volatilidade*, porém se rejeitou esta hipótese.

4.1.3
Previsões

O gráfico 4.4 descreve as estimativas pontuais ajustadas (dentro e fora da amostra), valores reais e intervalo de confiança (fora da amostra).

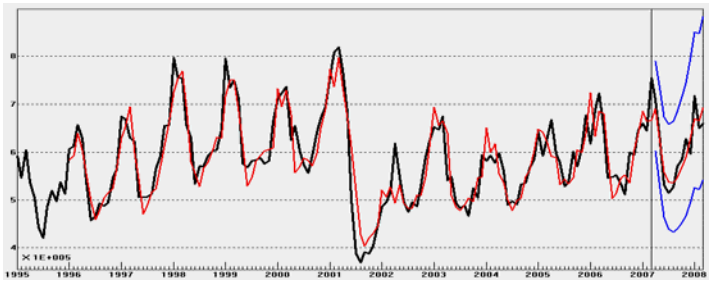

Gráfico 4.4 - Série Temporal e Previsões Pontuais e Intervalares.

Tabela 4.7 - MAPE's do ARIMA

Dentro da Amostra	Fora da Amostra
5,66%	7,88%

4.2
Aplicação do Modelo Holt-Winters

Como não há restrições quanto à sua utilização, consideraram-se as estatísticas R^2 e MAPE para a escolha do melhor modelo.

4.2.1
Modelagem

Inicialmente, testou-se o modelo com sazonalidade aditiva, visto que a série é homocedástica, e o *damped trend*, porém não obtiveram o melhor ajuste. O modelo com melhor ajustamento foi o Holt-Winters com sazonalidade multiplicativa. A seguir, características estatísticas do modelo estimado.

Tabela 4.8 - Valores dos Hiperparâmetros das Componentes

Componente	Hiperparâmetro
Nível	0.51347
Sazonalidade	0.49758

Tabela 4.9 - Valores dos Fatores Sazonais

Datas	Valores	Valores	Valores
Janeiro – Março	1.07286	1.10144	1.18919
Abril – Junho	1.12426	1.01458	0.94362
Julho – Setembro	0.89065	0.87161	0.89276
Outubro –	0.94374	0.96591	1.04651

4.2.2
Previsões

O gráfico 4.5 descreve as estimativas pontuais (dentro e fora da amostra), valores reais (dentro e fora da amostra) e intervalo de confiança (fora da amostra).

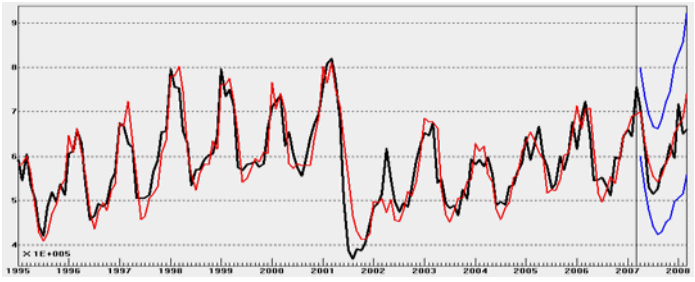

Gráfico 4.5 - Série temporal e Previsões Pontuais e Intervalares.

Tabela 4.10 - MAPE's do Modelo Holt-Winters

Dentro da Amostra	Fora da Amostra
5,82%	5,40%

4.3
Aplicação do Modelo de Redes Neurais Artificiais

Os critérios de escolha da RNA foram: MAPE, U-Theil. O código em MATLAB considerou noventa e seis configurações de redes, nas quais variaram

os parâmetros: funções de ativação, números de neurônios na camada escondida, padrões de entrada, algoritmo de treinamento e tamanho da janela. A melhor RNA possui as seguintes características:

- Tamanho da janela: 5;
- Padrão de entrada: *premnmx*:
- Número de camadas escondidas: 1;
- Algoritmo de treinamento: Levenberg-Marquardt (*trainlm*); e
- Número de neurônios na camada escondida: 5.

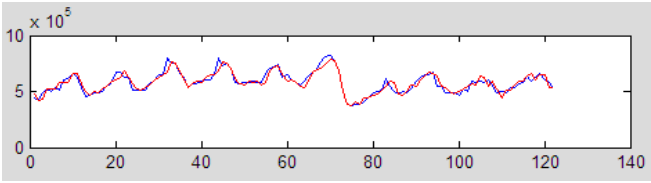

Gráfico 4.6 – Amostra (linha azul) e Previsões (linha vermelha) Dentro da Amostra

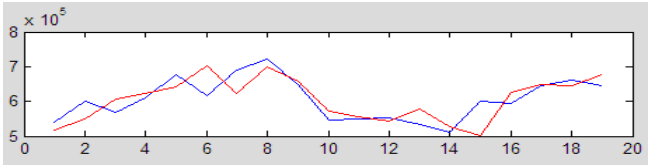

Gráfico 4.7 – Amostra (linha azul) e Previsões (linha vermelha) para Validação

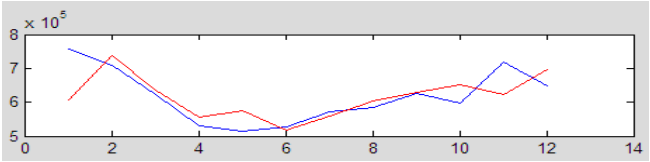

Gráfico 4.8 - Amostra (linha azul) e Previsões (linha vermelha) Fora da Amostra

Tabela 4.11 - MAPE's da RNA

Treino	Validação	Teste
4,36%	5,39%	7,01%

Aspectos relevantes como super-treinamento e capacidade de generalização, foram considerados. As amostras de validação e teste foram compostas, respectivamente, por seis e doze valores. As outras observações compuseram a amostra de treino.

4.4
Combinações dos Modelos

Combinaram-se as previsões pontuais oriundas dos modelos de RNA's, ARIMA e Holt-Winters. Como mencionado anteriormente, a função utilidade considerada foi o MAPE (*in sample*). As restrições e as justificativas atinentes aos pesos encontram-se na secção 3.3 (capítulo e). Os resultados da estatística de aderência das combinações encontram-se na seção 4.5 e subseções 4.5.1 e 4.5.2.

4.5
Comparação dos Modelos

Tabela 4.12 - Previsões dos Modelos Combinados e Valores Históricos

Data	Média Simples	Combinação Linear	Combinação Geométrica	Histórico
2007-12	704.373,80	705.924,64	698.840,73	708.970,08
2008-01	635.482,48	636.219,94	629.177,88	621.971,26
2008-02	571.079,09	558.901,86	553.330,10	531.089,20
2008-03	537.713,86	522.934,12	513.370,11	514.755,08
2008-04	559.897,50	566.244,47	554.748,89	527.584,12
2008-05	582.211,77	594.804,75	582.535,15	570.372,51
2008-06	592.510,00	593.691,08	586.504,64	585.272,81
2008-07	619.386,69	629.484,38	621.126,47	626.854,77
2008-08	619.766,17	599.962,62	597.689,15	596.979,50
2008-09	713.674,03	746.189,11	734.155,25	717.862,02
2008-10	645.633,87	626.630,27	633.326,40	649.547,68
2008-11	718.813,64	714.862,27	714.017,40	659.867,17

Comparando-se os três métodos, nota-se que o modelo geométrico obteve maior acurácia em sete lag's (em destaque). Em relação ao linear, exceto em dois meses, o geométrico obteve valores mais próximos aos reais. Contra o método de média simples, em oito lag's. Em 4.13, 4.14 e 4.15 expõem-se: APE, MAE e R^2

Tabela 4.13 – Erro Percentual Absoluto (APE) dos Modelos Combinados

Data	Média Simples	Combinação Linear	Combinação Geométrica
2007-12	0,65%	0,43%	1,43%
2008-01	2,17%	2,29%	1,16%
2008-02	7,53%	5,24%	4,19%
2008-03	4,46%	1,59%	0,27%
2008-04	6,12%	7,33%	5,15%
2008-05	2,08%	4,28%	2,13%
2008-06	1,24%	1,44%	0,21%
2008-07	1,19%	0,42%	0,91%
2008-08	3,82%	0,50%	0,12%
2008-09	0,58%	3,95%	2,27%
2008-10	0,60%	3,53%	2,50%
2008-11	8,93%	8,56%	8,21%

O modelo de combinação geométrica possui quatro de seus valores APE's inferiores a 1%, enquanto o linear e o de média simples, isso ocorre três vezes. A tabela 4.14 mostra que o valor de APE mínimo pertencente ao geométrico. No ponto de maior APE (lag 12), comum aos três métodos de combinação, o modelo geométrico também possui o menor valor percentual.

Tabela 4.14 - APE's dos Métodos Combinados

Mínimo	0,58%	0,42%	0,12%
Máximo	8,93%	8,56%	8,21%

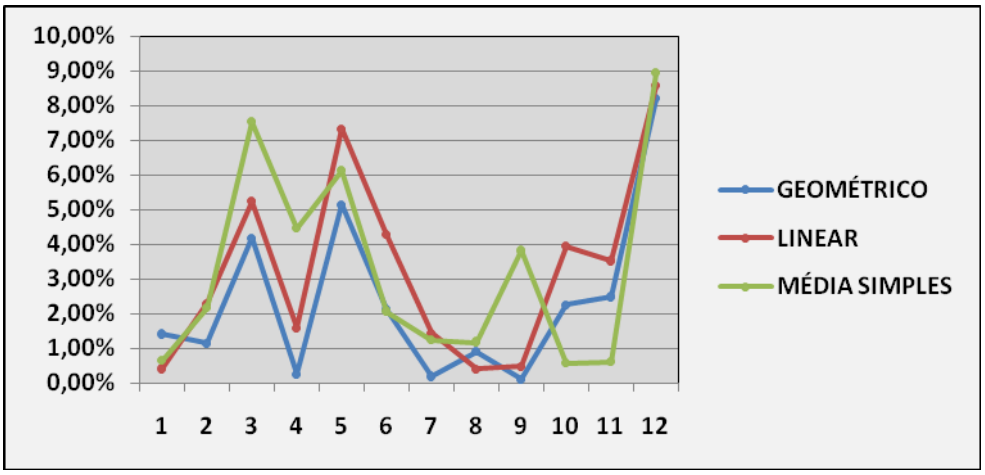

Gráfico 4.9 - Evolução dos APE`s das Combinações

O gráfico 4.9 mostra que, exceto no lag 1, o valor de APE do geométrico foi inferior a, pelo menos, um dos métodos combinados (a linha azul, a partir do lag 2, sempre está abaixo de, pelo menos, uma das linhas). A tabela 4.13 confirma este fato, em valores.

Tabela 4.15 - MAPE's dos Modelos Estimados

Métodos	**Dentro da Amostra**	**Fora da Amostra**
ARIMA	5,66%	7,88%
Holt-Winters	5,82%	5,40%
RNA	4,36%	7,02%
Média Simples	3,09%	3,28%
Combinação Linear	2,30%	3,30%
Combinação Geométrica	2,42%	2,48%

O valor de MAPE do modelo geométrico foi inferior a todas as outras metodologias (fora da amostra). Além disso, foi o único a apresentar valor abaixo de 3%, dentro e fora da amostra.

Tabela 4.16 - MAE's dos Modelos Estimados

Métodos	Dentro da Amostra	Fora da Amostra
ARIMA	32.684,26	45.354,33
Holt-Winters	32.760,43	32.765,21
RNA	26.801,18	37.057,95
Média Simples	17.586,29	19.145,75
Combinação Linear	13.231,77	19.720,75
Combinação Geométrica	13.832,50	14.551,98

O modelo geométrico possui menor valor de MAE, fora da amostra, mostrando maior capacidade de generalização.

Tabela 4.17 – Coeficiente de Explicação (R^2) dos Modelos Estimados

Métodos	Dentro da Amostra	Fora da Amostra
ARIMA	81,03%	38,40%
Holt-Winters	79,11%	66,13%
RNA	85,13%	53,54%
Média Simples	93,12%	84,81%
Combinação Linear	95,53%	84,94%
Combinação Geométrica	95,27%	90,00%

Considerando o coeficiente de explicação, tem-se que o modelo geométrico foi superior. Em relação ao linear (o segundo melhor), fora da amostra foi superior a 5,06%. Assim, foi capaz de explicar 90,00% da variabilidade fora da amostra.

Mais especificamente, foram feitas análises individuais contrapondo o modelo geométrico contra: combinação linear e média simples, respectivamente.

4.5.1
Comparação entre os Métodos Geométrico e Linear

Tabela 4.18 - MAPE`s dos Modelos

Métodos	MAPE	
	DENTRO	FORA
Combinação Linear	2,30%	3,30%
Combinação Geométrica	2,42%	2,48%

O método de combinação linear foi, em media percentual, 0,12% melhor dentro da amostra. Fora da amostra, o geométrico foi, em média, 0,82% superior.

Fora da amostra, o desempenho percentual relativo foi 33,06% superior ao linear.

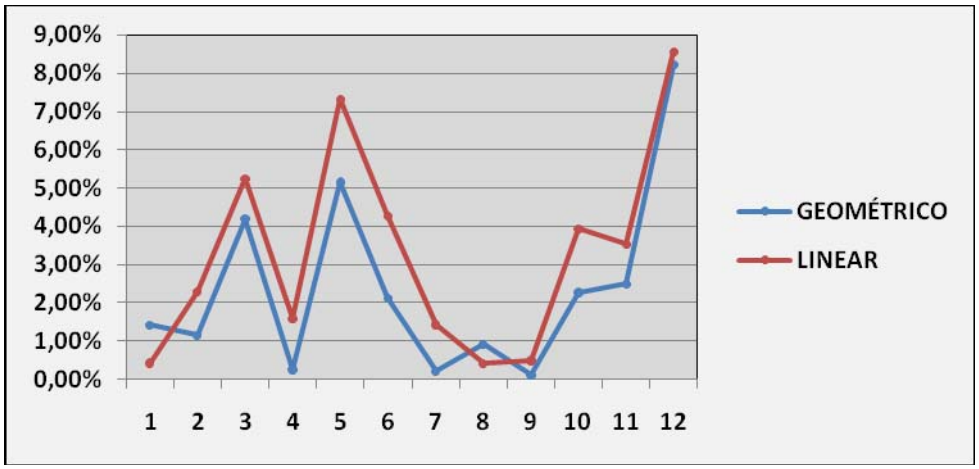

Gráfico 4.10 - Evolução Temporal dos APE`s

Destaca-se que o modelo de combinação geométrica obteve valores de APE mais próximos de zero em 84% do tempo considerado.

4.5.2
Comparação entre os Métodos Geométrico e Média Simples

Tabela 4.19 - MAPE`s dos Modelos

Métodos	MAPE	
	DENTRO	FORA
Média Simples	3,09%	3,28%
Combinação Geométrica	2,42%	2,48%

Os valores de MAPE do modelo geométrico, dentro e fora da amostra, são inferiores. Respectivamente, os valores são, em média, 0,67% e 0,80% inferiores. Em termos percentuais, foi 27,68% melhor *in sample* e 32,25%, *out of sample*.

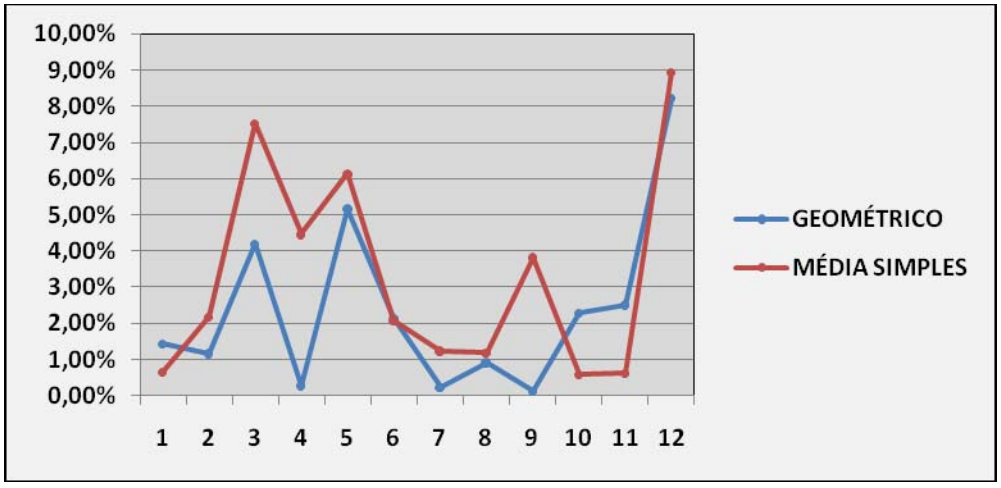

Gráfico 4.11 - Evolução Temporal dos APE`s

A evolução temporal dos APE's do método geométrico assume valores menores, na maioria do tempo, em relação à média simples. Assim, em oito lag's, são mais próximos de zero.

Analisando-se os gráficos 4.13 e 4.14, verifica-se que a combinação geométrica obteve melhor desempenho que os outros modelos, fora da amostra, quanto à estatística MAPE. Dentro da amostra, foi sensivelmente inferior ao linear

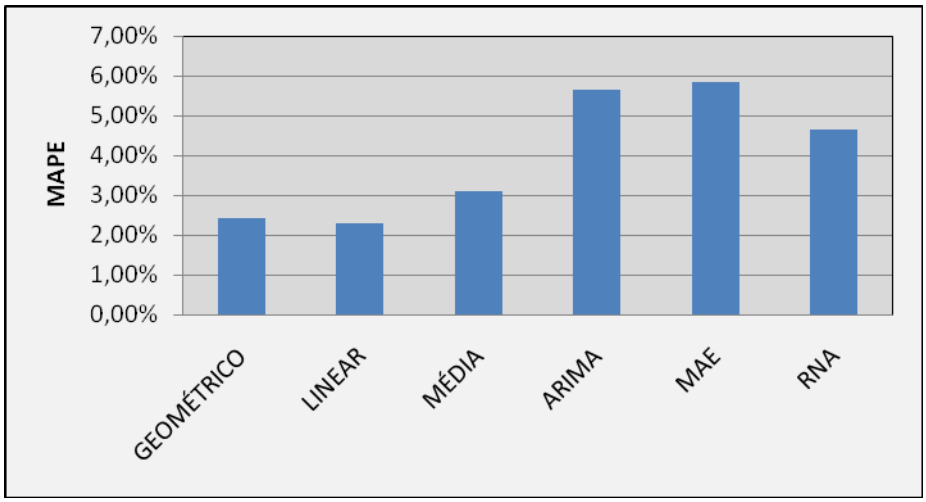

Gráfico 4.12- Valores dos MAPE`s dos Métodos (Dentro da Amostra)

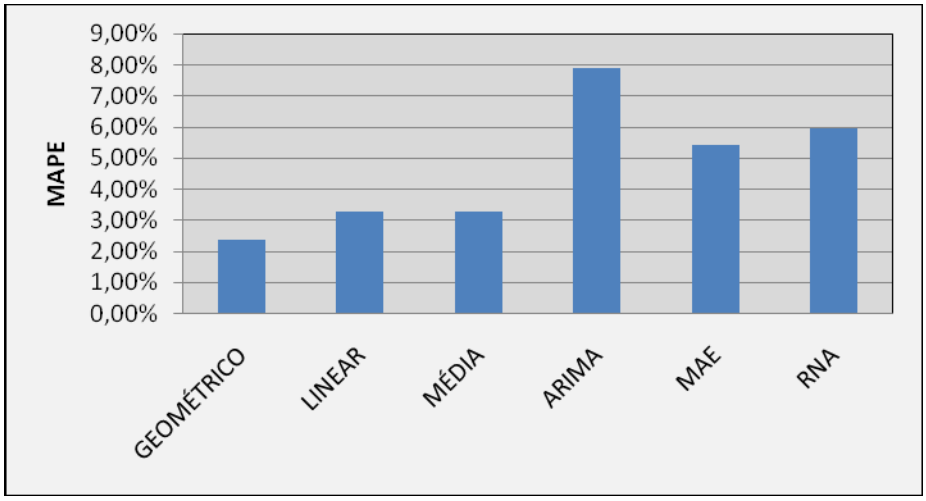

Gráfico 4.1.3 - Valores dos MAPE`s dos Métodos (Fora da Amostra)

Por sua vez, intervalos de confiança dos métodos foram construídos e comparados na seção 4.5.3.

4.5.3
Comparação dos Intervalos de Confiança dos Modelos Combinados

Os modelos utilizados à análise de intervalos de confiança foram somente os modelos combinados, visto suas, na maioria dos períodos, que suas variâncias foram menores que os individuais.

A maneira como foram construídos encontram-se na seção 3.4.4.

As tabelas 4.20, 4.21 e 4.22 explicitam, respectivamente, as combinações Linear, Geométrica e de Média Simples.

Tabela 4.20 - Intervalos de Confiança da Combinação Linear e Amplitude dos Limites do Intervalo de Confiança em Valores Absolutos.

Instante	Limite Inferior	Limite Superior	Amplitude Absoluta
1	654.511,80	757.337,49	102.825,69
2	572.957,30	699.482,58	126.525,28
3	485.596,38	632.207,33	146.610,95
4	437.577,92	608.290,33	170.712,41
5	473.859,45	658.629,49	184.770,04
6	496.363,64	693.245,85	196.882,21
7	487.249,85	700.132,31	212.882,47
8	517.027,82	741.940,95	224.913,12
9	478.012,73	721.912,50	243.899,77
10	619.381,27	872.996,94	253.615,67
11	486.674,15	766.586,40	279.912,25
12	555.988,67	873.735,87	317.747,20

Destaca-se que todos os valores reais, para o horizonte considerado, fora da amostra, encontram-se dentro do intervalo de confiança.

Além disso, o intervalo de confiança da combinação linear, na maioria dos lag's (fora da amostra), obteve amplitudes menores que os modelos individuais.

Tabela 4.21 - Intervalos de Confiança da Média Simples e Amplitude dos Limites do Intervalo de Confiança em Valores Absolutos.

Instante	Limite Inferior	Limite Superior	Amplitude Absoluta
1	646.130,09	762.617,51	116.487,42
2	572.659,13	698.305,83	125.646,70
3	505.456,85	636.701,33	131.244,48
4	447.827,20	627.600,52	179.773,32

5	465.371,43	654.423,57	189.052,14
6	484.129,11	680.294,44	196.165,33
7	487.905,21	697.114,78	209.209,57
8	510.573,55	728.199,83	217.626,28
9	502.080,84	737.451,49	235.370,65
10	592.621,31	834.726,74	242.105,43
11	517.407,15	773.860,60	256.453,45
12	576.010,92	861.616,36	285.605,44

Todos os valores reais se encontram no intervalo de confiança fora da amostra, considerando o horizonte previsão adotado.

As amplitudes absolutas foram, na maioria dos instantes, inferiores aos modelos individuais (fora da amostra).

Tabela 4.22 – Intervalos de Confiança da Combinação Geométrica e a Amplitude dos Limites do Intervalo de Confiança em Valores Absolutos.

Instante	**Limite Inferior**	**Limite Superior**	**Amplitude Absoluta**
1	646.450,70	751.230,76	104.780,06
2	566.451,26	691.904,50	125.453,24
3	479.450,77	627.209,42	147.758,66
4	425.257,31	601.482,91	176.225,60
5	465.062,80	644.434,98	179.372,18
6	489.343,39	675.726,92	186.383,53
7	483.429,62	689.579,66	206.150,04
8	515.683,51	726.569,44	210.885,92
9	473.163,39	722.214,90	249.051,51
10	606.314,34	861.996,15	255.681,81
11	495.452,42	771.200,39	275.747,97
12	567.610,13	860.424,67	292.814,53

De igual modo ao modelo linear e ao de média simples, todos os valores reais, para o horizonte considerado, fora da amostra, encontram-se dentro do intervalo de confiança do modelo geométrico.

O intervalo de confiança da combinação geométrica também obteve amplitudes menores que os modelos individuais, na maioria dos instantes.

Em relação ao linear, as amplitudes absolutas também foram menores em nove instantes de tempo. No caso da média simples, em sete.

O geométrico, portanto, produz menos incerteza quanto ao futuro, visto que suas distribuições preditivas normais produzem menos incerteza (na maioria dos lag's), visto suas amplitudes de seus intervalos de confiança, dada a série temporal supracitada.

Estes resultados podem ser verificados na tabelas 4.23.

Tabela 4.23 - Amplitudes dos Limites do Intervalo de Confiança dos Modelos Combinados em Valores Absolutos.

Instante	Geométrico	Linear	Média
1	104.780,06	102.825,69	116.487,42
2	125.453,24	126.525,28	125.646,70
3	147.758,66	146.610,95	131.244,48
4	176.225,60	178.712,41	179.773,32
5	179.372,18	184.770,04	189.052,14
6	186.383,53	196.882,21	196.165,33
7	206.150,04	212.882,47	209.209,57
8	210.885,92	224.913,12	217.626,28
9	249.051,51	243.899,77	235.370,65
10	255.681,81	253.615,67	242.105,43
11	275.747,97	279.912,25	256.453,45
12	292.814,53	317.747,20	285.605,44

No gráfico 4.13, descreve-se a evolução temporal das observações e das previsões pontuais, dentro da amostra.

No gráfico 4.14, explicita-se a série temporal de consumo residencial mensal de energia elétrica observada, fora da amostra, bem como suas respectivas

previsões pontuais e o intervalo de confiança do modelo de combinação geométrica.

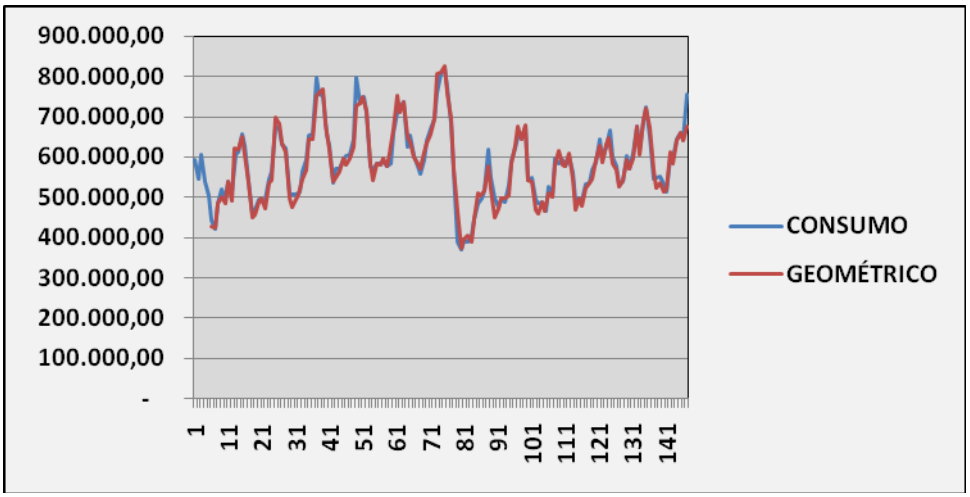

Gráfico 4.14 – Valores Reais e as Previsões Pontuais (*in sample*)

Nota-se que o método geométrico possui um bom desempenho quanto à dinâmica passada observada da série temporal de consumo, pois as curvas encontram-se, visualmente, a maior parte do tempo sobrepostas à das observações.

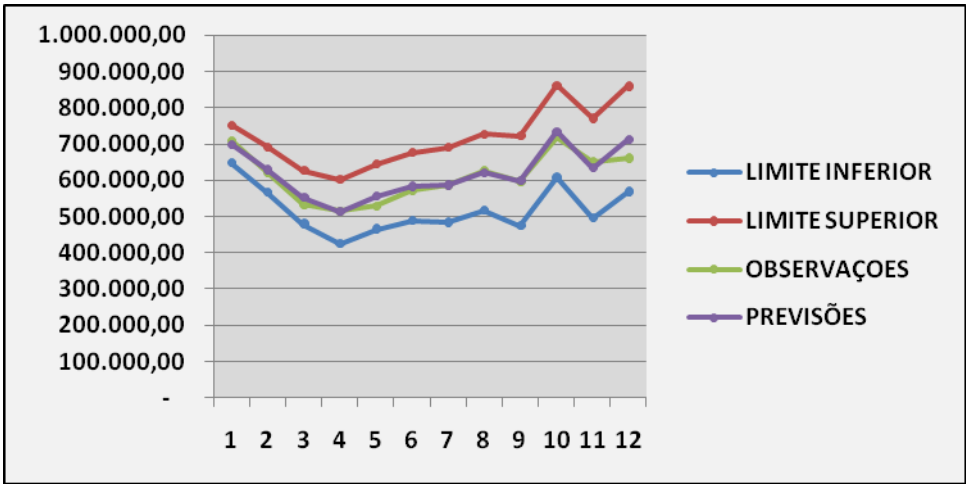

Gráfico 4.15 – Valores Reais e as Previsões Pontuais e Intervalares do Modelo de Combinação Geométrica

Analisando visualmente, alguns pontos, no gráfico 4.14, merecem ser comentados:

i. Nota-se que os pontos dos valores reais se encontram sobrepostos ao das previsões pontuais durante boa parte do horizonte considerado, mostrando novamente acurácia das previsões pontuais geométricas;

ii. Quanto aos intervalos de confiança, verifica-se que a amplitude dos mesmos não cresce de maneira exacerbada, ou seja, a variância (incerteza) não possui um crescimento exagerado, à medida que se aumentam os lag's; e

iii. Ainda segundo os intervalos de confiança, os mesmos mostram certa simetria entre os valores reais e os limites, superior e inferior - que é sempre desejável. Isso indica que a densidade preditiva gaussiana do modelo geométrico contém valores reais muito próximos seu centro de massa, ou seja, o evento que possui probabilidade máxima de acontecer, segundo o modelo, é muito próximo ao valor real, confirmando i. Pode-se dizer também que o valor real encontra-se entre os valores mais prováveis desta densidade preditiva, mostrando que os cenários produzidos são eficientes.

Em última análise, destaca-se que foram combinadas três modelos-base. Quando inseridos mais métodos, ocorreu perda de acurácia das previsões combinadas. Além disso, modelos multivariados (VEC) e causais (Regressão Dinâmica e Função de Transferência) foram testados, mas as outras variáveis disponíveis possuíam baixa correlação com a série de consumo residencial de energia elétrica. Embora tenham contribuído à modelagem, quando utilizados três métodos, não ficaram entre os três melhores métodos individuais.

Como mencionado anteriormente, as restrições impostas aos pesos adaptativos, bem como a constante de ajuste (caso geométrico), deveram-se à observância de perda de desempenho quando não assumidas, ainda que sensível em alguns casos. O mesmo critério foi válido à escolha da função utilidade MAPE *in sample*, que obteve desempenho superior (fora da amostra) à função MSE.

5
CONCLUSÃO

No presente estudo, combinaram-se modelos preditivos univariados aplicados à previsão pontual e intervalar da série de consumo residencial mensal de energia elétrica. Os três modelos combinados foram superiores aos individuais, quanto às medidas de acurácia consideradas. Em particular, a combinação geométrica de modelos individuais mostrou-se superior aos individuais e aos de combinação linear e de média simples.

Além da combinação dos três métodos individuais, outras foram testadas. Por exemplo, o ARIMA com MAE, sem restrições de pesos adaptativos. Em todas as possíveis, o método geométrico obteve melhor desempenho em, pelo menos, a maioria das estatísticas de aderência consideradas. Muito embora os outros resultados tenham sido também satisfatórios, mostraram-se apenas os das combinações dos três modelos individuais com restrições, devido ao melhor desempenho fora da amostra.

Devido à contribuição de cada método individual nos métodos combinados, cada um fornece informações distintas acerca da dinâmica temporal da série analisada. Assim, cada método individual funciona como uma componente do modelo geométrico combinado responsável por fornecer diferentes informações da série de tempo:

- Modelo ARIMA: fornece informações acerca da estrutura de dependência linear simples e sazonal presentes na série;
- Modelo Holt-Winters: contribui com informações atinentes à variabilidade e sazonalidade temporal; e
- Modelo de Redes Neurais Artificiais: fornece informações contidas na não-linearidade da série.

Assim sendo, o modelo combinado torna-se um agregador de informações, pois um modelo individual pode não capturar determinadas estruturas temporais.

Afinal, podem existir informações na estrutura de dependência linear, não-linear e / ou na frequência da série temporal relevantes à modelagem temporal. O modelo ARIMA, por exemplo, não captura estruturas não lineares de dependência, porém as RNA o fazem. Logo, uma ponderação adequada pode fornecer melhores previsões, pois, em um mesmo modelo, informações sobre as estruturas de dependência linear e não linear que caracterizam o processo estocástico gerador podem ser simultaneamente capturadas.

Quanto ao número (saturação) de modelos utilizados e às suas características (que fazem com que os mesmos forneçam informações úteis e independentes contidas na amostra), foram sustentados nos resultados de LIBBY & BLASHFIELD (1978) e MAKRIDAKIS & WINKLER (1983), respectivamente.

Dentre os modelos combinados, o geométrico possui a vantagem de reproduzir aproximadamente o efeito que o de combinação linear produz. O contrário, porém, não é possível. Assim, a combinação linear calcula (quando os pesos são normalizados e a soma é igual a uma unidade) o valor esperado das previsões, somente. O geométrico, por sua vez, é capaz de aproximá-lo e, além disso, é flexível quanto à combinação não linear de previsões. Este fato é relevante, visto que a melhor combinação de previsões pode não ser a linear (como no caso proposto).

Outro ponto importante é que algum modelo pode individualmente apresentar desempenho melhor que os outros em determinados períodos. Suponha que a RNA seja mais precisa nos primeiros quatro lag's; o ARIMA, nos três subseqüentes, e o MAE, nos últimos. O método geométrico é capaz de agregar estes desempenhos, de forma a descrever e generalizar a dinâmica temporal, com maior eficácia, e, ao mesmo tempo, atenuar determinadas deficiências contidas nos métodos individuais. Em síntese, o método agrega os pontos favoráveis dos modelos individuais e atenua os desfavoráveis, simultaneamente.

Além disso, considerar pesos ótimos foi importante, pois possibilitou testar a minimização de outra estatística de maneira simples, rápida e eficaz. Embora exista correlação entre as estatísticas MAPE e MSE, o ponto de mínimo de uma, não necessariamente é o da outra. Minimizar o MSE não significa ter a melhor combinação de pesos. Como verificado, a minimização do MAPE forneceu maior

poder acurácia do modelo de combinação geométrica fora da amostra. Ou seja, o ponto de mínimo do MSE não foi ótimo. Evidentemente, existem situações que o MSE pode ser melhor. Esta restrição é flexível e os pesos ótimos possibilitam esta flexibilidade.

Em última análise, o método proposto possibilitou que, mesmo considerando modelos individuais com desempenhos apenas razoáveis, fosse possível produzir previsões consideravelmente mais acuradas fora da amostra. O que reforça a afirmação de que o modelo geométrico é um agregador de informações.

5.1
Sugestões de Estudos Futuros

Diante dos resultados desta dissertação, inúmeras possibilidades podem ser estudadas e aplicadas em outros estudos de casos. Eis alguns:

- O método pode ser aplicado a modelos de volatilidade. Pode-se, por exemplo, combinar geometricamente modelos da família GARCH, EWMA, Volatilidade Estocástica com Redes Neurais Artificiais, ponderados por pesos ótimos;

- No caso dos métodos individuais, pode-se usar o mesmo método combinado geometricamente, sendo modelado no domínio no tempo e na frequência. Suponha dois modelos ARIMA distintos. Um utilizando a série original (no domínio do tempo) e a outra transformada no método *Wavelet* (no domínio da frequência). Ainda é possível projetar os resíduos via redes neurais artificiais, utilizando a transformada *Wavelet*, ao invés de assumi-los com valores iguais a zero no futuro;

- Em aplicações de projeção de retornos de ativos financeiros, podem-se utilizar o método combinado geometricamente modelos não lineares de redes neurais artificiais com modelos causais e / ou multivariados.

Constantes aditivas podem ser inseridas no modelo. Em alguns casos, melhores resultados podem ser alcançados;

- A combinação geométrica de redes neurais artificiais, em muitos casos, fornece projeções consideravelmente mais acuradas;

- É interessante combinar geometricamente combinações lineares. O oposto também é sugestivo. Ambas as alternativas com ou sem constantes de ajuste;

- As médias das previsões oriundas das combinações geométrica e linear são outra alternativa, em muitos casos, pode ser eficiente. Os resíduos deste modelo podem ser projetados, utilizando-se um método individual ou um híbrido;

- A modelagem híbrida utilizando os modelos não-lineares *STAR* e as RNA, com ou sem projeção de resíduos, utilizando a combinação geométrica é interessante;

- Aplicação do método em modelos de alta frequência à série de preços e retornos com pesos fixos ou variáveis;

- O método pode utilizar modelos dinâmicos e redes neurais combinados, com pesos fixos ótimos;

- A utilização de previsões subjetivas (feita por especialista ou grupos de estudo) combinadas geometricamente com previsões objetivas (feita por modelos) é factível; e

- Outros métodos de otimização, bem como outras funções utilidades a serem otimizadas, podem ser testados. Por exemplo, utilizar algoritmo genético para estimação dos pesos adaptativos.

Portanto, estas e várias outras possibilidades podem ser estudadas e testadas em diferentes situações, podendo fornecer resultados mais acurados que os modelos utilizados individuais.

REFERÊNCIAS BIBLIOGRÁFICAS

ALBUQUERQUE, A.R. & SOUZA, R.C. (2008). **Fluxo de Caixa em Risco: Uma Nova Abordagem para o Setor de Distribuição de Energia.** Dissertação de Mestrado. PUC-RIO.

ANDRADE, T. & SÁFADI, G.A. **Abordagem Bayesiana de Modelos de Séries Temporais**. ESTE (2007).

ARMSTRONG, S.J. (2004). **Principles of Forecasting: a Handbook for Researchers and Practitioners. Massachusetts**: Eletronic Services, http://www.wkap.nl. Acesso: 15 de fevereiro 2009.

ARMSTRONG, J.S. **Principles of Forecasting: A Handbook for Researchers and Practitioners.** Kluwer Academic Publishers. 2001.

ARMSTRONG, J.S. Combining Forecasting. In: ARMSTRONG, J. S. **Principles of Forecasting: A Handbook for Researchers and Practitioners.** Kluwer Academic Publishers. 2001a.

ASKU, C. & GUNTER, S.I. **An Empirical Analysis of the Accuracy of AS, OLS, ERLS, and NRLS Combination Forecasts.** International Journal of Forecasting, v. 8, 1992, p.27-43.

BATCHELOR, R. & DUA, P. **Forecaster Diversity and the Benefits of Combining Forecasts.** Management Science, v.41, 1995, p.68-75. 151.

BATES, J.M. & GRANGER, C.W.J. **The Combining of Forecasts.** Operational Research Quarterly, v.20, n.4, 1969, p. 451-468.

BOX, G.E.P. & JENKINS, G.M. **Time Series Analysis. Forecasting and Control**. Holden-Day. Edição revisada. San Francisco, 1976.

BUNN, D.W. **A Bayesian approach to the linear combination of forecasts.** Operational Research Quarterly, 26, 1975, p.325-329.

CHAN, C.K.; KINGSMAN, B.G. & WONG, H. **The Value of Combining Forecasts in Inventory Management – a Case Study in Banking.** European Journal of Operational Research, v.117, 1999, p.199-210.

CAVALERI, R. **Combinações de Previsões Aplicadas à Volatilidade**. Dissertação de Mestrado. Departamento de Economia, UFRS.

CHAMBERS, J. C., MULLICK, S. K. & SMITH, D. D. **How to Choose the Right Forecasting Technique.** Harvard Business Review. July-August, 1971, p.45-74.

CLEMEN, R.T. **Combining Forecasts: A Review and Annotated Bibliography.** International Journal of Forecasting. v. 5, 1989, p.559-583.

CLEMEN, R.T. & WINKLER, R.L. **Combining Economic Forecasts.** Journal of Business and Economic Statistics. v.4, 1986, p.39-46.

DIAS, M.A.G. **Quasi-Monte Carlo Simulation**. Disponível em: <https://www.puc-rio.br/marco.ind/quasi_mc.html>. Acesso: 15 de abril 2009.

DICKINSON, J.P. **Some Comments on the Combination of Forecasts.** Operational Research Quarterly, v.26, 1975, p. 205-210.

FARIA, A. E., (1996). **Graphical Bayesian Models in Multivariate Expert Judgements and Conditional External Bayesianity,** University of Warwick, UK.

FARIA, A. E. & MUBWANDARIKWA, E., (2008). **Multimodality on the Geometric Combination of Bayesian Forecasting Models**, International Journal of Statistics and Management System, 3, 1-25

FARIA, A. E. & MUBWANDARIKWA, E., (2007). **The Geometric Combination of Bayesian Forecasting Models**, Journal of Forecasting.

FARIA, A. E., SOUZA, R. C., (1995). **A Re-evaluation of the Quasi-Bayes Approach to the Linear Combination of Forecasts,** Journal of Forecasting, 14, 533-542.

FAVA, V.L. **Metodologia de Box-Jenkins para Modelos Univariados In: Manual de Econometria**. Vasconcelos, M. A. S. & Alves, D. Editora Atlas, São Paulo, 2000.

FLORES, B.E. & WHITE, E.M. **A Framework for the Combination of Forecasts.** Journal Academic Marketing Science, v.16 (3-4), 1988, p.95-103.

FLORES, B.E. & WHITE, E.M. **Subjetive versus Objective Combining of Forecasts: An Experiment.** Journal of Forecasting, v.8, 1989, p.331-341.

GOODWIN, P. **Correct or Combine? Mechanically Integration Judgmental Forecasts with Statistical.** International Journal of Forecasting, v.16, 2000, p.261-275.

BATES, J. M. & GRANGER, C. W. J. **Combination Forecasts**. Operations Research Quarterly, 1969

GRANGER, C.W.J. & RAMANATHAN, R. **Improved Methods of Forecasting.** Journal of Forecasting, v.3, 1984, p.197-204.

GUJARATI, D. **Econometria Básica**. Makron Books. 3a Ed., São Paulo, 2000.

GUPTA, S. & WILTON, P.C. **Combination of Forecasts: An Extension.** Management Science. v.33, n.3, March, 1987, p.356-372.

HAMILTON, J. (1994). **Time Series Analysis**. Princeton University Press, 1994.

HAYKIN, S. **Redes Neurais: Princípios e Prática**. Trad. Paulo Martins Engel. 2.ed. Porto Alegre: Bookman, 2001.

HILL, R.C. & GRIFFITHS, W.E. e JUDGE, G.G. (1999). **Econometria**. São Paulo: Saraiva, 1999.

HILL, C.; GRIFFITHS, W. & JUDGE, G. **Econometria**. Ed. Saraiva. São Paulo, 2000.

HOLLAUER, G. ; ISSLER, J. V.; NOTINI, H. H.. **Novo Indicador Coincidente para a Atividade Industrial Brasileira**. Revista de Economia Aplicada, 2008

JOHNSON, D. & KING, M. **Basic Forecasting Techniques**. Butterworth & Co., London, 1988.

KRYKOVA, I. **Evaluaring of Path-Depedent Securities with Low Discrepancy Methods.** Thesis. Faculty of the Worcester Polytechnic Institute, London (UK), 2003.

LEMOS, F.O. & FOGLIATTO, F.S. **Integração de Métodos Quantitativos e Qualitativos de Previsão para Desenvolvimento de um Sistema de Previsão de Demanda de Novos Produtos**. Tese de Doutorado. Departamento de Engenharia de Produção, PUC-RS, 2007.

LeSAGE, J.P. & MAGURA, M. **A Mixture-Model Approach to Combining Forecasts.** Journal of Business & Economic Statistics, v.10, n.4, 1992, p.445-452.

LIBBY, R. & BLASHFIELD, R. K. **Performance of a Composite as a Function of the Number of Judges.** Organizational Behavior and Human Performance, v.21, 1978, p.121-129.

LOBO, G.J. **Alternative Methods of Combining Security Analysts' and Statistical Forecasts of Annual Corporate Earnings.** International Journal of Forecasting, v.7, 1991, p.57-63.

LUTKEPOHL, H. (2006). **New Introduction to Multiple Time Series Analysis**. Springer.

MAHMOUD, E. **Combining Forecasts: Some Managerial Issues.** International Journal Forecasting, v.5, p.599-600.

MAHMOUD, E. **Combining Forecasts: Some Managerial Issues.** International Journal of Forecasting, v.5, 1989, p.599-600.MILLER, C.M.; CLEMEN, R.T. &

MAKRIDAKIS, S. **Why Combining Works?** International Journal of Forecasting, v.5, 1989, p.601-603.

MAKRIDAKIS, S., WHEELWRIGHT, S. C. & HYNDMAN, R. J. **Forecasting. Methods and Applications**. Third Edition. John Wiley & Sons. New York, 1998.

MAKRIDAKIS, S. & WINKLER, R.L. **Averages of Forecasts: Some Empirical Results.** Management Science, v.29, n.9, 1983, p.987-996.

MORETTIN, P.A. e TOLOI, L.M.C, **Análise Séries Temporais**, 2ª Ed. ABE Projeto Fisher, Ed. Edgard Blucher (2006).

MUBWANDARIKWA, E., (2007). **Modality Conditions and Prior Weights in the Geometric Combination of Bayesian Forecasting Models**. Thesis. The Open University Walton Hall Milton Keynes, London (UK), 2003.

NEWBOLD, P. & GRANGER, C.W.J. **Experience with Forecasting Univariate Time Series and Combination of Forecasts.** Journal Royal Statistical Society, series A, v.137, n.2, 1974, p.131-165.

NIEDERREITER H. **Random Number Generation and Quasi-Monte Carlo Methods**, CBMS-NSF, Vol. 63. Philadelphia: SIAM, 1992.

RAUSSER, G. C. & OLIVEIRA, R. A. **An Econometric Analysis of Wilderness Area Use.** Journal of the American Statistical Association, v.71, n.354, Application Section. June, 1976.

REEVES. G.R. & LAWRENCE, K.D. **Combining Multiple Forecasts given Multiple Objectives.** Journal of Forecasting, v.1, 1982, p.271-279.

RUSSELL, S.J. e NORVIG, P. **Artificial Intelligence: A Modern Approach**. New Jersey: Prentice-Hall Inc., 1995.

RIPLEY, B.D**. Stochastic Simulation**. John Wiley & Sons, Inc., 1987, 237 p

SOUZA, G.P. **Previsão do Consumo Industrial de Energia Elétrica no Estado de Santa Catarina: Uma Aplicação da Combinação de Previsões entre Modelos Univariados e de Regressão Dinâmica.** Dissertação de Mestrado. Departamento de Engenharia Elétrica, Santa Catarina, 2005.

SOUZA, R.C. & CAMARGO, M.E.. **Análise e Previsão de Séries Temporais: Os Modelos Arima**. 1. ed. Brasil: Inijuí, Ijuí, RS, 1996. 220 p.

SOUZA, R.C. & CAMARGO, M.E. **Análise e Previsão de Séries Temporais: Os Modelos Arima** - 2a. edição. 2a. ed. Rio de Janeiro: Gráfica e Editora Reginal, 2004. v. 01. 187 p.

TAFNER, M.A. **Redes Neurais Artificiais: Introdução e. Princípios de Neurocomputação**. - Blumenau : EKO, 1996.

WEBBY, R. & O'CONNOR, M. **Judgement and Statistical Time Series Forecasting: a Review of the Literature.** International Journal of Forecasting, 12, 1996, p.91-118.

WINKLER, R.L. **Combining Forecasts: A Philosophical Basis and Some Current Issues.** International Journal of Forecasting, v.5, 1989, p. 605-609.

WINKLER, R.L. **The Effect of Nonstationarity on Combined Forecasts.** International Journal of Forecasting, v.7, 1992, p.515-529.

WINKLER, R.L. & MAKRIDAKIS, S. **The Combination of Forecasting.** Journal of the Royal Statistical Society, series A, v.146, 1983, p.150-157.

ZANDONADE, E. (1993), **Aplicação da Metodologia de Redes Neurais em Previsão de Séries Temporais**. Dissertação de Mestrado, Departamento de Engenharia Elétrica, PUC -RJ.

ZOU, H. & YANG, Y. **Combining Time Series Models for Forecasting.** International Journal of Forecasting, v.20, n.1, 2004, p.69-84.

APÊNDICE A

A.1
Operadores de Séries Temporais

Alguns operadores que facilitam a expressão matemática dos modelos construídos à modelagem de séries temporais.

De acordo com CAMARGO e SOUZA (2004), para melhor defini-los, os mesmos são dados por:

- Operador de retardo ou de translação para o passado: representa a defasagem de *k* períodos de tempo para trás. É denotado pelo B, definido por: $B^k = Z_{t-k}$;

- Operador de avanço ou translação para o futuro: representa uma defasagem de *k* períodos de tempo para frente. É denotado pelo F, definido por: $F^k = Z_{t+k}$;

- Operador diferença: é a transformação nos dados que consiste em tomar diferenças sucessivas da série de tempo original. É denotado por $\nabla^d{}_s = (1-B^s)^d$, sendo *s* o número de períodos de *d*, o de diferenças;

- Operador soma ou inverso: é denotado por *S*, definido por: $S^d{}_s = (\nabla^d{}_s)^{-1}$

Exemplificando, assuma a equação: $\nabla^d{}_s = (1-B^s)^d$, para $\nabla = (1-B)$. Assim, $W_t = \nabla^d{}_s Z_t = (1-B^s)^d Z_t$. Para $s = 1$ e $d = 1$, tem-se a primeira diferença. Portanto: $\nabla^1{}_1 W_t = W_t - W_{t-1} = Z_t + 2 Z_{t-1} + Z_{t-2}$

A.2
Processos Estocásticos Ergódicos

Um processo estocástico é dito ergódico se apenas uma realização do mesmo é o suficiente para se obter todas as estatísticas que o caracterizam. Este resultado é conhecido como *teorema ergódigo*. Assim, o mesmo pode ser subdividido em duas classes: *estritos e amplos*. Tais seguem as classificações de estacionariedade. Portanto, um processo ergódigo é também estacionário, visto que uma realização de um não estacionário não poderá conter todas as informações necessárias para sua especificação. Mas nem todo processo estocástico estacionário é necessariamente ergódico.

Os conceitos de ergodicidade na média e de 2º ordem são importantíssimos, visto que equivalem à consistência fraca dos estimadores da média e covariância, respectivamente:

- Ergodicidade na média: $\frac{1}{T}\sum_{t=1}^{T} y_t \xrightarrow{p} \mu$; e (A.1)
- Ergodicidade de 2º ordem: $\frac{1}{T-k}\sum_{t=k+1}^{T} (y_t - \overline{y})\,(y_{t-k} - \overline{y}) \xrightarrow{p} \gamma_i$. (A.2)

A.3
Medidas de Aderência

As medidas de aderência adotadas, no capítulo 4, para a comparação dos métodos, dentro e fora da amostra:

- Erro: $e_t = y_t - \hat{y}_t$; (A.3)
- Erro Absoluto Médio: $MAE_n = \sum_{t=1}^{n} \frac{1}{n}\left|y_t - \hat{y}_t\right|$; e (A.4)
- Erro Percentual Absoluto Médio: $MAPE_n = \sum_{t=1}^{n} \frac{1}{n} * \left[\frac{\left|y_t - \hat{y}_t\right|}{y_t}\right]$; (A.5)
- Erro Percentual Absoluto: $APE_t = \left[\frac{\left|y_t - \hat{y}_t\right|}{y_t}\right]$; e (A.6)

- Coeficiente de Explicação: $R^2 = \left[1 - \frac{\sum_{t=1}^{n} e_t^{\,2}}{\sum_{t=1}^{n} (y_t - \bar{y})^2} \right]$, onde

$$\bar{y} = \left(\sum_{t=1}^{n} y_t \right) * \frac{1}{n}. \qquad \text{(A.7)}$$

APÊNDICE B

B.1
Teste de Raiz Unitária

Existem duas maneiras de verificar estacionariedade fraca. A primeira delas (informal, porém não menos utilizada) são os recursos gráficos: análise do gráfico da série e do correlograma da função de autocorrelação. Assim, caso o gráfico da autocorrelação apresente decaimento abrupto nos primeiros lag's, indica que existe evidência de estacionariedade do processo gerador. Caso contrário, o gráfico decai gradualmente na forma exponencial ou senoidal amortecida.

Na outra, aplica-se do teste de raiz unitária, que tem, comumente, como hipótese nula, a série não ser estacionária. Por exemplo, assuma um modelo AR (*1*) da equação (B.1), cujo polinômio característico é $(1 - \alpha B)$ com raiz $1/\alpha$ situada sobre o círculo unitário (mas não dentro), ou seja, $\alpha = 1$. O processo AR (*1*) torna-se um modelo não estacionário de passeio aleatório ou *randon walk*.

Assim, tem-se que:

$$Z_t = c + \alpha Z_{t-1} + \varepsilon_t \qquad \text{(B.1)}$$

Em que $\varepsilon_t \sim i.i.d\ (0, \sigma^2_{\varepsilon_t})$. A estacionariedade, por sua vez, pode ser evidenciada com a estabilidade do modelo, a qual se dá com $|\alpha| < 1$. ALEXANDER (2004) prova que $|\alpha| < 1$ gera estabilidade no modelo AR (*1*). Particularmente, o AR (*1*) depende da magnitude do parâmetro α, conforme aumenta em valor absoluto, a velocidade da reversão diminui.

Outro resultado interessante ocorre quando $\alpha = 0$, pois o processo $\{y_t\}$ torna-se um ruído branco, causando reversão à média instantânea. Por fim, se $|\alpha| > 1$, o PE é explosivo, isto é, $y_t \to \pm \infty$, conforme $t \to \infty$. Evidentemente, estende-se a condição de estabilidade do AR (*1*) para o AR (*p*), para *p > 1*, tendo seu polinômio característico $(1 - \alpha_1 B^1 - ... - \alpha_p B^p)$ com suas raízes reais e / ou a norma das complexas fora do círculo unitário para estabilidade do modelo.

Como os modelos populacionais são desconhecidos, fazem-se necessários testes estatísticos. Não obstante, para testar se $\alpha = 1$, não é suficiente estimar α e, posteriormente, usar o teste *t*. Pois existe sério viés no caso de não estacionariedade do processo gerador. (ALEXADER, 2004)

Segundo ALEXADER, 2004, usando-se o operador de primeira diferença, o viés é corrigido, de modo que o modelo pode ser reescrito como: $\nabla Z_t = c + (\alpha - 1)\ Z_{t-1} + \varepsilon_t$.

Reescrevendo novamente, tem-se:

$$\nabla Z_t = c + \beta\ Z_{t-1} + \varepsilon_t \qquad \text{(B.2)}$$

Nota-se que o teste é unicaudal, pois a hipótese alternativa de que o processo de ser estacionário é $|\alpha| < 1$ ou $\beta < 0$, visto que o coeficiente na variável defasada da equação (B.2) é menor que zero. Seus valores críticos devem ser aumentados, dependendo do tamanho da amostra. Os refinamentos subsequentes do teste são referidos como teste Dickey-Fuller Aumentado - *ADF (m)*. Tais podem adicionar um intercepto ou um intercepto mais uma tendência determinística, resultando em outros processos adequados para testar a raiz unitária. Além disso, para corrigir uma possível presença de autocorrelação nos resíduos das equações (B.3), (B.4) e (B.5), é adicionado o somatório $\sum_{i=1}^{i=p} \beta_i \nabla y_{t-i}$ O número de defasagens incluídas deve ser tal que remova qualquer autocorrelação dos resíduos, a fim de que a regressão MQS forneça um estimador não-viesado do coeficiente de Z_{t-1}.

Valores críticos ligeiramente diferentes são aplicados ao teste *ADF (m)*, embora o princípio geral de teste da significância do coeficiente em Z_{t-1} seja similar ao DF. (ALEXADER, 2004)

Portanto, sejam as ADF (*m*):

- ADF-I: $\nabla y_t = \beta y_{t-1} + \sum_{i=1}^{m} \beta_i \nabla y_{t-i} + \varepsilon_t$; (B.3)

  H_0: $\beta = 0 \rightarrow y_t \sim$ não estacionária (RW)

  H_1: $\beta < 0 \rightarrow y_t \sim I(0)$ (estacionária)

- ADF-II: $\nabla y_t = \alpha_0 + \beta y_{t-1} + \sum_{i=1}^{m} \beta_i \nabla y_{t-i} + \varepsilon_t$; (B.4)

  H_0: $\beta = 0 \rightarrow y_t \sim$ RW mais um *drift*

  H_1: $\beta < 0 \rightarrow y_t \sim I(0)$ (estacionária)

- ADF-III: $\nabla y_t = \alpha_0 + \alpha_1 t + \beta y_{t-1} + \sum_{i=1}^{m} \beta_i \nabla y_{t-i} + \varepsilon_t$; e (B.5)

  H_0: $\beta = 0 \rightarrow y_t \sim$ RW mais um *drift* em torno de uma tendência determinística

  H_1: $\beta < 0 \rightarrow y_t \sim I(0)$ (estacionária)

A estatística de teste *ADF* (*m*) é dada por: $ADF\ (m) = b - 0 / s.e\ (b)$, em que *b* é a estimativa de β da regressão ADF (*m*) e $s.e\ (b)$, o erro padrão de *b*. A escolha de qual estrutura ADF (*m*) segue as condições:

- Se a série parece ser estacionária com média nula, então a estrutura adequada é ADF-I;
- Se a série parece ser estacionária com média não nula, então a estrutura adequada é ADF-II;
- Se a série parece ser não estacionária, mas sem tendência sustentada, positiva ou negativa, então a estrutura adequada é ADF-II; e
- Se a série parece ser não estacionária, com tendência sustentada, positiva ou negativa, então a estrutura adequada é ADF-II ou ADF-III.

Através da figuras 6, definida por Camargo e Souza (1996), pode-se visualizar melhor o processo de identificação, em que há duas situações quanto ao parâmetro ϕ_1, para em os modelos AR (*1*).

APÊNDICE C

C.1
Simulação de Quase-Monte Carlo

A simulação utilizando o método de QMC é similar ao tradicional Monte-Carlo (MC), exceto na geração da seqüência de números aleatórios, onde este utiliza sequências pseudo-aleatórias e aquele, quase-aleatórias.

A vantagem de usar a seqüência quase-aleatória é que a mesma tem a finalidade de gerar números distribuídos de forma a não formarem *clusters.* Com isso, a convergência, em relação ao MC, ocorre mais rapidamente, na ordem de $(ln*N)^s / N$ (sendo *s*, o número de dimensões), contra $1/\sqrt{N}$ do método de MC. (KRYKOVA, 2003)

A seqüência de baixa discrepância é um conjunto *s*-dimensional de pontos que preenchem uma região de modo eficiente e sem grandes distâncias entre um ponto e outro. Em vista disso, a sequência a quase aleatória é dita de baixa discrepância. Destaca-se, porém, que alguns métodos QMC possuem limitações determinadas dimensionalidades. (KRYKOVA, 2003)

De acordo com KRYKOVA (2003), as vantagens principais do Método de QMC sobre o MC são:

- O cálculo da integral é obtido por aproximação através de uma seqüência de pontos;

- O tempo de computação na simulação de QMC é quase igual ou menor (devido ao menor erro gerado) ao gasto no método MC;

- Em várias aplicações, geram-se resultados melhores, principalmente nas caudas das distribuições; e

- Possui melhor desempenho quando se trabalha com número elevado de dimensões.

C.1.1
Seqüência de Baixa Discrepância

Como visto, o conceito de baixa discrepância está associado à propriedade de geração, evitando a formação de *clusters*. Pode-se dizer também que o intuito é obter uma distribuição dos pontos cubra graficamente toda a área do plano (considerando duas dimensões) de maneira eficiente e sem discrepâncias.

Segundo DIAS (2008), a seqüência de Van der Corput é a mais conhecida e utilizada na literatura para geração de números com baixa discrepância, em caso unidimensionais.

De acordo com KRYKOVA (2003), o algoritmo de geração de seus valores segue três passos:

1. O número *n* na base decimal é expandido à base *b*. Sendo n = 4 na base binária, por exemplo, obtém-se o valor 100 ($4 = 1 \times 2^2 + 0 \times 2^1 + 0 \times 2^0$);
2. O valor na base *b* é refletido. No exemplo, 001; e
3. Assim, o número 0,001 corresponde ao número decimal 1/8.

Pode-se generalizar o método utilizando duas equações:

- A primeira, para transformação de qualquer número *n*, em qualquer base *b*. Por exemplo, na base 2, utilizam-se seqüências de zeros e uns; 3, seqüências de zeros, uns e dois).
  Portanto:

$$n = \sum_{j=0}^{m} a_j(n) b^j \qquad \text{(C.1)}$$

- A segunda, para gerar o número correspondente à seqüência de Van der Corput, na base *b*:

$$b(n) = \Phi_b(n) = \sum_{j=0}^{m} a_j(n) b^{-j-1} \qquad \text{(C.2)}$$

Onde *m* é o menor valor que torna $a_j(n) = 0$, para todo $j > m$. Existem outros tipos de seqüências de baixa discrepância, mas para o caso

multidimensional: HALTON (1960), FAURE (1982), NIEDERREITER (1987), SOBOL (1967), Quase-Monte-Carlo Híbrido. (KRYKOVA, 2003)

C.1.2
Sequência de Quase-Monte-Carlo Híbrida

Segundo (KRYKOVA, 2003), a técnica QMC Híbrida é basicamente um processo de *n* simulações e *d* dimensões obtido através dos seguintes passos:

1. Gera-se um vetor-coluna de *n* números quase-aleatórios, pode-se usar a sequência de van der Corput;
2. Para cada dimensão *d*, faz-se uma permutação independente e aleatória dos elementos do vetor-coluna da anterior, inserindo-o na posterior.

$$a_j^d(n) = \frac{\sum_{j \geq i}^{m} \frac{j!}{i!(j-i)!} a_j^{d-1}(n)}{b} \tag{C.3}$$

Por seguinte, segue-se a definição:

$$x_n^{\ d} \ = \ \Phi_b^{\ d}(n) \ = \ \sum_{i=0}^{m} \ a_j^{\ k}(n) \ p^{i+1} \tag{C.4}$$

Tabela C.1: Valores das Três Primeiras Dimensões na Base Dois (Fonte: KRYKOVA, 2003)

N	**Dimensão 1 (base 2)**	**Dimensão 2 (base 2)**	**Dimensão 3 (base 2)**
1	1/2	1/2	1/4
2	1/4	1/8	3/8
3	3/4	3/4	1/2
5	1/8	5/8	1/16
6	5/8	3/8	5/8
7	3/8	1/4	7/8
8	7/8	1/16	3/4
9	1/16	7/8	1/8

O gráfico C.1 mostra os pares ordenados para as dimensões quarenta e nove e cinqüenta.

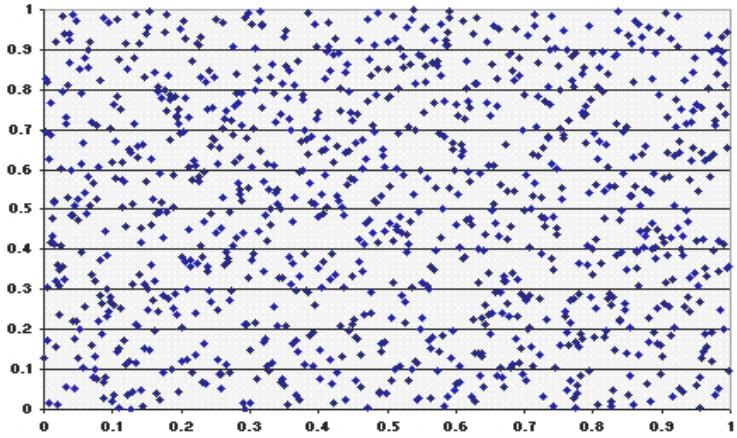

Gráfico C.1 – Sequência Quase Aleatória Híbrida na Base Binária – Dimensões 49 x 50. (Fonte: http://www.puc-rio.br/marco.ind/quasi_mc.html#hybrid)

Foram simuladas mil amostras quase aleatórias independentes de van der Corput na base binária para cada dimensão. Nota-se uniformidade na distribuição dos pontos no plano, mesmo analisando as dimensões 49 e 50.

Segundo KRYKOVA (2003):

> *Visto que os pontos quase aleatórios híbridos foram construídos como permutações de van der Corput na base binária, este método não apresenta problemas de correlação em ambientes de altas dimensões.*

A permutação elimina as correlações, pois preserva as propriedades de baixa discrepância da primeira coluna que ocorre em virtude do baixo valor usado para a base. Assim, os elementos das dimensões subsequentes são os mesmos, porém permutados. (KRYKOVA, 2003)

Determinados algoritmos possuem o problema de altas correlações entre os vetores de altas dimensões, contribuindo à formação de *clusters*. Isso implica a necessidade de um maior número de simulações a fim de que a distribuição uniforme convirja, o que é desfavorável ao processo de simulação. (KRYKOVA, 2003)

Portanto, o método QMC híbrido elimina a correlação cíclica decorrente, em aplicações que envolvem dimensionalidade elevada.

C.1.3
Geração da Distribuição Normal Padrão

No processo de simulação, os números *pseudo* ou quase aleatórios são gerados através de uma distribuição Uniforme. Tal se deve à maior facilidade na geração de outras distribuições de probabilidade através de algoritmos inversores. O principal modo, portanto, de realizar esta transformação é utilizando a inversão da distribuição *U [0,1]* para uma determinada função de probabilidade desejada.

Em simulação, tem-se número *Y(x*) oriundo de uma uniforme, isto é, *Y(x) ~ U [0,1]*. O valor de interesse, porém, é a amostra *x***,** correspondente a um valor de quantil de determinada distribuição, no caso, a normal-padrão. Portanto, *Y(x*) pode ser interpretado como uma probabilidade acumulada.

$$Y(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \qquad \text{(C.5)}$$

Segundo KRYKOVA (2003), uma das maneiras para transformar números quase-aleatórios oriundos da distribuição *U [0,1]* em uma amostra normalmente distribuída é através do algoritmo de Box & Muller. No entanto, para seqüências de baixa discrepância, não é o mais apropriado, pois altera justamente a propriedade principal de números de baixa discrepância, a uniformidade da distância entre um ponto e outro. Além disso, o método não tem um desempenho satisfatório, quanto aos valores nas caudas da distribuição normal.

Visto que inversor de Chebychev possui um bom desempenho nas caudas das normais e o de BEASLEY & SPRINGER (1977), na região central. O algoritmo de MORO (1995) (ou *Inversão de Moro*) apresenta uma solução híbrida, dividindo o domínio da distribuição Uniforme do seguinte modo:

1. A região central da distribuição, $|U| \leq 0,42$ - $U = Y(x)$ - 0,5 - é modelada pelo o algoritmo de BEASLEY & SPRINGER (1977):

$\Phi^{-1}(x) = U \frac{\sum_{n=0}^{3} a_n U^{2n}}{\sum_{n=0}^{4} b_n U^{2n}}$, onde a_n e b_n são constantes que assumem os valores da tabela C.2.

Tabela C.2: Valores de a_n e b_n (Fonte**:** KRYKOVA, 2003)

N	$\boldsymbol{a_n}$	$\boldsymbol{b_n}$
0	2,50662823884	1,00
1	-18,61500062529	-8,4735109309
2	41,39119773534	23,08336743743
3	-25,44106049637	-21,06224101826
4		3,13082909833

2. As caudas da distribuição ($|U| > 0,42$), porém, são modeladas por séries truncadas de Chebyshev:

$$\Phi^{-1}(x) = \sum_{n=0}^{8} c_n T_n(z) - \frac{c_0}{2}, \ \ se\ U > 0$$

$$z = k_1\ [2\ln(-\ln(0,5 - |U|)) - k_2]$$

$$\frac{c_0}{2} - \sum_{n=0}^{8} c_n T_n(z), \ \ se\ U \le 0$$

As constantes k_1 e k_2 são escolhidas de forma que $z = -1$ quando, $\Phi(x) = 0,92$,e $z = 1$, quando $\Phi(x) = 1 - 10^{-12}$. Para cada valor de *n*, a tabela C.3 mostra os valores que *c* e *k* assumem.

Tabela C.3: Valores de c_n e k_n. (Fonte: KRYKOVA, 2003)

n	c_n	k_n
0	7,7108870705487895	
1	2,7772013533685169	0,4179886424926431
2	0,3614964129261002	4,2454686881376569
3	0,0373418233434554	
4	0,0028297143036967	
5	0,0001625716917922	
6	0,0000080173304740	
7	0,0000003840919865	
8	9,9999999129707170	

Portanto, o uso deste procedimento para construção das caldas da normal é factível, dado o uso de sequências QMC.

APÊNDICE D

D.1
Modelo de Combinação Linear de Previsões

Considerando o modelo de GRANGER & BATES (1969), tem-se a equação:

$$y_{T+h,CL} \;=\; \omega\; y_{T+h,1} + (1-\omega)\; y_{T+h,2},\;\; (h = 1,2,3,...) \qquad \text{(D.1)}$$

Onde ω é o peso da previsão do modelo 1 ($y_{T+h,1}$) e $(1-\omega)$, o peso do modelo 2 ($y_{T+h,2}$). Assim, a variância da previsão combinada é dada por:

$$\sigma^2{}_{CL} \;=\; \omega^2\; \sigma^2{}_1 + (1-\omega)^2\; \sigma^2{}_2 + 2\rho\; \omega\; \sigma_1\; (1-\omega)\; \sigma_2 \qquad \text{(D.2)}$$

Onde $\sigma^2{}_1$ e $\sigma^2{}_2$ são as variâncias dos erros das previsões a serem combinadas; ρ , o coeficiente de correlação entre os erros de previsão e ω, o peso associado linearmente à previsão do modelo 1.

Para a minimização da variância $\sigma^2{}_{CL}$, procede-se a diferenciação da equação (D.2), com relação a ω, igualando-a a zero (condição de primeira ordem). Desse modo, o mínimo de $\sigma^2{}_{CL}$ ocorre quando ω assume o valor dado pela equação (D.3).

Por minimizar a variância, este método ficou conhecido como *Método da Variância Mínima*.

$$\omega \;=\; \frac{\sigma^2{}_2 - \rho\; \sigma_1\; \sigma_2}{\sigma^2{}_1 + \sigma^2{}_2 - 2\rho\; \sigma_1\; \sigma_2} \qquad \text{(D.3)}$$

Caso os erros sejam descorrelacionados ($\rho \;=\; 0$), tem-se que ω fica reduzido pela equação (D.4).

$$\omega \;=\; \frac{\sigma^2{}_2}{\sigma^2{}_1 + \sigma^2{}_2} \qquad \text{(D.4)}$$

Como as variâncias dos erros não são conhecidas, GRANGER & BATES propuseram cinco procedimentos de estimação, tendo como base os erros de previsão.

Inclusive NEWBOLD & BATES (1974) utilizaram estes procedimentos, mas para mais de duas previsões.

i. Procedimento 1:

$$\hat{\omega}_{i,T} = \left(\sum_{t=T-\nu}^{T-1} e^2{}_{i,t} \right)^{-1} \Bigg/ \sum_{j=1}^{P} \left(\left(\sum_{t=T-\nu}^{T-1} e^2{}_{j,t} \right)^{-1} \right)$$

Onde *i* (*i = 1,...,p*) e *j* (*j = 1,...,p*) são índices relativos às previsões-base e *T*, o período de tempo atual.

ii. Procedimento 2:

$$\hat{\omega} = \left(\hat{\Sigma}^{-1} 1 \right) \Big/ \left(1' \hat{\Sigma}^{-1} 1 \right), \quad \text{onde:} \quad 0 \le \omega_{i,t} \le 1, \quad \text{para todo } i$$

$$, \hat{\Sigma}_{i,j} = \nu^{-1} \sum_{t=T-\nu}^{T-1} e_{i,t}\, e_{j,t} \text{ e } \mathbf{1} = (1,1,...,1)'$$

iii. Procedimento 3:

$$\hat{\omega}_{i,T} = \alpha\, \hat{\omega}_{i,T-1} \left((1-\alpha) \left(\sum_{t=T-\nu}^{T-1} e^2{}_{i,t} \right)^{-1} \Bigg/ \sum_{j=1}^{P} \left(\left(\sum_{t=T-\nu}^{T-1} e^2{}_{j,t} \right)^{-1} \right) \right), \quad \text{onde:}$$

$0 \le \alpha \le 1$.

iv. Procedimento 4:

$$\hat{\omega}_{i,T} = \hat{\omega}_{i,T-1}\left(\left(\sum_{t=T-\nu}^{T-1} z_t\, e^2{}_{i,t}\right)^{-1} \Big/ \sum_{j=1}^{P}\left(\left(\sum_{t=T-\nu}^{T-1} z_t\, e^2{}_{j,t}\right)^{-1}\right)\right)$$

Onde $z_t \geq 1$, sendo z_t um peso que atribui maior relevância à variância dos erros recentes.

v. Procedimento 5:

$$\hat{\omega} = \left(\hat{\Sigma}^{-1}\, 1\right) \Big/ \left(1'\hat{\Sigma}^{-1} 1\right)$$

$$\hat{\Sigma}_{\hat{i},j} = \left(\left(\sum_{t=1}^{T-1} z_t\, e_{i,t}\, e_{j,t}\right)^{-1} \Big/ \sum_{j=1}^{P}\left(\left(\sum_{t=T-\nu}^{T-1} z_t\right)^{-1}\right)\right)$$

. Onde $0 \leq \hat{\omega}_{i,T} \leq 1$, para todo *i*, e, com $z_t \geq 1$.

Dado o exposto, é importante salientar dois pontos:

- Pequenos valores de ν limitam o processo de estimação dos pesos adaptativos lineares, utilizando apenas valores mais recentes;

- Pequenos valores de α e grandes de z_t significa que é dado maior peso aos valores mais recentes

Livros Grátis

(http://www.livrosgratis.com.br)

Milhares de Livros para Download:

Baixar livros de Literatura
Baixar livros de Literatura de Cordel
Baixar livros de Literatura Infantil
Baixar livros de Matemática
Baixar livros de Medicina
Baixar livros de Medicina Veterinária
Baixar livros de Meio Ambiente
Baixar livros de Meteorologia
Baixar Monografias e TCC
Baixar livros Multidisciplinar
Baixar livros de Música
Baixar livros de Psicologia
Baixar livros de Química
Baixar livros de Saúde Coletiva
Baixar livros de Serviço Social
Baixar livros de Sociologia
Baixar livros de Teologia
Baixar livros de Trabalho
Baixar livros de Turismo