

GUILHERME MATIUSSI RAMALHO

**UMA ABORDAGEM ESTATÍSTICA PARA O MODELO DO PREÇO
SPOT DA ENERGIA ELÉTRICA NO SUBMERCADO
SUDESTE/CENTRO-OESTE BRASILEIRO**

**São Paulo
2014**

GUILHERME MATIUSSI RAMALHO

**UMA ABORDAGEM ESTATÍSTICA PARA O MODELO DO PREÇO
SPOT DA ENERGIA ELÉTRICA NO SUBMERCADO
SUDESTE/CENTRO-OESTE BRASILEIRO**

Dissertação apresentada ao Departamento de Engenharia de Telecomunicações e Controle da Universidade de São Paulo para obtenção do título de Mestre em Ciências.

Área de Concentração: Engenharia de Sistemas
Orientador: Prof. Dr. Oswaldo Luiz do Valle Costa

**São Paulo
2014**

Dedicatória

À minha família Ivete, Paulo, Ian Arthur e Izabel e à minha companheira Renata pelo apoio incondicional e pela compreensão nos momentos de ausência necessários à conclusão deste trabalho.

Agradecimentos

Ao professor Dr. Oswaldo L.V. Costa, pela grande compreensão, orientação e disponibilidade necessária para conclusão desta Dissertação.

Ao prof. Dr. Rodrigo de Barros Nabholz e à profa. Dra. Virginia Parente pelo direcionamento nas fases iniciais deste trabalho.

À Escola Politécnica da USP por me permitir concluir mais um curso em suas dependências e todos os professores e profissionais que tornam essa instituição uma referência internacionalmente conhecida. Espero conseguir contribuir para a sociedade brasileira à altura do título que carrego.

Aos amigos Vitor Chiamente, Marcelo Parodi e Paulo Mayon pela pronta disponibilidade em conversas e esclarecimentos sobre o mercado de energia elétrica.

*“If the only tool you have is a hammer, you tend to see every problem as a
nail.”*

Abraham Harold Maslow

Resumo

O objetivo deste trabalho é o desenvolvimento de uma ferramenta estatística que sirva de base para o estudo do preço *spot* da energia elétrica do subsistema Sudeste/Centro-Oeste do Sistema Interligado Nacional, utilizando a estimação por regressão linear e teste de razão de verossimilhança como instrumentos para desenvolvimento e avaliação dos modelos. Na análise dos resultados estatísticos descritivos dos modelos, diferentemente do que é observado na literatura, a primeira conclusão é a verificação de que as variáveis sazonais, quando analisadas isoladamente, apresentam resultados pouco aderentes ao preço *spot* PLD. Após a análise da componente sazonal é verificada a influência da energia fornecida e a energia demandada como variáveis de entrada, com o qual conclui-se que especificamente a energia armazenada e produção de energia termelétrica são as variáveis que mais influenciam os preços *spot* no subsistema estudado. Entre os modelos testados, o que particularmente ofereceu os melhores resultados foi um modelo misto criado a partir da escolha das melhores variáveis de entrada dos modelos testados preliminarmente, alcançando um coeficiente de determinação R^2 de 0.825, resultado esse que pode ser considerado aderente ao preço *spot*. No último capítulo é apresentada uma introdução ao modelo de predição do preço *spot*, possibilitando dessa forma a análise do comportamento do preço a partir da alteração das variáveis de entrada.

Índice de Termos - Modelo de mínimos quadrados, Estimação de máxima verossimilhança, Modelos de predição, Análise de regressão e Processos estocásticos.

Abstract

The objective of this work is the development of a statistical method to study the spot prices of the electrical energy of the Southeast/Middle-West (SE-CO) subsystem of the The Brazilian National Connected System, using the Least Squares Estimation and Likelihood Ratio Test as tools to perform and evaluate the models. Verifying the descriptive statistical results of the models, differently from what is observed in the literature, the first observation is that the seasonal component, when analyzed alone, presented results loosely adherent to the spot price PLD. It is then evaluated the influence of the energy supply and the energy demand as input variables, verifying that specifically the stored water and the thermoelectric power production are the variables that the most influence the spot prices in the studied subsystem. Among the models, the one that offered the best result was a mixed model created from the selection of the best input variables of the preliminarily tested models, achieving a coefficient of determination R^2 of 0.825, a result that can be considered adherent to the spot price. At the last part of the work It is presented an introduction to the spot price prediction model, allowing the analysis of the price behavior by the changing of the input variables.

Index Terms - Least squares approximations, Maximum likelihood estimation, Mean square error methods, Prediction methods, Regression analysis and Stochastic processes.

Lista de Tabelas

Diagrama I : Resultados do modelo Nord Pool	20
Diagrama II : Resultados do modelo australiano	26
Tabela I : Estatística PLD	43
Diagrama III : Correlações Entre Preditores e PLD	46
Tabela II : Variação Sazonal Mensal	48
Tabela III : Efeito Sazonal	49
Tabela IV : Parâmetros de Suprimento	51
Tabela V : Efeitos do Suprimento	53
Tabela VI : Efeitos do Suprimento 2	54
Tabela VII : Parâmetros de Demanda	56
Tabela VIII : Efeito da Demanda	57
Tabela IX : Efeito da Demanda 2	58
Tabela X : Modelo Final	60

Lista de Figuras

3.1	Resíduos x PLD de resposta	38
3.2	Gráfico Q-Q	39
4.1	Boxplot do PLD histórico	44
4.2	PLD histórico	45
4.3	Densidade PLD	45
4.4	Gráfico Q-Q - Modelo Sazonal	50
4.5	Gráfico Q-Q - Modelo de Suprimentos Inicial	54
4.6	Gráfico Q-Q - Modelo de Suprimentos Final	54
4.7	Suprimentos : Resíduos x PLD	55
4.8	Relação Energia Armazenada x ENA	56
4.9	Modelo Demanda	58
4.10	Hidráulica e Carga de Demanda	60
5.1	Predição do Valor Médio de PLD	62
6.1	Utilização de Energia Termelétricas	67

Lista de Abreviaturas

ACL - Ambiente de Contratação Livre
ACR - Ambiente de Contratação Regulada
ANEEL - Agência Nacional de Energia Elétrica
CBOT - Chicago Board of Trade
CCEE - Câmara de Comercialização de Energia Elétrica
CD - Carga de Demanda (Abreviatura em equações)
CE - Carga de Energia (Abreviatura em equações)
CMO - Custo Marginal de Operação
E - ENA (Abreviatura em equações)
ENA - Energia Natural Afluente
ESS - Encargo de Serviços do Sistema
FERC - Comissão Reguladora Federal de Energia
H - Hidrelétrico (Abreviatura em equações)
IE - Intercâmbio de Energia (Abreviatura em equações)
MBG - Movimento Browniano Geométrico
MCP - Mercado de Curto Prazo
MGE - Minneapolis Grain Exchange
MWh - Megawatts por hora
N - Nuclear (Abreviatura em equações)
NEC - Nordic Electricity Clearing
NEMMCO - National Electricity Market Management Company
NPE - Nordic Power Exchange
NSW - New South Wales
NYMEX - New York Mercantile Exchange
OIS - Operadores Independentes de Sistemas
ONS - Operador Nacional do Sistema Elétrico
PJM - Pennsylvania - New Jersey - Maryland Interconnection
PLD - Preço de Liquidação das Diferenças
QLD - Queensland
QNI - The Queensland e o New South Wales Interconnector
S - Energia Armazenada (Abreviatura em equações)
SA - South Australia

SE/CO - Sudeste/Centro-Oeste
SIN - Sistema Interligado Nacional
T - Termoelétrico (Abreviatura em equações)
VIC - Victoria

Conteúdo

1	Introdução	1
2	Mercados de Energia Elétrica	5
2.1	Mercado Brasileiro de Energia Elétrica	9
2.1.1	Estudo de caso de precificação	11
2.2	Mercado dos Países Nórdicos	12
2.2.1	O sistema de geração do Nord Pool	14
2.2.2	Estudo de caso	14
2.2.3	Modelo para preço spot	15
2.3	Mercado Australiano	21
2.3.1	Dados e estatísticas	23
2.3.2	Especificação do modelo	23
2.4	Mercado dos Estados Unidos	27
3	Método Empregado	31
3.1	Introdução teórica ao método empregado	31
3.1.1	Estimação dos parâmetro de regressão	32
3.1.2	Abordagem geométrica para o método	33
3.1.3	Qualidade da regressão	33
3.1.4	Teste de hipótese sobre os parâmetros do modelo	34
3.2	Método empregado para avaliação do preço <i>spot</i>	36
3.2.1	Procedimento de estimação por mínimos quadrados	36
3.2.2	Normalidade do termo de erro	37
3.2.3	Verificação dos preditores significativos	40
3.3	Passos do método empregado	40
4	Análise dos Modelos	42
4.1	Origem dos Dados	43
4.2	Caraterísticas Gerais do Preço Spot	43
4.3	Resposta a Parâmetros Sazonais	46
4.3.1	Originação dos Dados - Parâmetros Sazonais	46
4.3.2	Modelo - Parâmetros Sazonais	47
4.3.3	Resultados - Parâmetros Sazonais	47

4.4	Resposta a Parâmetros de Suprimento	49
4.4.1	Originação dos Dados - Parâmetros de Suprimento	51
4.4.2	O Modelo - Parâmetros de Suprimento	52
4.4.3	Resultados - Parâmetros de Suprimentos	52
4.5	A Resposta a Parâmetros de Demanda	55
4.5.1	Originação de Dados - Parâmetros de Demanda	56
4.5.2	O Modelo - Parâmetros de Demanda	57
4.5.3	Resultados - Parâmetros de Demanda	57
4.6	Modelo Final	59
5	Introdução à Predição do PLD	61
6	Conclusão	64

Capítulo 1

Introdução

A energia elétrica, em diferentes abordagens e pontos de vista, é tema de muitos estudos atualmente. Aspectos como geração, otimização e racionalização de redes, expansão, análise de regulação são exemplos de linhas de estudo que demonstram o grau de importância que a energia elétrica tem no cotidiano das pessoas e o grau de influência que produz nas economias. O ativo base para qualquer modelo de precificação de energia elétrica é o seu preço *spot*, que, de forma generalizada, reflete o preço que atualmente um determinado mercado regionalizado está utilizando nas negociações da *commodity* energia elétrica para utilização no curto prazo ou imediato.

No sentido de área de abrangência de um preço *spot*, pode-se entender como região um estado de um determinado país, um conjunto de estados (como a região Sudeste/Centro-Oeste do território brasileiro), um país ou mesmo um conjunto de países, e reflete a área de atuação desse preço *spot* nas negociações dos mercados no qual a energia elétrica é operada e normalmente está relacionada à região de geração dessa energia. Nesse sentido são muitas as variáveis que afetam os preços *spot* da energia elétrica, entre elas podemos citar, por exemplo, as variações sazonais, os custos marginais de produção, restrições de transporte e interconexões, limites de armazenamento, regulamentação e padrões de consumo de energia. Essas variáveis possuem diferentes níveis de influência em cada mercado regional que atua, e dessa forma se observa que um modelo que procura explicar o preço *spot* para uma determinada região dificilmente irá ser aderente em outras regiões sem alterações, isso devido às especificidades de cada realidade.

A importância do estudo do preço da energia elétrica está no desenvolvimento do mercado de comercialização de energia, com aplicações em predição de preços e avaliação de derivativos, por exemplo. Dessa forma, nos estudos de precificação de energia, o maior desafio é encontrar ou produzir modelos que sejam aderentes suficientemente às realidades regionais das suas áreas de comercialização, com respeito à volatilidade, tendências, perfil de consumo, sazonalidades e outras características próprias desses preços, que os tornem confiáveis para as aplicações financeiras que objetivaram o seu desenvolvimento.

O mercado livre de energia elétrica, como será discutido no capítulo 2, é o mercado no qual as negociações são tratadas livremente entre as partes através de contratos bilaterais. Apesar desse mercado estar em constante desenvolvimento no Brasil com respeito ao incremento do número

de participantes, liquidez e polarização de bolsas de energia, a maior parte das negociações do mercado livre atualmente ainda ocorrerem em mercado de balcão, fora do ambiente de bolsa. Essa situação dificulta a análise quantitativa dos preços de energia elétrica nos mercados regionais brasileiros pois o acesso à informação das negociações é bastante limitado, sendo que as informações, na maioria dos casos, ficam restritas às partes e contrapartes da negociação. Como será explicado nos capítulos a seguir, no mercado brasileiro usa-se o Preço de Liquidação das Diferenças (PLD) como *proxy* do preço *spot* de energia, sendo através de ágios e deságios sobre esse preço de referência que o mercado faz a comercialização de energia no mercado livre.

Como será analisado nesse trabalho, o preço *spot* da energia elétrica do sistema brasileiro tende a possuir uma menor volatilidade quando comparada a outros mercados de bolsa mais desenvolvidos e líquidos de energia elétrica. Uma das justificativas para essa menor volatilidade, conforme verifica Jong 2005 [12], é relacionada principalmente pela forte presença de energia hidrelétrica e a sua capacidade de armazenamento de água, e conseqüentemente energia, nos reservatórios das usinas. Essa capacidade de armazenamento de água se reflete em uma menor volatilidade e uma maior estabilidade dos preços, porém influi também sobre o gerenciamento do Sistema Interligado Nacional por parte do Operador Nacional do Sistema, pois o controle de despacho de um sistema hidrotérmico é mais complexo do que um sistema de base termelétrica.

Atualmente, observa-se que estudos estão sendo feitos endereçando o preço *spot* como um processo estocástico na forma $P_t = f(t) + X_t$, onde P_t é o preço *spot*, $f(t)$ é uma função determinística e X_t é um processo estocástico. Frequentemente, $f(t)$ é uma função que captura um padrão sazonal ou uma mudança física no sistema que afeta o comportamento dos preços, influenciando profundamente o modelo por definir o preço base para um determinado momento. O processo estocástico X_t usualmente é modelado como um processo de reversão à média, ou um processo difusivo a saltos, pois dessa forma é possível caracterizar a alta volatilidade que normalmente esses preços possuem. Essa abordagem de modelo é discutida por Higgs, H., Worthington, A.C. 2006 [10], Deng, S.2003 [7] e Julia, L., Schwartz, E. 2001 [15] em diferentes situações e países, e será melhor discutida no capítulo 2.

Como será verificado pelo presente estudo, unicamente uma abordagem sazonal determinística não define bem o preço *spot* no mercado de energia elétrica brasileiro. O principal objetivo do presente estudo é definir, dessa forma, um modelo do preço *spot*, ou PLD, que se adapte bem a um submercado específico do mercado brasileiro. A abordagem é de identificar, dada uma base de dados histórica, as principais componentes que influenciam o comportamento do PLD. O estudo seguirá uma abordagem estatística, utilizando a regressão linear e o teste da razão de verossimilhança com o objetivo avaliar a aderência dos modelos e melhorar as respostas a partir de critérios definidos.

A base de dados mensal do sistema foi obtida da base pública oferecida pela ONS no seu *website* [17], e a base semanal de PLD, dividido nos quatro subsistemas e em três patamares diários de energia, foram obtidas do site da CCEE [4]. Esse trabalho centralizou o estudo no submercado Sudeste/Centro-Oeste uma vez que esse submercado que concentra atualmente a maior parte da geração e do consumo de energia no Brasil, compreendendo os mercados de

energia elétrica de São Paulo, Rio de Janeiro e Minas Gerais.

A introdução de modelos teóricos para o preço *spot* de energia é um desafio em todos os mercados onde a energia elétrica é comercializada em mercado livre, pois possui características de volatilidade e comportamento que a difere de todas as outras commodities. O capítulo 2 inicia-se com uma breve introdução dos motivos que tornam essa *commodity* diferente dos outros ativos operados no mercado financeiro. Após isso é verificado qual o atual estado da arte para o mercado brasileiro, onde são oferecidos alguns exemplos de estudos feitos no sentido de análise do PLD para utilização na precificação de derivativos sobre o preço *spot*. Na sequência são verificados alguns mercados que possuem características que poderiam se assemelhar às encontradas no mercado brasileiro, como o mercado do Nord Pool e o mercado Australiano. Na última seção desse capítulo é feita uma breve análise das características de alguns mercados organizados norte-americanos, uma vez observada a sua característica de mercados regionalizados e independentes.

O capítulo 3 tem o objetivo de introduzir os métodos matemáticos utilizados no capítulo subsequente para avaliação dos modelos de precificação do PLD. Inicialmente, seguindo uma abordagem teórica, são disponibilizados os instrumentos teóricos de análise de Regressão Linear, os critérios utilizados para avaliação dos modelos e seleção dos preditores, como o coeficiente de determinação e a razão de máxima verossimilhança. A segunda seção desse capítulo tem o objetivo de verificar como a teoria introduzida na primeira seção é interpretada na prática para a aplicação nas avaliações e modelos testados. Nessa seção são definidos também quais são os métodos utilizados para verificação da hipótese de normalidade do resíduo $\hat{\epsilon}$, condição necessária para aplicação dos critérios de seleção dos preditores utilizando testes de hipótese para validar a influência dos parâmetros de regressão β s nos modelos.

A seguir é apresentado o capítulo 4 com as avaliações dos modelos propriamente ditos. Inicialmente são analisados alguns parâmetros estatísticos básicos que são utilizados para verificar o comportamento do PLD, como média e mediana, volatilidade e parâmetros que indicam tendências de como é a distribuição da densidade de probabilidade pela medição da curtose e simetria. Após isso, parte-se para uma análise de um dos modelos mais utilizados para estudos do preço *spot* de energia elétrica, que utiliza a premissa de sazonalidade dos preços durante o ano. Na seção seguinte inicia-se uma busca de quais são os verdadeiros critérios, ou preditores, que influenciam os preços através da verificação dos suprimentos de energia elétrica. Nessa classe entram os preditores que englobam os métodos de suprimento de energia elétrica no sistema, como a ENA, a energia armazenada, a quantidade de energia hidráulica produzida, a quantidade de energia térmica convencional e a energia térmica nuclear.

Ainda no capítulo 4, após a verificação dos parâmetros de suprimento, são analisados os parâmetros de demanda, como a energia consumida, a energia transacionada entre submercados e uma forma de avaliação da influência dos picos de demanda de energia nos preços. Esse capítulo é concluído através da apresentação de um modelo misto que busca avaliar os resultados conseguidos por um modelo que coloca os melhores preditores de cada modelo testado em um único modelo, com o objetivo de encontrar uma reposta mais aderente ao PLD real.

No capítulo 5 é feita uma breve introdução teórica ao conceito de predição, com um exemplo aplicado ao Modelo Final discutido no capítulo 4. A seguir é apresentada a conclusão que se inicia com os principais resultados do trabalho, onde é constatado que, para o submercado em questão, o modelo de variação sazonal dos preços de energia elétrica não é aderente à realidade dos preços quando analisado individualmente. Constata-se também que os principais preditores que influenciam os preços são os relacionados aos suprimentos de energia elétrica, com destaque para a produção termoeleétrica e energia armazenada nos reservatórios das hidrelétricas. Ao final são sugeridas linhas de pesquisas futuras para aplicação em precificação de derivativos, predição de preços e verificação da sensibilidade de alteração das variáveis de entrada do sistema nos preços futuros.

Capítulo 2

Mercados de Energia Elétrica

Historicamente, na negociação de commodities e ativos financeiros, a volatilidade é quase sempre definida como o desvio padrão do retorno promovido pela variação dos preços desses ativos, sendo estimado através do levantamento de séries de preços históricos ordenados no tempo. Alguns autores, como Clewlow e Strickland 2000 [6], consideram que essa forma de cálculo da volatilidade, quando aplicada ao mercado de energia elétrica, não é exatamente apropriada devido às características que esses preços apresentam. O mercado de energia elétrica exige uma complexidade maior, no sentido de que um modelo que almeja descrevê-lo deve conseguir capturar as características fundamentais dos seus preços. O grande desafio é que essas características podem variar, por exemplo, de acordo com a região de estudo, com o horário de estudo ou com o modo que a negociação de energia elétrica é contabilizada, o que torna o estudo de precificação de energia elétrica bastante específico.

Na literatura, de forma geral, observa-se que na busca de uma estimativa da volatilidade do preço não enviesada, existe uma preocupação em se estimar a volatilidade da energia elétrica a partir da definição da abrangência da análise para uma determinada época (sazonalidade) e região (diferenciação geográfica). O que se nota é que ao se misturar diferentes épocas do ano, ou a tentativa de utilização da mesma volatilidade para mais de um sistema, muitas vezes é inapropriado devido ao comportamento diferenciado que o preço da energia possui para diferentes situações. Via de regra, observa-se que a aplicabilidade do estabelecimento de sazonalidades não é unânime para todos os mercados, quando se busca quantificar da melhor forma possível a volatilidade.

Os riscos encontrados no mercado de energia elétrica são variados e muito específicos, Mayo 2009 [16] caracteriza esses riscos, enfatizando que muitas vezes esses podem se mostrar interdependentes. Os principais riscos apontados são:

- Risco de preço de mercado: É o risco mais genérico, que tem relação com a alta volatilidade dos preços com motivações variadas, mas muitas vezes relacionado aos meios de produção, à flutuação da demanda, à disponibilidade de recursos (água, combustível, vento etc.) ou restrições inerentes ao sistema de cobertura em termos de distribuição e transmissão.
- Risco de preço de combustível: Esse risco é inerente a qualquer mercado que possui

uma parcela significativa do seu mercado de suprimento proveniente de termelétricas. Variações do preço de combustível afetam o custo de produção dessas usinas, o que para o mercado brasileiro, por exemplo, é fator chave para o Custo Marginal de Operação, que será caracterizado nos capítulos a seguir. Em muitos mercados esses riscos variáveis são absorvidos pelos próprios agentes de produção, porém para outros esse risco é repassado e é refletido no preço final da energia gerada por aquele produtor. Para os mercados que possuem o controle de despacho centralizado, como é o caso do mercado brasileiro, o preço do combustível muitas vezes reflete a ordem de prioridade que uma determinada usina será despachada, além do número de horas que aquela planta deve operar.

- Risco de contraparte: É o risco de crédito inerente aos mercados operados em mercado de balcão ou leilão. Esse tipo de risco reflete a incapacidade de um agente conseguir honrar em termos financeiros, ou produtivos, o que foi estabelecido em contrato. Em mercados de energia mais desenvolvidos, o risco financeiro de contraparte é minimizado em parte através de bolsas de energia elétrica que funcionam como *clearings* (ou associadas a *clearings*), intermediando as negociações de compra e venda de energia elétrica e honrando as posições assumidas por contrapartes inadimplentes.
- Risco volumétrico: O mercado de energia elétrica é basicamente feito de previsões de volume de consumo e volume de oferta, isso porque os contratos são em sua maioria estabelecidos a prazos e volumes que serão consumidos futuramente. Essas previsões podem ser afetadas por diversos motivos, como a flutuação na demanda, impedimentos físicos na rede, falta de combustível ou falta de água no sistema, perda de clientes e manutenções das fontes geradoras fora do cronograma. Esses riscos podem afetar o abastecimento de energia elétrica como um todo, influenciando dessa forma nos preços e volumes das negociações.
- Risco de congestionamento da transmissão: Esse risco está muito ligado ao risco volumétrico. Em determinados mercados, como o brasileiro, os custos relacionados à esse risco são rateados entre todos os agentes dentro de um mesmo submercado, porém as restrições entre submercados são refletidos nos preços finais operados. A transmissão eficiente é tão importante quanto a geração em si, pois determina onde, como e a que preço uma energia de determinada região chegará a outras regiões.
- Risco de liquidez de mercado: Uma vez que em muitos mercados de energia elétrica as negociações de energia são feitas através de mercados de balcão descentralizados, muitas vezes existe uma dificuldade de estabelecimento de preços de mercado uma vez que a informação do preço vigente é desconhecido ou pela falta de agentes negociando ou pela falta de transparência nas negociações.
- Risco hidrológico: Esse é um problema inerente aos mercados de base hidrológica, como o brasileiro, por exemplo. Em mercados com predomínio de geração termelétrica, sempre existe o risco de abastecimento por parte dos fornecedores de combustível, porém esse risco

normalmente é mitigado através do preço que se paga pelo combustível, ou seja, torna-se de fato um problema financeiro ao invés de um problema de abastecimento. No caso de sistemas hidráulicos, o produtor está totalmente exposto ao abastecimento hidrológico que, no fim da linha, foge do controle de qualquer agente do mercado. Dessa forma os produtores do sistema hidráulico sempre estão expostos aos riscos de em determinados momentos estarem sobrecontratados, o que o induz, na maioria dos casos, a honrar os seus contratos comprando energia no mercado *spot* (ou PLD no caso brasileiro), ou estarem subcontratados e nesse caso vender energia no mercado *spot*, ficando então expostos à volatilidade desses preços.

- Risco regulatório: Sendo a energia elétrica uma *commodity* amplamente utilizada e que envolve altas somas monetárias, os agentes reguladores estão constantemente agindo sobre a legislação a fim de estabelecer o que, no seu ponto de vista, seria melhor ao mercado, para o sistema ou para a instituição nacional. Essas mudanças podem oferecer impactos positivos ou negativos, dependendo do ponto de vista do agente do mercado. Esse risco é mais comum em mercados recentes e de bases pouco consolidadas, porém não é incomum ser enfrentado também em mercados já bem estabelecidos. É um dos principais riscos que define se um investidor irá investir ou não em um determinado mercado, pois em mercados com grande risco regulatório, qualquer tentativa de previsão de receita fica fracamente embasada.

Como reflexo desses riscos e das características intrínsecas da energia elétrica, como será verificado a seguir, tem-se que os preços da energia são particularmente voláteis e muitas vezes instáveis.

Como um dos modelos mais conhecidos para descrição do comportamento de ativos financeiros tem-se o Movimento Browniano Geométrico (MBG), propostos por Fisher Black e Myron Scholes em [1] no qual o movimento contínuo médio do ativo segue um crescimento médio geométrico, com desvios (ou “ruídos”) dados por uma distribuição de probabilidade gaussiana, também conhecida como função de distribuição lognormal.

Na observação empírica de muitos que acompanham os preços dessa *commodity*, incluindo nisso operadores de mercado e acadêmicos, é fato que o Movimento Browniano Geométrico não descreve de forma apropriada o seu comportamento. Algumas das características mais comuns que tornam os preços dessa *commodity* tão particular seguem a seguir [6]:

- Ativos de investimento vs. Bens Consumíveis: Uma observação que fica bastante evidente é que não se pode tratar os modelos de preço de energia elétrica como se fossem ativos financeiros puros. Como em muitos casos eles são usados como inputs para o desenvolvimento de outros tipos de negócios, inclusive em mercadorias e outras empresas, a *commodity* energia pode ser vista como um dos fatores que influenciam os ativos financeiros puros. Outro fator observado é que o preço pode variar conforme a fonte proveniente, ou seja, sendo a nossa matriz energética predominantemente de origem hidrotérmica, se olharmos o preço de geração de uma hidrelétrica quando uma termelétrica flexível está

produzindo parecerá que essa está oferecendo energia a um preço relativo negativo, apesar de estar produzindo, o que não seria previsto em um MBG.

- O preço da energia elétrica tem sazonalidade: Quando observadas no tempo, a volatilidade e os preços efetivos de energia elétrica possuem uma certa regularidade temporal. É bastante conhecido em mercados europeus ou americanos a sazonalidade na utilização da energia devido a estações de inverno (para uso na climatização de ambientes) e verão (no uso do ar-condicionado), refletindo esses fatos no preços da energia. No consumo diário, sabemos que existe uma demanda maior de potência em alguns períodos do dia, como em locais com alta concentração industrial onde existem dois picos de consumo, um pela manhã e outro no período da tarde. Para muitos casos, essa sazonalidade é bem quantificada, quando disponíveis, através da observação dos preços futuros dessas commodities. No caso brasileiro, porém, como a informação ainda não é totalmente disponibilizada por ser um mercado ainda em formação, muitas vezes é mais difícil de ser verificada quando comparada aos países que possuem a energia elétrica sendo negociada efetivamente em Bolsa. Uma das principais contribuições desse trabalho será verificar o quanto essa sazonalidade de fato está presente no mercado brasileiro.
- Os preços apresentam um comportamento de processos a salto: Devido às flutuações na demanda que não são acompanhadas na mesma velocidade pela oferta, frequentemente se verificam picos de demanda repentinos e conseqüentemente picos de preços no mercado spot. Esse comportamento é principalmente verificado em mercados de bolsa que operam com energia em tempo real, onde nesses casos é de fato observado o desacoplamento entre oferta e demanda de energia. Frequentemente se verifica também saltos nos preços devido às flutuações de previsão de abastecimento de energia (como no caso de chuvas para hidrelétricas) ou devido a aumentos bruscos nos preços de combustíveis que abastecem as termelétricas, movimentos esses devidos à limitada capacidade de armazenamento de energia, tornando os preços muito sensíveis aos custos de produção.
- Preços giram em torno dos custos de produção: No mercado norte americano e europeu, se nota uma clara tendência dos preços da *commodity* girarem ao entorno dos custos de produção, sobretudo do preço dos combustíveis das termelétricas. No mercado brasileiro esse custo tende a ser menos significativo dada a sua matriz hidrelétrica, porém em semelhança pode-se notar que os preços são caracterizados por uma componente de “Reversão à Média”, que age sobre o preço da energia elétrica limitando o seu crescimento exponencial no longo prazo. Villapana 2002 em [23] e Deng 2003 em [7], tiveram bons resultados ao modelar o preço da eletricidade adicionando saltos estocásticos ao processo de reversão à média.

A importância do estudo do preço *spot* é aplicado no desenvolvimento do mercado de comercialização de energia, com aplicações em predição de preços e avaliação de derivativos, por exemplo. Como visto, as variáveis que afetam os preços da energia elétrica são variadas, por

exemplo a sazonalidade, os custos marginais de produção, restrições de transporte e interconexões, limites de armazenamento, regulamentação e padrões de consumo de energia. Dessa forma, se observa que um modelo que se adapta bem em um país dificilmente irá funcionar em outros países sem adaptações, isso devido às especificidades de cada realidade.

A seguir será feita uma análise do mercado brasileiro, quais são suas principais características e como funciona a precificação de energia. Para efeitos comparativos serão analisados também outros mercados e algumas abordagens utilizadas para *pricing* do preço *spot* de energia elétrica. O estudo de outros mercados mais desenvolvidos de energia elétrica é importante por motivos diversos, podendo oferecer semelhanças ao mercado de energia elétrica brasileiro, seja como experiência de situações ou regulamentações já utilizadas e que possam ser aplicadas, seja com o objetivo de se encontrar padrões modelísticos aplicáveis para o mercado nacional.

2.1 Mercado Brasileiro de Energia Elétrica

No Brasil atualmente existem dois ambientes de contratação de energia elétrica. O primeiro é o Ambiente de Contratação Regulada (ACR) onde se realizam as operações de compra e venda da energia elétrica em ambiente regulado, em que vendedores (em geral geradores) e agentes de distribuição comercializam através de licitações, conforme regras e procedimentos definidos em lei e regulados pela Agência Nacional de Energia Elétrica (ANEEL). O segundo tipo de ambiente de contratação de energia é o Ambiente de Contratação Livre (ACL), no qual se fazem operações de compra e venda de energia elétrica através de contratos bilaterais livremente negociados, sem regulamentação impositiva sobre tarifas, porém sempre observando as regras e procedimentos de comercialização definidos pela ANEEL.

A entidade que contabiliza os contratos de transação de energia em ambos os mercados ACR e ACL Brasileiro é a Câmara de Comercialização de Energia Elétrica (CCEE). Esses registros, como será explicado a seguir, servem de base para a contabilização e liquidação entre os agentes do mercado nos ajustes de produção e consumo de energia dos contratos estabelecidos. Na CCEE, todos os agentes que estão envolvidos nas transações do mercado de energia elétrica registram os montantes de energia e o período de vigência dos contratos estabelecidos. Os preços não são registrados e são uma informação exclusiva dos contratos firmados entre as partes envolvidas. Os participantes do mercado, que fazem consequentemente os registros das negociações, são as empresas geradoras, distribuidoras, comercializadoras e consumidores livres de energia elétrica.

A CCEE centraliza as informações das negociações compreendendo os quatro mercados regionais (ou subsistemas): Sul, Sudeste/Centro-Oeste, Norte e Nordeste. Semanalmente as diferenças entre as quantidades de energia gerada, contratada e consumida são quantificadas no ambiente da CCEE, dividido nos três patamares de energia diários e nos quatro mercados regionais.

O preço de referência estabelecido para cada uma dessas subdivisões (de patamares e de submercados) é chamado de Preço de Liquidação das Diferenças (PLD), e representa o que é

conhecido como o *proxy* do preço *spot* da energia elétrica no mercado livre brasileiro. Esses preços são calculados semanalmente como o resultado do mínimo custo operacional para suprir toda a demanda e para oferecer um suprimento de energia futuro com preços e riscos de disponibilidade controlados. Dessa forma o PLD é um *output* gerado por modelos computacionais operados pelo Operador Nacional do Sistema Elétrico (ONS) para controle do despacho das usinas geradoras, com algumas adaptações necessárias para o estabelecimento de preços comuns de transação pela CCEE.

Sabendo-se que no mercado brasileiro existe uma preponderância de usinas hidrelétricas no parque de geração de energia elétrica, para o cálculo do PLD são utilizados algoritmos que de forma geral objetivam encontrar uma solução ótima para a questão de se precificar o equilíbrio entre armazenar água para uma utilização futura, ou utilizá-la no momento atual [4].

O dilema a ser sanado para a definição do PLD pode ser descrito, de forma resumida, da seguinte forma: A primeira vista a utilização da energia total acumulada nos reservatórios das usinas hidrelétricas é a forma mais econômica de se garantir a energia elétrica para atender a demanda, pois minimiza os custos de utilização de outros combustíveis mais caros do que a água, como combustíveis fósseis utilizados por usinas termelétricas, por exemplo. Essa estratégia, porém, deixaria o sistema mais exposto a riscos de abastecimentos futuros, dado que são imprevisíveis a quantidade de chuvas no longo prazo, influenciando no nível que os reservatórios estarão quando forem necessários. Dessa forma, tem-se a maior confiabilidade possível no fornecimento quando é deixado constantemente um elevado nível de água nos reservatórios, e isso só é possível poupando a energia hidrelétrica e utilizando-se a energia térmica que tem como característica altos custos de produção devido a custos de combustível e manutenção.

Valendo-se dessas variáveis, ou seja, das condições hidrológicas estimadas presentes e futuras, dos preços dos combustíveis, das simulações do custo da falta de energia, da possibilidade de entrada de novas fontes geradoras, das possíveis novas interligações entre submercados, da disponibilidade e custos relacionados à aquisição de novos equipamentos (geração, transmissão e distribuição), além de outros fatores, o modelo de precificação determina o despacho ótimo para cada central hidrelétrica, definindo o quanto cada central hidrelétrica ou termelétrica irá contribuir para o fornecimento final atual. Ao final desse processo estabelece-se o Custo Marginal de Operação (CMO) para o período, dividido nos diferentes horários de consumo (patamares) e submercados.

Os dados utilizados pelo ONS para otimização da operação do Sistema Interligado Nacional (SIN) e o cálculo do CMO são utilizados pelo CCEE para estabelecer o PLD, que é o valor determinado semanalmente para cada patamar, no qual o custo efetivo fica limitado a um preço máximo e mínimo vigentes para cada um dos submercados. Quando se calcula o PLD não são consideradas as restrições de transmissão internas em cada um dos submercados, fazendo com que a energia comercializada dentro de um mesmo submercado seja tratado como igualmente disponível em todos os pontos finais de consumo, já as restrições externas de interligação de submercados são consideradas nos preços finais. Como resultado tem-se que é gerado um preço único dentro de um mesmo submercado.

Existem usinas que apresentam limites para o cumprimento do despacho por parte do ONS devido a limites operacionais. Essas usinas são consideradas inflexíveis, e sua parte inflexível não é considerada para o estabelecimento do preço naquele submercado, só são abatidas no volume necessário para o abastecimento daquela região. Caso a usina ultrapasse o nível considerado inflexível, esse preço será considerado para a formação do preço.

Com isso o cálculo do preço é baseado no despacho “ex-ante”, sendo apurado com as previsões disponíveis das variáveis em comum acordo, antes da operação real do sistema, e fazendo-se uma análise da capacidade de produção do sistema, e da demanda que virá nas datas futuras para cada região. Para o cálculo efetivo do PLD são utilizados modelos computacionais, no caso o NEWAVE e o DECOMP, que tem como *outputs* o CMO de cada submercado. Esse nada mais é que o custo operativo do último gerador despachado, e por ser um sistemas de despacho por ordem de mérito, é o mais caro para atender uma determinada carga. Mais informações sobre o cálculo do PLD são providos em [2].

O PLD é o preço de referência utilizado para compra e venda de energia no mercado de curto prazo para um determinado submercado, ou seja, aquela energia que por algum motivo foi sobrecontratada ou subcontratada e precisa ser liquidada pelos compradores e vendedores no ambiente provido pelo CCEE. Esse preço também é a referência utilizada pelos mercados do ACL e ACR para a precificação de contratos futuros, através de leilões ou contratos bilaterais entre os agentes, operando com ágios ou deságios sobre o PLD de acordo com volumes, prazos e riscos de contraparte.

2.1.1 Estudo de caso de precificação

Na literatura brasileira não estão disponíveis muitos artigos que tratam especificamente da modelagem do preço da energia elétrica com objetivos de *pricing* no mercado Brasileiro. No que foi observado em [16] e [18], são apresentados os modelos com as seguintes componentes:

- Movimento Browniano Geométrico: Clássico processo difusivo log-normal.
- Processo de reversão à média: Processo difusivo com reversão à um valor médio.
- Processo com reversão à média com Saltos Aleatórios: Além da componente de reversão à média são incorporados ao modelo uma distribuição separada que representa picos esporádico de preços. Normalmente esses picos esporádicos são oriundos de um processo de Poisson incorporado ao modelo, porém podem ser introduzidos através de um modelo de troca de regime onde se consegue modelar fases nas quais os preços possuem uma volatilidade maior.
- Modelos estocástico de troca de regime: Através de um modelo de Mudança de Regime de Markov é introduzida uma mudança de regime estocástica, que incorpora ao modelo fases de reversão à média e fases de ”stress” com picos de preços.

Dessa forma, os modelos encontrados no mercado nacional são modelos muito semelhantes ao que será apresentado a seguir no modelo Australiano na seção 2.3, com relação a sua natureza

estocástica, ou seja, um modelo de troca de regimes de reversão à média e processos difusivos à saltos. Em [18] é concluído que o modelo que apresentou melhor aderência aos dados reais é o modelo estocástico com reversão à média, mudança de regime markoviano e difusão com saltos dependentes. Segundo esse artigo, o modelo tem sucesso em explicar a volatilidade encontrada no CMO de cada submercado do SIN.

Contudo, uma componente determinística bem definida que incorpore as principais tendências sazonais desses preços ao longo dos meses do ano não está presente em nenhum dos modelos verificados, como se encontra frequentemente para outros mercados. Tendo em vista que, para o mercado brasileiro, o preço *spot* é um reflexo do CMO, que por sua vez é altamente correlacionado com fatores como a distribuição de chuvas durante o ano, a primeira impressão é de se conseguir encontrar componentes determinísticas sazonais muito representativas para os modelos que procuram explicar os preços *spot* para os submercados brasileiros. Porém, como será verificado nos capítulos seguintes, isso não é o que acontece de fato.

2.2 Mercado dos Países Nórdicos

O mercado nórdico está em um processo de desregulamentação desde os anos noventa, na busca de abrir ao máximo os diversos setores da eletricidade para a competição. A rede de transmissão é operada por agentes independentes, que são monitorados e regulados por autoridades públicas. Todos os participantes do mercado podem fazer parte desse sistema de energia, podendo operar através de contratos bilaterais ou através de contratos na bolsa de energia local o *Nordic Power Exchange* (NPE).

Historicamente, o NPE iniciou suas operações em 1993 no mercado da Noruega, em seguida entrou no mercado a Suécia em 1996, a Finlândia em 1998, a parte oeste da Dinamarca em 1999 e a parte leste da Dinamarca em 2000. Basicamente o NPE organiza dois mercados, o mercado físico (*Elspot*) e o mercado financeiro (*Eltermin* e *Eloption*). O NPE também oferece serviços de *Clearing* na liquidação de seus contratos. Esses mercados funcionam todos os dias do ano (24/7).

Em 2007, com dados retirados de [16], os volumes operados no NPE foram de 1.352 TWh (movimentando cerca de EUR 52,8 bi) enquanto em contratos bilaterais foram negociados 1.309 (movimentando cerca de EUR 38,4 bi).

No Elspot existe a comercialização no mercado *spot* para energia a ser entregue fisicamente no dia seguinte. Essa energia é vendida para cada uma das 24 horas do dia seguinte. Cada um dos contratos no Elspot diz respeito ao volume a ser entregue em megawatts por hora (MWh), em uma determinada hora, por um determinado preço por MWh. O preço, chamado de "Sistema de Preço" é separado para cada uma das horas do dia seguinte, e é balanceado independentemente pela oferta e demanda de cada horário. Nesses mercados não são considerados os limites de carga suportados na rede entre as nações que estão cobertos pelo NPE. Com isso, o sistema de preços pode ser considerado um mercado respaldado por uma *clearing* onde os participantes operam eletricidade para toda a rede coberta, sem considerar as limitações

da rede transmissão. Esses preços, que são altamente confiáveis e transparentes, são também utilizados como referência nas negociações do mercado financeiro do NPE.

Para melhor balanceamento dos contratos físicos no curto prazo o Elspot oferece também um mercado intradiário (também denominado mercado H+1) que se encerra, para um determinado horário, algumas horas antes da entrega física. Ainda existe um mercado de ajuste fino *ex-post*, ou mercado de balanço, que ocorre após o fechamento do mercado *spot* para determinado horário, no qual os agentes fazem ofertas de compra e venda de volumes de energia para consumo imediato. Esse mercado permite o balanceamento das posições em tempo real.

Para lidar com os problemas de limite de transmissão, o sistema independente de cada nação desenvolveu estratégias próprias para cada área. Limites de transmissão entre duas nações, assim como limites dentro da Noruega, são gerenciados através de mecanismos de preço no mercado *spot*, implicando em ajustes de preços nas áreas envolvidas. Dessa forma, quando existem problemas de excesso do limite de transmissão, são criados preços por zonas, preços esses independentes ao Sistema de Preços. Na realidade, o NPE é separado em zonas de negociação para a oferta de geração e demanda para cada hora. Suécia, Finlândia e Dinamarca são sempre zonas de negociações independentes. Os limites de transmissão internos dentro de um mesmo país são gerenciados diretamente pelos operadores nacionais, e o custo da regulação são financiados através de tarifas no custo de transmissão.

Um sistema elétrico precisa ser constantemente balanceado. Para lidar com qualquer diferença não prevista entre o que foi planejado e o que é de fato intercambiado durante o dia, uma vez que o mercado *spot* do dia seguinte já está fechado, os operadores utilizam o recurso dos mercados de regulação (ou mercados de balanceamento) nos quais a necessidade de mais carga ou menos carga são obtidos em um curto espaço de tempo. No NPE isso acontece através do mercado chamado *Elbas*, um mercado físico de curto prazo lançado em 1999 pelo *Finish electricity exchange El-Ex* que permite aos operadores ajustarem as suas posições após o "mercado do dia seguinte" fechar.

No mercado financeiro do NPE, conhecidos como *Eltermin* e *Eloption*, é permitida a operação de contratos financeiros como futuros e termos com períodos de entrega de até três anos. Esses mercados não operam mercados físicos, e são todos acertados financeiramente baseados no sistema de preço do mercado *spot*. Eles são divididos em contratos de 1 MW para cada hora do dia para um período de entrega que pode ser de um dia, uma semana, um bloco (quatro semanas), uma temporada do ano ou um ano. Esses contratos podem ser negociados em bolsa, dependendo do tipo de contrato. Opções europeias e asiáticas também são operadas desde 1999.

Os contratos futuros sazonais são divididos em 3 estações no ano. O primeiro contrato são para as 16 primeiras semanas do ano, o segundo vai da semana 17 até a semana 40 e o terceiro da semana 41 até a semana 52/53. Esses são divididos em blocos (normalmente de 4 semanas) que por sua vez são divididos em contratos semanais. Esses por sua vez são divididos em contratos diários. Cada um desses contratos podem ser operados até alguns dias antes do seu início efetivo, como para o caso dos contratos de estações, que podem ser operados até o início de duas estações para trás. Já nos contratos diários podem ser operados até um dia antes

da sua entrega efetiva.

Contratos a termo com periodicidade anuais podem ser operados até 2 dias antes da sua entrega efetiva. Os contratos a termo sazonal podem ser operados até o dia da sua entrega efetiva, sendo que esses contratos são divididos em 3 períodos do ano também. São eles, Inverno 1 (com período de entrega de 1º de Janeiro a 30 de Abril), Verão (de 1º de Maio a 30 de Setembro) e Inverno 2 (de 1º de Outubro a 31 de Dezembro).

O mercado de opções tem como característica a abertura de uma posição no mercado a termo, no caso de exercício. As opções asiáticas são liquidadas financeiramente, onde é utilizado como referência a média aritmética do preço para o período de vigência do contrato base.

Os serviços de clearing são oferecidos por uma empresa separada chamada *Nordic Electricity Clearing* (NEC), que garante o pagamento e a entrega de todos os *trades* feitos no mercado, entrando como contraparte em todos os contratos fechados nesse mercado, tanto para venda como para compra. Ela pode prestar serviços de clearing também em contratos bilaterais.

2.2.1 O sistema de geração do Nord Pool

De acordo com os dados do operador do sistema do Nord Pool (TSO e Nord Pool 1999 [21]) o sistema de geração varia entre os países do Nord Pool. Quase toda a energia da Noruega é de origem hidrelétrica (99%). A Suécia possui uma distribuição quase igual entre energia hidrelétrica (48%) e energia nuclear (46%), com uma produção residual de energia térmica (gás e óleo) (6%). A Finlândia já tem uma matriz mais variada utilizando energia hidrelétrica (22%), nuclear (31%) e energia termelétrica baseadas em vários tipos de combustíveis (47%). Já na Dinamarca a base energética mais forte é a termal (93%) utilizando gás natural e carvão como combustível e também energia eólica (7%).

2.2.2 Estudo de caso

Em suas pesquisas Lucia e Schwartz [15], utilizam dados do *Elspot* de 24 séries temporais, uma para cada hora do dia, para todos os dias da semana, registrado por 7 anos começando de 1 de Janeiro de 1993 e indo até 31 de Dezembro de 1999. Nessas séries eles encontraram que qualquer série de preços para um determinado horário quando comparado em pares se mostra muito correlacionados entre si, com todas as correlações acima de 0.94 e com uma média de 0.98. Isto é, os preços em geral são altamente correlacionados e não haveria perda significativa em considerar o preço *spot*, em análises de longo prazo, como uma média dos preços *spot* intradiários. Dessa forma, Lucia e Schwartz consideram em seus estudos o preço *spot* como uma média aritmética de todos os preços de um determinado dia.

Lucia e Schwartz verificam também uma alta volatilidade anualizada de 189% ao ano, com as temporadas frias (252%) 2 vezes mais voláteis que as temporadas temperadas (128%). A estatística sobre os dados reflete que altos preços são relativamente frequentes, pois a série histórica possui uma curtose de 3.5, enquanto de uma amostra que segue uma distribuição normal é próxima a 3. Isso significa que preços muito altos ou muito baixos são mais frequentes

que o esperado para uma amostra normal com o mesmo desvio padrão. Sobre a simetria, pode-se afirmar que picos de preços altistas são mais frequentes que os picos negativos, devido ao sinal positivo desse fator.

Como será verificado na seção 4.2, muitas dessas características, como a volatilidade relativamente alta, o perfil leptocúrtico (distribuição com caudas pesadas, características de amostras com curtoses superiores a 3) e simetria positiva são encontradas também para o mercado brasileiro. Essas são características de distribuições de preços apresentam modelos mais aderentes levando-se em conta processos a saltos e de reversão à média, como é muito frequente em modelos de preço *spot* de energia elétrica.

Uma análise da curtose separada para as estações do ano demonstra que os maiores saltos do preço da energia estão nas temporadas frias (com curtose quatro vezes mais altas que as temporadas quentes). Esses saltos em preços são provocados pela alta demanda de energia nas temporadas mais frias, obrigando o sistema a utilizar geradores de energia elétrica com custos marginais de operação mais altos, principalmente devido aos custos de combustíveis mais caros, necessário para supri-lo.

Uma observação importante que se faz é que os preços do mercado futuro e a termo dos mercados de bolsa incorporam também a sazonalidade que se observa no preço spot, isso é verificado através observações empíricas das curvas de mercado futuras, onde se observa de forma clara a característica sazonal dos preços.

2.2.3 Modelo para preço spot

Lucia e Schwartz [15] elaboram modelos que incorporam dois tipos de componentes. Uma componente determinística que modela qualquer movimento periódico genérico, ou uma tendência já esperada, como o movimento sazonal do preço da energia elétrica. A segunda componente é estocástica e é modelada através de um processo em tempo contínuo.

A componente determinística é formada por uma componente totalmente previsível dada pela função $F = f(t)$. A segunda componente é formada por um processo estocástico difusivo X_t . Dessa forma tem-se o modelo dado por:

$$P_t = f(t) + X_t \quad \text{onde } t \in [0, \infty) \quad (2.1)$$

O modelo de X_t segue o seguinte processo estocástico contínuo:

$$dX_t = -\alpha X_t dt + \sigma dZ \quad (2.2)$$

onde $\alpha > 0$, $X(0) = x_0$ e dZ é o incremento de um processo de movimento Browniano padrão Z_t . Dessa forma, X_t segue um processo de reversão à média com o preço de longo termo igual a zero e taxa de reversão à média α . Considerando que $F = f(t)$ satisfaz as propriedades de regularidade, podemos escrever:

$$d(P_t - f(t)) = \alpha(f(t) - P_t)dt + \sigma dZ \quad (2.3)$$

A partir dessa equação verifica-se que quando o preço P_t desvia do termo determinístico $f(t)$ ele tende a voltar a uma taxa de reversão à média proporcional ao desvio. Nesse modelo, a única fonte de incerteza é o fator X_t , e com isso podemos chamá-lo de variável de estado.

Pode-se reescrever a equação 2.3 da seguinte forma:

$$dP_t = \alpha(h(t) - P_t)dt + \sigma dZ \quad (2.4)$$

onde $h(t)$ é dado pela equação determinística abaixo:

$$h(t) = \frac{1}{\alpha} \frac{df}{dt}(t) + f(t) \quad (2.5)$$

Essa é uma versão estendida do modelo de Vasicek, sugerido por Hull and White (1990) [11].

Nesse caso, tem-se que a equação 2.2 é de solução já conhecida e pode ser dada pela seguinte equação, combinado com 2.1 :

$$P_t = f(t) + X_0 e^{-\alpha t} + \sigma \int_0^t e^{\alpha(s-t)} dZ(s) \quad (2.6)$$

Dessa forma a distribuição condicional de P_t é normal com média e variância dadas por:

$$\begin{aligned} E_0(P_t) &\equiv E(P_t|X_0) = f(t) + (P_0 - f(0))e^{-\alpha t} \\ Var_0(P_t) &\equiv Var(P_t|X_0) = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}) \end{aligned} \quad (2.7)$$

A partir disso fica claro que o preço P_t tende ao seu fator determinístico no longo prazo, dado P_0 , e a convergência será mais rápida conforme α for maior. A variância tem um limite finito no longo prazo.

Como o objetivo é a precificação de derivativos, seguindo os argumentos de Cox e Ross em [5] e Harrison e Kreps em [8] torna-se necessário a construção de um processo neutro ao risco para X_t . Para isso se utiliza argumentos de não arbitragem entre derivativos, permitindo assim obter o processo neutro ao risco X_t . Isso é dado pela equação:

$$\begin{aligned} dX_t &= \alpha(\mu^* - X_t)dt + \sigma dZ^* \\ \text{sendo } \mu^* &\equiv -\lambda\sigma/\alpha \end{aligned} \quad (2.8)$$

Nesse caso tem-se que dZ^* é um incremento de Z_t^* , um movimento Browniano na medida de probabilidade livre de risco, e λ é uma constante que representa o preço de mercado por unidade de risco ligado à variável de estado X_t .

A partir do processo neutro a risco também chega-se uma solução analítica para a equação 2.8, obtendo-se:

$$P_t = f(t) + X_0 e^{-\alpha t} + \mu^*(1 - e^{-\alpha t}) + \sigma \int_0^t e^{\alpha(s-t)} dZ^*(s) \quad (2.9)$$

nesse caso, tem-se que P_t é normal na medida livre de risco, com o valor esperado dado abaixo:

$$E_0^*(P_t) \equiv E^*(P_t|X_0) = f(t) + (X_0) e^{-\alpha t} + \mu^*(1 - e^{-\alpha t}) \quad (2.10)$$

A partir da construção da medida de probabilidade livre de risco, o valor de qualquer derivativo será igual ao *payoff* descontado até presente pela taxa livre de risco, que para o presente estudo é constante. Dessa forma o valor no tempo zero de um contrato a termo, sobre um ativo *spot* com vencimento em T , será dado por:

$$\nu_0(X_T, T) = e^{-rT} E_0^*[P_T - F_0(P_0, T)] \quad (2.11)$$

Nesse caso o preço $F_0(P_0, T)$ é o preço em $t = 0$ para o termo com vencimento em T , sendo r é a taxa de juros livre de risco. Uma vez que o preço do contrato a termo deve ser zero quando ele é criado pela primeira vez, pode-se concluir a seguinte solução fechada para os preços dos contratos a termo (ou futuros), para $t = 0$, utilizando-se 2.10 e 2.1:

$$\begin{aligned} F_0(P_0, T) &= E_0^*(P_T) = f(T) + (P_0 - f(0))e^{-\alpha T} + \mu^*(1 - e^{-\alpha T}) \\ \mu^* &= -\lambda\sigma/\alpha \end{aligned} \quad (2.12)$$

Com relação ao termo determinístico, a ideia é especificar o termo $F = f(t)$ da equação 2.1. O principal objetivo da função é o de capturar qualquer padrão que seja comum ou previsível no seu comportamento futuro. Como pensamento inicial poderia ser considerada uma função constante, por exemplo, que desconsideraria qualquer comportamento sazonal relevante.

Em seu livro, Pilipovic 1998 [19] propõe um modelo senoidal como guia para a componente determinística do seu modelo. Em sua pesquisa quando analisa outros ativos do mercado americano, como do Índice S&P 500 e do preço do gás natural, fica muito evidente a presença de componentes senoidais na composição dos preços, indicando a presença de uma sazonalidade para o comportamento da energia elétrica. Em seu livro é proposto também a utilização de duas senoidais de normas diferentes, para tentar capturar os picos de inverno e verão nos preços, devido ao costume de utilização de calefação e ar condicionado respectivamente. Dessa forma, é possível se observar que as funções de comportamento determinístico que caracterizam o comportamento dos preços e energia elétrica podem estar relacionadas a questões geográficas, principalmente devido a questões climáticas específicas de uma determinada região.

Verifica-se na literatura também a utilização de funções degraus para aproximar os componentes periódicos (como em Jaillet, P., e Tompaidis S. (1998)[13]), através de variáveis *dummies* para fixar o comportamento característico nos diferentes meses do ano. Um problema observado no uso das *dummies*, é que apesar da sua fácil interpretação e flexibilidade, elas são muito sensíveis a ruídos na amostra e *outliers*. Além disso, os pontos onde são fixados os degraus das *dummies* são escolhidos arbitrariamente por nós, muitas vezes através de observação empírica ou comodidade de interpretação temporal (como os meses do ano), e essa escolha pode levar a afetar negativamente a parametrização.

No modelo inicial proposto pelo artigo, foram colocados dois termos na função determinística além da constante. Os termos colocados tentam capturar a influência dos preços entre dias úteis e não úteis e a evolução dos preços devido a fatores sazonais de acordo com os meses do ano. Dessa forma, são propostas duas versões para a função determinística, sendo a primeira delas:

$$f_1(t) = \mu_0 + \beta D_t + \sum_{i=2}^{12} \beta_i M_{it} \quad (2.13)$$

Onde as variáveis *dummies* são dadas por:

Uma para capturar as diferenças nos preços entre dias da semana e finais de semana e feriados D_t :

$$D_t = \begin{cases} 1, & \text{se a data } t \text{ é um feriado ou fim de semana} \\ 0, & \text{se for dia de semana} \end{cases}$$

E outra para capturar as diferenças nos preços entre os meses do ano M_{it} :

$$M_{it} = \begin{cases} 1, & \text{se a data } t \text{ corresponde ao mês } i \\ 0, & \text{se não} \end{cases}$$

os parâmetros μ_0 , β e β_i for $i=2\dots 12$ são todos parâmetros constantes.

A interpretação para os parâmetros β_s é que eles procuram capturar as mudanças provocadas pelos finais de semanas e feriados, e dos meses de fevereiro a dezembro, quando comparados com o modelo base (constante) representado pelos dias úteis do mês de janeiro.

A função determinística unicamente através de variáveis *dummies* é uma solução viável pela fácil interpretação, porém na sua pesquisa Lucia e Schwartz 2001 [15] propõe também outro modelo para a função determinística, através de uma função cossenoidal que diminui bastante o número de variáveis em comparação com o modelo baseado em variáveis *dummies* mensais. Essa função é representada por:

$$f_2(t) = \mu_0 + \beta D_t + \gamma \cos\left((t + \tau) \frac{2\pi}{365}\right) \quad (2.14)$$

onde D_t segue a mesma definição do modelo acima.

A função cosseno é medida em radianos e μ_0 , β , γ e τ são todos constantes. Da mesma forma, o coeficiente β captura as modificações provocadas pelos finais de semana e feriados. A função cosseno procura refletir a variação sazonal dos preços, dentro de um período anual.

A partir dos preços *spot*, estima-se o processo estocástico para o modelo. Por se tratarem de observações discretas, o modelo deverá ser reescrito totalmente em formato discreto. Reescrevendo-se então a equação 2.2 em termos discretos:

$$X_t = (1 - \alpha_0)X_{t-1} + \varepsilon_t \quad (2.15)$$

sendo $t = 0, 1, 2, \dots, N$ e ε_t uma função independente e identicamente distribuída seguindo distribuição normal de média zero e variância σ^2 .

Dessa forma tem-se 2 modelos que se diferem da forma como a função determinística é fixada:

Modelo 1 :

$$\begin{aligned} P_t &= \mu_0 + \beta D_t + \sum_{i=2}^{12} \beta_i M_{it} + X_t \\ X_t &= (1 - \alpha_0)X_{t-1} + \varepsilon_t \end{aligned} \quad (2.16)$$

Modelo 2 :

$$\begin{aligned} P_t &= \mu_0 + \beta D_t + \gamma \cos\left((t + \tau) \frac{2\pi}{365}\right) + X_t \\ X_t &= (1 - \alpha_0)X_{t-1} + \varepsilon_t \end{aligned} \quad (2.17)$$

Os modelos foram estimados simultaneamente através do método dos mínimos quadrados não linear, e são escritos de maneira genérica da seguinte forma:

$$\begin{aligned} y_t &= f(\Theta, x_t) + X_t \\ X_t &= \phi X_{(t-1)} + \varepsilon_t \end{aligned} \quad (2.18)$$

Onde $\phi = 1 - \alpha_0$

Nessas equações y_t representa a variável preço, como uma função dos parâmetros vetores Θ e das variáveis x_t . A segunda equação representa o ruído e é uma equação autorregressiva de primeira ordem. Rearranjando a equação 2.18 tem-se:

$$y_t = \phi y_{t-1} + f(\Theta, x_t) - \phi f(\Theta, x_{t-1}) + \varepsilon_t \quad (2.19)$$

E nesse caso todos os parâmetros ϕ e Θ são estimados simultaneamente usando técnicas numéricas do método de mínimos quadrados não linear. Estima-se o parâmetro α_0 a partir de $\hat{\alpha}_0 = 1 - \hat{\phi}$ e usa-se o desvio padrão da regressão para se estimar σ .

Nos dois modelos o parâmetro constante μ_0 é significativamente diferente de zero, ou seja, existe um preço médio que reflete realmente os custos operacionais de produção de energia (custos básicos). Para todos os parâmetros β tem-se valores diferentes de zero também, porém nem todos possuem níveis de significância representativos, o que pode configurar a eliminação de alguns parâmetros na busca de melhorar a aderência do modelo (conforme seção 3.2.3).

A estimação dos parâmetros β , ϕ e σ para os dois modelos são muito semelhantes entre si. O teste de hipótese para $\phi = 1$, o que representaria que o fator de reversão à média é zero, ou seja, não representativo, é rejeitado, ou seja, apesar de pequeno ele é representativo para o modelo apresentado.

Os principais resultados estatísticos relativos à sazonalidade do modelo são apresentados no Diagrama I abaixo, extraído e adaptado do estudo Lucia e Schwartz [15]. Os valores correspondentes aos p-values do Diagrama I foram aproximados a partir das estatísticas-t presentes no estudo.

Diagrama I: Resultados do modelo Nord Pool		
Parâmetro	Estimação (NOK)	p-value
β_0	153.06	< 0.0025
β_2	-2.53	0.01
β_3	-4.51	0.10
β_4	-3.84	0.25
β_5	-13.25	0.01
β_6	-12.66	0.01
β_7	-7.04	0.10
β_8	-8.11	0.10
β_9	-10.06	0.05
β_{10}	-9.60	0.05
β_{11}	-7.30	0.05
β_{12}	-6.02	0.05
Desvio padrão		9.001

No diagrama acima, a partir dos p-values relativamente baixos apresentados, nota-se de forma geral que o modelo possui uma boa aderência a aspectos sazonais. O p-value do diagrama é o teste de Fisher-Snedecor (ou *F-statistics*) apresentado na seção 3.1.4 feito para as variáveis β individualmente. Nesse teste, o que se revela na prática é que quanto menor o p-value, dado um nível de significância de referência, menos se consegue aceitar a hipótese H_0 de rejeição do parâmetro que está sendo testado. Dessa forma, de forma exemplificativa para o parâmetro β_4 , por possuir um valor de p-value alto de 0.25, esse parâmetro poderia ser retirado do modelo sem grandes perdas de aderência. O mesmo ocorre para os parâmetros β_3 , β_7 e β_8 que, dependendo do nível de significância escolhido, poderiam também serem excluídos sem grandes prejuízos.

Através da pesquisa feita, por Lucia e Schwartz (2001) [15] pode-se concluir que para o *Nord Pool* existe uma clara tendência do comportamento dos preços seguirem fatores sazonais, e isso acaba influenciando diretamente na formação dos preços dos contratos futuros e a termo para esses mercados. Encontrou-se porém muita dificuldade na estimação dos parâmetros na busca de se encontrar a verdadeira tendência desses modelos. Em seus resultados, mostrou-se necessário uma grande amostra de dados para se conseguir parametrizar corretamente os modelos, e não foram para todos os casos que esses dados estavam disponíveis suficientemente.

A técnica utilizada foi estimar a componente determinística dos modelos para o preço *spot*, e a partir disso utilizar esse modelo para precificação de contratos derivativos. Uma técnica alternativa que poderia ter sido utilizada é a utilização dos próprios preços futuros para se parametrizar um modelo, e a partir disso utilizar esse modelo para precificar outros derivativos mais complexos como opções ou *swaps*. Esse modelo porém esbarra falta de liquidez de alguns contratos futuros, tornando-se difícil a montagem de uma curva futura estimada bem embasada.

Outra conclusão apresentada é que a volatilidade para as temporadas frias e quente são claramente diferentes, sendo que no modelo de reversão à média essa tendência precisou ser considerada para a própria aderência do modelo. Nota-se que em seus estudos não foram

utilizados processos a saltos, que podem também estar presentes e afetarem significativamente os preços.

Mostrou-se também que fatores geográficos e sazonais afetam diretamente os preços dos ativos, sendo que cada região dentro do Nord Pool precisou ter um modelo próprio especificado. Por ser um mercado com forte presença de geração hidrelétrica, esperava-se encontrar semelhanças capazes de tornarem o modelo dos países nórdicos aplicável ao brasileiro, nota-se porém que existe uma forte diferença no padrão de consumo principalmente devido a fatores climáticos e industriais. No capítulo 4 será verificado que o modelo sazonal determinístico não é de fato aplicável quando visto isoladamente no mercado brasileiro.

2.3 Mercado Australiano

Segundo Higgs e Worthington (2006) [10], o mercado de eletricidade australiano passou por um processo de desregulamentação de forma pioneira, na mesma tendência do Reino Unido, americanos e o próprio brasileiro. Em seu artigo, Higgs e Worthington especificam as principais características observáveis no preço *spot* de energia no mercado australiano:

- Alta volatilidade, devido a alta variância dos seus preços históricos
- Uma forte reversão à média, com os preços variando ao redor de um equilíbrio de longo prazo
- Saltos, ou picos altistas que decaem rapidamente. Esses picos estão associados à inelasticidade característica do mercado de energia elétrica entre a oferta e a demanda.

Essas características definem o mercado de energia como um todo e dessa forma o preço reflete os padrões verificados no mercado australiano, sendo eles a sazonalidade, o baixo custo marginal, o impacto de possíveis quedas de energia, problemas na interconexão de sistemas, capacidade de armazenamento limitada e as eventuais manipulações de mercado.

Higgs e Worthington defendem que a gestão do risco devido às incertezas do sistema fazem com que todos os participantes do mercado, entre eles geradores, distribuidores e usuários finais, despendam muitos recursos. Isso para se proteger da volatilidade do preço *spot* no mercado Australiano, que ultrapassa 900% quando calculados da forma costumeira (desvio padrão dos diferenciais de preços logarítmicos).

Dessa forma, o modelo de Higgs e Worthington parte da premissa já conhecida de que os preços para o mercado de eletricidade demonstram uma alta volatilidade, forte reversão à média e frequentemente apresenta saltos em seus preços. A característica do modelo apresentado em seu artigo é que ele possui uma componente estocástica, uma componente de reversão à média e uma componente de troca de regime, como serão melhor definidos a seguir. A troca de regime é uma tentativa de modelar as especificidades do mercado australiano, composto pelos submercados de *New South Wales (NSW)*, *Queensland (QLD)*, *South Australia (SA)* e *Victoria (VIC)*.

Um ponto interessante observado é que os preços tendem a possuir uma forte reversão à média após os saltos de preços, e a volatilidade é quatorze vezes maior em períodos com frequência de saltos. Observou-se também que para o mercado australiano, a probabilidade de haver saltos variam entre 5.16% e 9.44% para um dado dia, dependendo do submercado que se está analisando.

No mercado australiano existe a utilização das *peaking plants*, que são as usinas que geram energia nos momentos de picos de preços da energia elétrica. Essas usinas possuem como características altos custos marginais e a capacidade de recuperar custos incorridos em curtos períodos de tempo. Segundo Higgs e Worthington o sistema australiano se beneficiaria com um bom modelo para os saltos nos preços da energia elétrica proporcionados pela utilização dessas usinas menos eficientes do ponto de vista econômico. Por exemplo, as indústrias conseguiriam fazer uma melhor gestão de sua energia nesses momentos de picos, além dos próprios revendedores que poderiam se proteger desses preços através de derivativos no mercado. Com um modelo eficiente que leva a utilização dos *peaking plants* em consideração, poderiam ser precificados derivativos para se proteger dos elevados custos de energia nos períodos de maior utilização e os seus altos preços associados.

Os processos de reversão à média com saltos são limitados em alguns aspectos quando estão presentes em modelos de energia elétrica. Primeiramente, percebe-se que esses modelos são aderentes para se modelar operação com moedas, e ações, por exemplo, onde existem saltos eventuais e depois uma reversão vagarosa na direção de um equilíbrio de longo prazo. Para o mercado de energia elétrica, pelo contrário, o retorno de picos de preços ao seu valor de equilíbrio ocorre em poucos dias ou mesmo horas. Higgs e Worthington acreditam que com modelos simples de reversão à média com saltos, o comportamento do preço de energia elétrica só poderia ser simulado através de altos e irrealísticos parâmetros de reversão à média. Dessa forma, como será verificado em maiores detalhes a seguir, eles apresentam um modelo que utiliza a reversão a média, porém conseguem controlar de melhor forma os saltos e retornos dos períodos de maior volatilidade através de um modelo de troca de regime.

Outro ponto levantado em Higgs e Worthington 2006 [10] é que os processos de saltos são assumidos constantes no tempo, porém o que se observa no mercado de energia elétrica é que tipicamente existe uma alternância entre períodos de alta e baixa frequência de saltos.

Frequentemente encontram-se autores dissociando os processos de saltos dos processos de reversão a média, utilizando para isso modelos com mudança de regime. Seguindo essa linha Higgs e Worthington propõe um modelo que comporta três regimes: Um primeiro regime com moderada reversão à média e volatilidade, um segundo regime que caracterizaria um período de saltos altistas e na sequência um terceiro regime com saltos baixistas, ou reversão à média mais forte, que forçam os preços a ficarem próximos aos preços médios característicos da sazonalidade. A vantagem desse modelo é que são separados cada um dos comportamentos que se observam nos preços. Uma desvantagem é que comportamentos anormais, como sucessivas fases de saltos altistas, por exemplo, não são bem capturados por esses modelos.

2.3.1 Dados e estatísticas

No estudo de Higgs e Worthington 2006 [10] foram utilizados os preços *spot* do Mercado Nacional de Eletricidade Australiano para o período de 5 anos (de 01/01/1999 até 31/12/2005). Os dados de preços são do *National Electricity Market Management Company* (NEMMCO 2005) em uma base de meia hora, representando 48 intervalos de negociação em cada período de 24 horas. Utilizou-se a média aritmética para cada um desses dias como base para o trabalho.

O modelo foi baseado em dados altamente voláteis, com a média aritmética diária girando no entorno dos US 35,00/MWh. A análise estatística dos dados apresentou uma simetria positiva, o que indica haver também para esses modelos uma tendência a haver saltos altistas maiores e mais intensos, além de seguir uma distribuição leptocúrtica, indicando uma maior frequência para valores muito próximos à média e mais distantes da média e menos frequentes em valores intermediários.

O teste de hipótese para verificar a possibilidade de distribuição normal nos preços foi rejeitado a um nível de confiança de 10%. Além disso, o trabalho afirma que para o mercado Australiano, os preços possuem uma tendência de serem estacionários.

2.3.2 Especificação do modelo

Na metodologia utilizada por Higgs e Worthington 2006 [10], os preços são modelados por duas componentes, uma determinística representada pela função $f(t)$, e uma componente estocástica representada pela função $X(t)$. Considerando $P(t)$ o logaritmo natural do preço *spot* no tempo t , essa função será dada por:

$$P(t) = f(t) + X(t) \quad \text{onde } t = 1, 2, \dots, T \quad (2.20)$$

A componente determinística tem o objetivo de capturar os padrões já conhecidos para o comportamento dos preços no tempo. O modelo mais simples é uma componente constante no tempo, o que daria como resultado um comportamento exponencial para o modelo de preços. Foi feito um teste também colocando-se uma componente temporal no modelo, porém no momento da parametrização essa componente ficou com o seu coeficiente muito baixo e foi retirado sem grandes prejuízos.

Através da análise dos dados, chegou-se a conclusão que as variações mais significativas de preços se encontram nos dias de semana, quando comparado aos finais de semana e que os preços são mais altos durante os dias da semana e durante os meses mais próximos dos extremos quentes e frios do ano, principalmente devido à utilização residencial do ar condicionado e da calefação.

No modelo de Higgs e Worthington 2006 [10] a sazonalidade nos preços é incorporada através de variáveis *dummies*. Dessa forma, foram incluídos três conjuntos de variáveis *dummies*:

- Uma para capturar as variações entre dias úteis e dias não úteis;
- Uma segunda para capturar as variações sazonais no ano (mensais);

- Uma terceira para incluir a inserção de duas novas interconexões entre mercados: O The Queensland e o New South Wales Interconnector (QNI) que começou a operar em 18 de Fevereiro de 2001 e a interconexão Murraylink entre a South Australia e Victoria, iniciada em 2 Setembro de 2002.

Assim, a componente determinística $f(t)$ é dada por:

$$f(t) = \mu_0 + \beta_1 D_t + \sum_{i=2}^{12} \beta_i M_{it} + \gamma_1 INT \quad (2.21)$$

onde D_t são variáveis *dummies* para os dias da semana, tendo valor 1 para os feriados e dias de fim de semana e zero para os dias de semana. M_{it} , com $i = 2, \dots, 12$ são 11 variáveis *dummies*, uma correspondente a cada mês, janeiro é o mês de referência e está implicitamente relacionado ao intercepto μ_0 . A variável *dummy* de interconexão INT possui valor 1 depois de 18 de Fevereiro de 2001 para QLD e NSW e depois de 2 de Setembro de 2002 para SA e VIC e zero para outros intervalos de datas. As variáveis β_i para $i = 1, 2, 3, \dots, 12$, γ_1 e μ_0 são coeficientes a serem estimados.

O passo da componente estocástica é definido da seguinte forma:

$$dX(t) = dP(t) - df(t) \quad \text{onde } t = 1, 2, \dots, T \quad (2.22)$$

onde $dX(t) = X(t) - X(t-1)$.

Higgs e Worthington 2006 [10] avaliam a componente estocástica em três regimes dinâmicos, sendo eles (i) um modelo estocástico básico, (ii) um modelo de reversão à média e (iii) um modelo de mudança de regime.

O modelo básico estocástico é dado por: $dX(t) = \sigma_0 \xi(t)$. Nesse caso $\xi(t) \sim N(0, 1)$ e σ_0 é o desvio padrão dos retornos logarítmicos dos preços.

O modelo de reversão à média se justifica em sua utilização pelo comportamento observado de que os preços *spot* tendem a flutuar até encontrar o equilíbrio em um nível ao redor de um preço de longo prazo, no caso chamado de μ_0 , sendo que esse preço reflete o custo marginal de produção de eletricidade. Outro fator no preço é a taxa de reversão, que é um parâmetro que mensura a velocidade que o preço é forçado a retornar ao nível do equilíbrio de longo prazo depois que foi desviado desse equilíbrio. Esse fator age de forma negativa se o preço *spot* for maior que o preço de equilíbrio de longo prazo, e positivo caso o contrário. Dessa forma o modelo de reversão à média é dado por:

$$dX(t) = -\alpha_0 X(t-1) + \sigma_0 \xi(t) \quad \text{onde } t = 1, 2, \dots, T \quad (2.23)$$

neste caso $\xi(t) \sim N(0, 1)$ e α_0 é a taxa de reversão à média.

O terceiro modelo introduz um fator que leva os preços a ficarem em um patamar muito acima do situado em seu estado "normal", sendo esse fator controlado através de um processo de mudança de regime. No artigo de Higgs e Worthington 2006 [10] esse tipo de comportamento no preço se justifica com fatores do tipo interrupção por geradores ou altas e baixas temperaturas.

Nessas situações os preços irão retornar aos seus patamares de equilíbrio muito rapidamente, assim que o gerador voltar a funcionar ou as temperaturas voltarem a seus níveis considerados comuns, por exemplo. Com isso, saltos altistas são seguidos muito rapidamente por saltos baixistas, pois o fator de reversão à média força os preços a ficarem no equilíbrio de longo prazo.

O modelo final proposto está na linha do terceiro sugerido, sendo um modelo de mudança de regime de Markov que separa um período de reversão a média normal sem saltos, de um período com saltos. Dessa forma em todos os dias do ano o preço *spot* segue um dos seguintes regimes:

1. O regime 0 - um regime normal na dinâmica dos preços;
2. O regime +1 - um regime em que os preços repentinamente crescem, ou decrescem, em um pico de preço;
3. O regime -1 - quando o preço reverte ao normal após o pico ocorrer.

Assim a componente estocástica $dX(t)$ é modelada da seguinte forma, tendo em vista capturar a mudança de regime dos preços:

$$\begin{aligned}
 dX(t) &= -\alpha_0 X(t-1) + \sigma_0 \xi(t) & \text{no regime 0} & \text{onde } \xi(t) \sim N(0, 1) \\
 dX(t) &= \mu_1 + \sigma_1 \xi(t) & \text{no regime +1} & \text{onde } \xi(t) \sim N(0, 1) \\
 dX(t) &= -\alpha_{-1} X(t-1) + \sigma_{-1} \xi(t) & \text{no regime -1} & \text{onde } \xi(t) \sim N(0, 1)
 \end{aligned} \tag{2.24}$$

Para se mover de um regime a outro é utilizada uma matriz de transição de Markov que contém as probabilidades de se saltar de um regime de um determinado dia a outro no dia seguinte. É utilizada estimação por máxima verossimilhança para determinar os parâmetros do modelo dentro de cada regime.

Dessa forma, seja $\pi(i, j)$ a probabilidade do preço da energia elétrica passar do regime j no período t , para o regime i no período $t + 1$. As probabilidades são selecionadas da perspectiva de um dia para o dia seguinte, ou seja, estando hoje um dia normal, qual a probabilidade de ter-se um pico nos preços amanhã. Sendo $\pi(0, 0)$ a probabilidade de não ocorrência de picos, a probabilidade de ocorrerem picos será $\pi(+1, 0) = 1 - \pi(0, 0)$.

É definida uma probabilidade zero de se mover do regime normal para o regime de reversão do período de picos, ou seja, $\pi(-1; 0) = 0$. Estando no regime de picos +1 no dia t , o modelo proposto assume que os picos serão de períodos muito curtos, ou seja, estando hoje no período de regime de picos, amanhã se inicia o período de recuperação dos preços (reversão). Isso é representado por $\pi(-1; +1)$ igual a um, e $\pi(0, +1)$ e $\pi(+1, +1)$ iguais a zero. Estando em regime de reversão à média no dia t , o processo de preço é esperado que volte ao regime normal no dia seguinte, com isso $\pi(0, -1)$ é igual a um, e $\pi(+1, -1) = 0$ e $\pi(-1, -1) = 0$. Com isso, adotando-se as hipóteses descritas, a única probabilidade que é necessário estimar-se é a $\pi(0, 0)$, que foi feita através de observação empírica, ou seja, definiu-se um movimento diário mínimo que caracterizaria um salto e fez-se a estatística a partir da série histórica desse evento.

Os principais resultados estatísticos das simulações feitas por Higgs e Worthington 2006 [10] sobre o modelo logarítmico $\ln(P(t))$ com mudança de regime apresentam-se no Diagrama II abaixo, no qual são apresentados os resultados estatísticos para os submercados de *New South Wales* (NSW) e *South Australia* (SA).

Diagrama II: Resultados do modelo australiano				
	NSW		SA	
Parâmetro	Estimaco (DOL)	p-value	Estimaco (DOL)	p-value
β_0	3.33	0.01	3.7	0.01
β_1	-0.15	0.01	-0.27	0.01
β_2	0.01	sn	0.12	0.05
β_3	-0.20	0.05	-0.11	0.05
β_4	-0.19	0.01	-0.09	0.10
β_5	0.14	0.01	0.10	0.05
β_6	0.26	0.01	0.11	0.05
β_7	0.17	0.01	0.10	0.05
β_8	0.04	sn	0.01	sn
β_9	0.13	0.01	-0.11	0.05
β_{10}	-0.10	0.05	0.02	sn
β_{11}	-0.11	0.05	0.08	sn
β_{12}	-0.12	0.05	-0.05	sn
Desvio padro				0.0023

No diagrama, os valores "sn" so aqueles no apresentados na tabela do estudo e provavelmente se referem a valores com baixa significncia e que poderiam ser rejeitados do modelo sem prejuzos. A anlise que precisa ser feita pode ser exemplificada atravs do parmetro β_4 o qual oferece uma estimaco de -0.19 com um p-value associado de 0.01 no modelo de NSW que quando comparado ao modelo de SA que oferece uma estimaco de -0.09 com um p-value associado de 0.10. Nessa anlise comparativa o p-value de 0.01 de NSW possui uma maior significncia que o p-value 0.10 de SA, dado que esse p-value  o teste de Fisher-Snedecor (ou *F-statistics*) apresentado na seo 3.1.4 feito para a varivel β_4 individualmente. Nesse teste, o que se revela na prtica  que quanto menor o p-value, dado um nvel de significncia de referncia, menos se consegue aceitar a hiptese H_0 de rejeio do parmetro que est sendo testado.

Dessa forma, apesar de ambos os modelos apresentados no diagrama oferecem uma boa aderncia de sazonalidade, verifica-se que de forma geral o modelo de NSW apresentou uma melhor aderncia quando comparada ao modelo SA, pois apresentou menores p-values. Quando comparados os modelos apresentados no estudo, o que ofereceu a melhor aderncia foi realmente o de NSW, seguido ao modelo de VIC, QLD e SA.

Em suas anlises Higgs e Worthington 2006 [10] chegam s seguintes concluses: Primeiro, que apesar da frequncia dos saltos serem relativamente altas, a maior  de que 9.4% dos dias

apresentaria um salto em VIC, o retorno ao regime normal é bastante veloz, mostrando que esses picos de preço possuem tempo de vida muito curto. A reversão à média nesses períodos de stress são mais velozes do que em períodos normais, indicado por um parâmetro de reversão à média mais representativo para esses modelos. Outra conclusão é que os picos de preços influenciam muito em sua volatilidade, mostrando que em dias de regime normal a volatilidade é até bastante baixa, com o preço variando ao redor do custo marginal de operação. A terceira conclusão é que o tamanho dos saltos podem variar muito dependendo do mercado estudado.

Na parte determinística do modelo, se concluiu que a influência dos parâmetros *dummies* de dias da semana e dos meses variam muito de mercado para mercado, e muitos resultados divergem de um mercado para o outro. Uma verificação interessante é que o fator de interconexão, que influiu positivamente nos preços de QLD e SA, apresentou efeitos negativos em NSW.

O artigo sugere que uma das limitações do modelo é que ele não permite sucessivos saltos, e isso pode ser melhorado. Outra limitação é que, ao se utilizar os preços diários o período mais curto para um salto que se consegue impor é um dia, porém em observações empíricas se observa que existem muitos casos que os picos chegam a durar até meia hora, ou seja, perde-se muita informação utilizando-se preços diários. Pode ser valiosa a investigação da dinâmica dos saltos no decorrer dos dias, pois os preços possuem uma dinâmica regular entre as próprias horas do dia devido aos padrões de consumo (em semelhança aos "patamares" característicos do mercado brasileiro), o que implicaria em uma modelagem adicional para se capturar esses padrões comportamentais regulares diários nos preços.

Escolheu-se o estudo do mercado australiano na tentativa de comparação com o mercado brasileiro devido às dimensões continentais dos dois países, provocando restrições de transmissão equiparáveis e os regimes sazonais de temperatura serem semelhantes. Uma grande diferença porém é a forma de contabilização dos preços, onde na Austrália existe um mercado de energia com registro de negociações no decorrer do dia, onde consegue-se de fato definir um preço *spot* de energia a preços de mercado, diferentemente do mercado brasileiro que é de definição de preços em base semanal, calculados através de algoritmos. Uma outra diferença importante entre os mercados é a base hidrotérmica, com forte viés hidráulico do mercado brasileiro, e o termelétrico da australiano. Nota-se que a base da energia possui uma influência muito forte no regime dos preços, com relação aos seus valores médios e o regime de volatilidade.

2.4 Mercado dos Estados Unidos

O mercado norte-americano, diferentemente do mercado de outros países que sofreram uma desregulamentação do seu mercado de energia elétrica, possui uma maior segregação dos mercados regionais, sem a atividade de um operador em escala nacional, e dessa forma com uma característica operacional e regulatória independente para cada estado. Antes da desregulamentação, que teve seu início no final dos anos noventa, o mercado já possuía uma forte operação privada em seu sistema, o que se por um lado era um fator positivo, pois não incluiria a privatização de

empresas públicas, por outro havia a forte preocupação que a concorrência gerada pela abertura abaixassem as tarifas de tal forma que tornasse inviável a operação, principalmente devido a custos incorridos e não amortizados (dívidas) que as empresas haviam em seus balanços e que anteriormente a se pensar na desregulamentação tinham perspectivas de pagamento em um mercado monopolizado e mais previsível no longo prazo.

Essas perspectivas de redução de preços de tarifas provinham de experiências de países como Inglaterra, que com o aumento competitivo potencializado pelos avanços tecnológicos de controle do sistema elétrico e de comercialização de energia viram como consequência da desregulamentação um aumento do equilíbrio competitivo entre os agentes do mercado, com benefícios para os consumidores e para o mercado em geral.

A presença de bolsas de energia regionalizadas tem como característica a forte presença em geral nos estados mais liberalizados. A NYMEX, cobrindo principalmente a região de Nova Iorque e estados vizinhos, começou a operar derivativos de eletricidade em 1996. Isso foi seguido pela liberação para negociação no Chicago Board of Trade (CBOT) e a Minneapolis Grain Exchange (MGE). Logo depois de sua implantação essas bolsas viram um grande aumento do volume de negócios, com uma variedade de tipos de futuros de energia, opções e derivativos criados também crescente. Porém, após a crise energética da Califórnia [16] em 2002 houve um declínio de negociações, culminando com a suspensão devido ao baixo volume de negócios.

Assim que implantados, os novos formatos do mercado americano se baseavam no poder do consumidor conseguir escolher entre as diferentes empresas geradoras de energia elétrica. Assim como se observa no Brasil, as geradoras não são mais reguladas, porém a transmissão e a distribuição ainda possuem suas tarifas e negociações reguladas. O principal racional por trás da regulamentação da transmissão é assegurar que todos os geradores de energia possam ter acesso ao sistema independentemente da origem da produção, possibilitando dessa forma aos produtores concentrarem seus esforços na redução dos custos de produção e não em fatores alheios que possam aumentar os seus custos produtivos, como o acesso à rede de transmissão. Com relação à distribuição regulamentada, aí está uma das principais diferenças com relação ao processo de desregulamentação que se passou na Inglaterra, o qual permitiu que as tarifas do consumidor final flutuassem e fossem definidos por fatores mercadológicos de competição.

A partir de 2003 a Comissão Reguladora Federal de Energia (FERC) permitiu aos estados decidirem pelo modelo a ser adotado em suas jurisdições pela forma de operação. Dessa forma deixou aos operadores independentes de sistemas (OISs) decidirem pela melhor abordagem de administração do sistema a critério próprio. Atualmente os mercados mais desenvolvidos estão na PJM (Pennsylvania - New Jersey - Maryland Interconnection), na Califórnia, na Nova Inglaterra e em Nova Iorque. Esses modelos são bastante diversos, sem um padrão definido de forma de negociação comum a todos os estados.

De forma geral, a implantação dos mercados de longo prazo a partir de derivativos e contratos futuros esbarra na necessidade de definição dos preços de curto prazo que consigam refletir um preço *spot* unânime e reconhecido por todos os agentes de mercado. Se os dados da comercialização de energia são de baixa qualidade com relação aos preços e volumes operados devido

à falta de transparência no mercado, o mercado desses produtos financeiros de energia tende a ruir. Segundo Mayo 2009 [16] no mercado norte americano, apenas dez dos maiores centros de comercialização de energia possuem grandes e líquidos mercados de curto prazo com registro correto e divulgado de preços de liquidação para mercados D+1 e de tempo real. Essa situação permite aos OIS dessas regiões o desenvolvimento de índices de preços credíveis, permitindo dessa forma a operação eficiente de mercados financeiros de negociação de energia.

A seguir serão descritas as principais jurisdições de energia dos Estados Unidos:

A Pennsylvania-New Jersey-Maryland ISO (JPM-ISO) foi implantada em 1998, serve 13 estados e o Distrito de Colúmbia. Opera nos mercados de entrega de energia em D+1 e em tempo real, contratos futuros mensais de liquidação financeira negociados no NYMEX, opções sobre posições futuras, mercados de capacidade e serviços ancilares. Mercado bastante transparente, com disponibilização de informações na Internet, sendo que em 2007 possuía 500 participantes do mercado para uma demanda total anual de 763 TWh (Mayo 2009 [16]).

O New York ISO - NYISO foi implantado em 1999, servindo o estado de Nova Iorque. Opera nos mercados de D+1 e em tempo real, contratos futuros liquidados financeiramente pela NYMEX, contratos de congestionamento de transmissão, mercado de capacidade e mercados de serviços ancilares. O pregão de bolsa é eletrônico, com uma relação bem dividida entre negociação em bolsa pelo NYMEX (50%) e a negociação bilateral em mercado de balcão (50%). Para o ano de 2007 apresentou um volume de negociações de US\$ 9,5 bi, com 340 participantes de mercado e uma demanda total de 167 TWh (Mayo 2009 [16]).

O ISO New Englad - ISO NE foi implantado em 1997 e possui uma abrangência em seis estados do nordeste estadunidense. Opera nos mercados de D+1 e em tempo real, contratos futuros liquidados financeiramente pela NYMEX, contratos de congestionamento de transmissão, mercado de capacidade e mercados de serviços ancilares. Para o ano de 2007 apresentou um volume de negociações de US\$ 10 bi, com mais de 300 participantes de mercado e uma demanda total de 134 TWh (Mayo 2009 [16]).

A California ISO - CAISO foi implantada em 1998 e abrange o estado da Califórnia e os estados vizinhos. Entre os produtos operados estão a entrega de energia em tempo real, contratos de congestionamento e transmissão em base regional e mercado de serviços ancilares. Em 2007 operou um volume financeiro de US\$ 2,5 bi para uma demanda total de 242 TWh (Mayo 2009 [16]).

O estudo do mercado e do histórico norte americano é emblemático no sentido verificação de como um processo de desregulamentação mal planejado pode trazer consequências negativas para a população e para o mercado em geral. Seu processo de desregulamentação mal conduzido, iniciado no estado da Califórnia, atrelado a práticas comerciais criminosas pelos agentes de geração em sua conduta de operação do sistema e manipulações contábeis, acarretou em uma situação de total descontrole na Califórnia, culminando com uma ambiente de calamidade econômica provocando *black-outs* e falências de empresas distribuidoras que não conseguiam repassar aos seus clientes finais os altos preços que pagavam pela energia. Maiores informações sobre as causas e os impactos da crise energética da Califórnia são disponibilizadas por Taylor

e VanDoren 2002 [22].

Capítulo 3

Método Empregado

Este capítulo apresenta o método utilizado para avaliação dos diferentes modelos a serem estudados. O método é uma abordagem estatística que utiliza por base a análise de regressão linear, que a partir dos métodos dos mínimos quadrados e testes de razão de verossimilhança, como serão apresentados na seção a seguir, oferecem alguns parâmetros que auxiliam a verificação da qualidade desses modelos, evidenciando a aderência das respostas dos modelos utilizados aos números observados na realidade. A partir desse método, após uma categorização das variáveis de entrada dos modelos, é possível avaliar quais entradas realmente contribuem para a resposta quando comparada com os números reais, refinando dessa forma as repostas apresentadas e consequentemente os modelos propriamente ditos.

O objetivo é oferecer um modelo de regressão linear que a priori explique qual é o comportamento do preço *spot* de energia e avaliar quais dessas entradas realmente contribuem para oferecer um modelo aderente ao preço da energia que é observado na realidade. As entradas dos modelos, aqui chamados de preditores, serão categorizados em fatores sazonais, fatores de suprimento e fatores de demanda e a avaliação dos modelos propriamente ditos será feita no capítulo 4.

3.1 Introdução teórica ao método empregado

No presente estudo serão utilizados modelos de regressão linear como base para os procedimentos de análise e avaliação. Esses modelos estarão na seguinte forma:

$$PLD = \beta_0 + \beta_1 M_1 + \beta_2 M_2 + \dots + \beta_r M_r + \varepsilon \quad (3.1)$$

Neste modelo, PLD é a variável resposta observada e já conhecida, as variáveis M_1, M_2, \dots, M_r são os r preditores do modelo também já conhecidos, com $r = 1, \dots, n$, e os parâmetros $\beta_1, \beta_2, \dots, \beta_r$ são as variáveis de regressão desconhecidas, incluindo o intercepto β_0 . Estando os parâmetros desconhecidos de regressão na forma da equação 3.1, a equação será linear e por esse motivo utiliza-se o nome de *modelo de regressão linear*.

Para cada observação histórica do variável resposta PLD , o modelo terá os valores de M_i associada a essa resposta, o que torna o modelo da seguinte forma, com $i = 1, \dots, n$, sendo n

são o número de observações históricas:

$$PLD_i = \beta_0 + \beta_1 M_{1i} + \beta_2 M_{2i} + \dots + \beta_r M_{ri} + \varepsilon_i \quad (3.2)$$

Seguindo o procedimento de Johnson e Wichern 2002 em [14], assume-se as seguintes propriedades para o vetor do termo de erro ε , vetor esse de dimensões $(n \times 1)$:

1. $E(\varepsilon) = 0$
2. $Cov(\varepsilon) = E(\varepsilon\varepsilon') = \sigma^2 I$

Como será observado a seguir, será necessária também a hipótese de distribuição normal para o termo de erro, para que se torne possível a construção de intervalos de confiança de valores e a avaliação através de testes de hipóteses para as variáveis β_i .

3.1.1 Estimação dos parâmetro de regressão

Na análise por regressão linear, dada uma base de dados previamente disponibilizada, o objetivo inicial é estimar os valores dos coeficientes de regressão β_i de 3.1 e o desvio padrão correspondente ao termo de erro σ . Neste trabalho para definir um valor estimado para os coeficientes β_i , é utilizado o método dos mínimos quadrados. Através desse método são definidas estimações para os coeficientes do vetor β tal que minimizem a diferença quadrática entre o valor verificado das observações PLD e o seu valor estimado \widehat{PLD} . Esses estimadores, quando definidos por essa técnica, são chamados de estimadores de mínimos quadrados e, neste trabalho, serão identificados por $\hat{\beta}$.

Os vetores de preditores observados M_1, M_2, \dots, M_r , podem ser acoplados em uma matriz única, chamada Matriz Desenho:

$$M = \begin{bmatrix} 1 & M_{11} & M_{12} & \dots & M_{1r} \\ 1 & M_{21} & M_{22} & \dots & M_{2r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & M_{n1} & M_{n2} & \dots & M_{nr} \end{bmatrix} \quad (3.3)$$

onde r é o número de preditores, e n é o número de observações históricas feitas.

Observa-se que nesta matriz, a primeira coluna de "1"s corresponde ao preditor acoplado ao intercepto β_0 dos parâmetros de regressão.

No caso de possuímos uma matriz desenho M de rank completo e tal que $r + 1 \leq n$, seguindo-se Johnson e Wichern 2002 [14] pode-se verificar que a estimação de mínimos quadrados do vetor β é dado por:

$$\hat{\beta} = (M' \cdot M)^{-1} \cdot M' \cdot PLD \quad (3.4)$$

Neste caso, sendo PLD o vetor das variáveis respostas observadas, verifica-se que o estimador para o vetor ε , também chamado de *resíduo*, será dado por:

$$\hat{\varepsilon} = PLD - \widehat{PLD} = PLD - M\hat{\beta} \quad (3.5)$$

isto é, a diferença entre o histórico observado PLD e o estimado através do modelo de regressão linear.

3.1.2 Abordagem geométrica para o método

Esta seção procura evidenciar qual o conceito por trás da análise por regressão linear, utilizando para isso uma abordagem geométrica dos resultados. O que ocorre é que o valor estimado de um vetor variável resposta \widehat{PLD} específico nada mais é do que uma projeção ortogonal da variável resposta original PLD (em sua forma vetorial) no sub-espaço formado pelos vetores de preditores M_i . Isso se torna claro na seguinte forma:

$$\widehat{PLD} = \hat{\beta}_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \hat{\beta}_1 \begin{bmatrix} M_{11} \\ M_{21} \\ \vdots \\ M_{n1} \end{bmatrix} + \cdots + \hat{\beta}_r \begin{bmatrix} M_{1r} \\ M_{2r} \\ \vdots \\ M_{nr} \end{bmatrix} \quad (3.6)$$

Observa-se nessa equação que o valor estimado \widehat{PLD} é de fato uma combinação linear dos vetores M_i . O método dos mínimos quadrados oferece os parâmetros estimados $\hat{\beta}$ s que proporcionam exatamente a mínima distância quadrática entre o vetor de observações PLD e o plano formado pelos vetores de preditores M_i , ou seja, a sua projeção ortogonal \widehat{PLD} neste hiperplano. Como consequência, temos que o vetor resíduo $\hat{\varepsilon} = PLD - \widehat{PLD}$ é um vetor que é perpendicular a esse hiperplano e sua norma é na verdade uma métrica da qualidade dessa estimação quando comparada ao seu valor real PLD .

Como será observado na próxima seção, uma das formas de medição da qualidade da regressão é justamente através da soma quadrática dos vetores resíduos (norma) de cada uma das observações do PLD , quando comparada aos valores estimados \widehat{PLD} . Quanto menor for essa soma, mais próximo os pontos observados PLD estão do hiperplano formado pelos vetores de preditores M_i e dessa forma melhor é a qualidade da regressão.

3.1.3 Qualidade da regressão

Uma forma usualmente utilizada para medição da qualidade da regressão é através do *coeficiente de determinação*, abreviado pela sigla R^2 . Para se estimar esse coeficiente é necessário um resultado precedente, seguindo-se a abordagem de Johnson e Wichern (2002) [14], e valendo-se da equação 3.4:

$$\begin{aligned} \widehat{PLD}'\hat{\varepsilon} &= \hat{\beta}'M'\hat{\varepsilon} = \hat{\beta}'M'(PLD - \widehat{PLD}) = \hat{\beta}'M'(PLD - M\hat{\beta}) \\ &= \hat{\beta}'M'(I - M((M'M)^{-1}M'))PLD = \hat{\beta}' * (M' - M')PLD = 0 \end{aligned}$$

A partir disso, extrai-se algumas propriedades da soma dos quadrados da variável reposita:

$$PLD'PLD = (\widehat{PLD} + \hat{\varepsilon})'(\widehat{PLD} + \hat{\varepsilon}) = \widehat{PLD}'\widehat{PLD} + \hat{\varepsilon}'\hat{\varepsilon} \quad (3.7)$$

A partir desta equação, e sabendo-se que as médias $\overline{\widehat{PLD}} = \overline{PLD}$ obtêm-se, com os seguintes passos, o R^2 :

$$\begin{aligned} PLD'PLD - n\overline{PLD}^2 &= \widehat{PLD}'\widehat{PLD} - n\overline{\widehat{PLD}}^2 + \hat{\varepsilon}'\hat{\varepsilon} \\ \sum_{i=1}^n (PLD_i - \overline{PLD})^2 &= \sum_{i=1}^n (\widehat{PLD}_i - \overline{PLD})^2 + \sum_{i=1}^n \hat{\varepsilon}_i^2 \end{aligned} \quad (3.8)$$

Dividindo-se todas as equações por $\sum_{i=1}^n (PLD_i - \overline{PLD})^2$, obtêm-se:

$$\begin{aligned} \frac{\sum_{i=1}^n (\widehat{PLD}_i - \overline{PLD})^2}{\sum_{i=1}^n (PLD_i - \overline{PLD})^2} &= \frac{\sum_{i=1}^n (PLD_i - \overline{PLD})^2}{\sum_{i=1}^n (PLD_i - \overline{PLD})^2} - \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{\sum_{i=1}^n (PLD_i - \overline{PLD})^2} \\ R^2 &= 1 - \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{\sum_{i=1}^n (PLD_i - \overline{PLD})^2} \end{aligned} \quad (3.9)$$

onde \overline{PLD} é a média de todos os valores do vetor resposta.

O coeficiente R^2 é indicativo da qualidade da regressão e a sua utilização será explicitada na seção 3.2.1.

3.1.4 Teste de hipótese sobre os parâmetros do modelo

A análise de inferência estatística sobre os parâmetros do modelo parte da hipótese de que ε possui uma distribuição normal. A verificação dessa hipótese é conduzida, na prática, através da investigação da distribuição da amostra da base de dados que está sendo analisada. Os métodos utilizados no presente trabalho para verificação da normalidade serão a construção de gráficos do termo de erro contra a variável resposta, o gráfico Q-Q e o teste de normalidade de Shapiro-Wilk, os quais serão explicados na seção 3.2.2.

Assumindo, dessa forma, que ε possui uma distribuição normal, Johnson e Wichern (2002) [14] demonstram que o estimador de máxima verossimilhança de β é dado por $\hat{\beta} = (M' \cdot M)^{-1} \cdot M' \cdot PLD$, ou seja, igual ao estimador de mínimos quadrados e sua distribuição será dada por:

$$N_{r+1}(\beta, \sigma^2(M'M)^{-1}) \quad (3.10)$$

Pode-se verificar também que o estimador de máxima verossimilhança de σ é dado por :

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} = \frac{(PLD - M\hat{\beta})'(PLD - M\hat{\beta})}{n} \quad (3.11)$$

A condição necessária de distribuição normal para o parâmetro de regressão β é utilizada para se fazer inferências para averiguação da relevância dos preditores M_i no modelo de regressão. Uma maneira de se verificar se um preditor influencia ou não a variável resposta é através de testes de hipótese. Nessa abordagem divide-se os parâmetros de regressão em dois grupos, um formado pelos parâmetros que realmente influenciam na variável resposta e outro

grupo dos que não influenciam. O modelo linear de regressão pode então ser escrito da seguinte forma:

$$PLD = M\beta + \varepsilon = M_1\beta_1 + M_2\beta_2 + \varepsilon$$

onde M_2 corresponde ao grupo de preditores que desconfia-se não influenciam na variável resposta PLD , associados aos parâmetros β_2 de regressão.

Define-se a matriz \mathbf{C} como uma matriz de dimensões $(r - q) \times (r + 1)$, onde r é o número de preditores do modelo e $(r - q)$ é o número de parâmetros de regressão testados, da seguinte forma:

$$\mathbf{C} = \begin{bmatrix} 0_{(r-q) \times (q+1)} & \vdots & I_{(r-q) \times (r-q)} \end{bmatrix}$$

onde I é a matriz identidade.

Utilizando-se a matriz C , é verificada a seguinte hipótese:

$$H_0 : \mathbf{C}\beta = 0$$

A partir da equação 3.10 é possível derivar-se a distribuição de $\mathbf{C}\hat{\beta}$:

$$\mathbf{C}\hat{\beta} \sim N_{r-q}(C\beta, \sigma^2 C(M'M)^{-1}C')$$

Dessa forma, seguindo-se Johnson e Wichern 2002 [14], o teste de hipótese H_0 pode ser rejeitado a um nível α de significância se a elipsoide confiança da distribuição do vetor $C\beta$ não contém o ponto $\mathbf{0}$. Isso significa que não pode ser atribuído o valor "0" a um parâmetro de regressão " β_i " a um nível de confiança α , o que na prática representa que o preditor correspondente ao " β_i " é significativo para a variável resposta PLD e não pode ser eliminado do modelo.

O teste de hipótese H_0 pode ser rejeitado se for verificado o seguinte teste de razão de verossimilhança:

$$\frac{(\mathbf{C}\hat{\beta})'(\mathbf{C}(M'M)^{-1}\mathbf{C}')^{-1}(\mathbf{C}\hat{\beta})}{s^2} > (r - q)\mathcal{F}_{p,n-r-1}(\alpha) \quad (3.12)$$

onde $s^2 = \hat{\varepsilon}'\hat{\varepsilon}/(n - (r + 1))$ é uma estimação do σ^2 e $\mathcal{F}_{p,n-r-1}(\alpha)$ é a distribuição de Fisher-Snedecor (ou F -statistics) com p e $n - r - 1$ graus de liberdade no (100α) ésimo percentil superior.

Outra maneira de se analisar o mesmo problema é através da verificação do quanto os resíduos, em sua forma quadrática, do modelo são incrementados quando é feita a comparação entre os modelos sem os preditores investigados e o modelo completo. Se a diferença é alta (dado um nível de confiança α) pode-se rejeitar H_0 .

Para se fazer essa análise, define-se a soma dos quadrados dos resíduos de M , ou $SS_{res}(M)$, da seguinte forma:

$$SS_{res}(M) = \hat{\varepsilon}'\hat{\varepsilon} = (PLD - M\hat{\beta})'(PLD - M\hat{\beta})$$

Considerando-se M a matriz de preditores originais do modelo, M_1 a matriz com o subgrupo de preditores que influenciam a variável resposta e M_2 o subgrupo de preditores que desconfia-se não influenciarem a variável resposta, o teste de hipótese $H_0 : \beta_2 = 0$ e a razão de verossimilhança associada a ela, da mesma forma que 3.12, pode ser escrita da seguinte forma:

$$\frac{(SS_{res}(M_1) - SS_{res}(M))}{s^2} > (r - q)\mathcal{F}_{p,n-r-1}(\alpha) \quad (3.13)$$

No caso de a diferença $(SS_{res}(M_1) - SS_{res}(M))$ ser significativa, a um nível α de confiança, pode-se concluir que existe perda qualidade no modelo ao se retirar os preditores M_2 . Na prática isso significa que, com a retirada desses preditores, houve um crescimento significativo do resíduo $\hat{\varepsilon}$, ou seja, o modelo está menos aderente à variável resposta que o modelo completo M e dessa forma pode-se rejeitar a hipótese H_0 .

3.2 Método empregado para avaliação do preço *spot*

Nesta seção será apresentado o método aplicado para avaliação dos modelos. Pela abordagem descrita a seguir, será introduzida uma ferramenta que permite a construção e a comparação de modelos para o preço *spot*. Os modelos são diferenciados entre si pelas variáveis que foram utilizadas como entradas no sistema, representado pela matriz desenho de preditores M . Dessa forma, dado um conjunto de variáveis que supostamente influenciam os preços *spot* PLD , pode-se verificar quais são as variáveis que influenciam mais fortemente a variável resposta.

Como será verificado no capítulo 4, a análise se iniciará a partir da medição de algumas estatísticas básicas sobre a base de dados dos preços, como média, valores máximos e mínimos históricos, medidas de desvio padrão e volatilidade, medidas de curtose e assimetria, na busca de capturar informações gerais sobre os preços *spot* de energia elétrica. Após essas medições, se inicia a aplicação do método propriamente dito com a definição de um modelo de regressão linear para cada um dos modelos, como descrito a seguir.

3.2.1 Procedimento de estimação por mínimos quadrados

O modelo de regressão linear básico é dado pela equação 3.1, em sua forma matricial:

$$PLD = M \cdot \beta + \varepsilon \quad (3.14)$$

onde PLD é um n -vetor de observações da variável resposta, M é uma $n \times (r + 1)$ matriz desenho (n é o número de observações e r é o número de preditores somado ao intercepto). β é um vetor de $(r + 1)$ coeficientes de regressão (intercepto incluído) e ε é um n -vetor dos termos de erro. Os vetores β e ε serão ambos estimados pelo método dos mínimos quadrados.

Nos modelos testados, PLD é um vetor que assume o valor do preço *spot* da energia elétrica ao longo do tempo em base mensal. M é uma matriz que possui em suas colunas os preditores, desenhada de acordo com as variáveis que serão testadas em cada verificação de modelo, podendo assumir, por exemplo, variáveis *dummy* correspondente aos meses em um modelo sazonal, volumes de fornecimento de energia, a energia armazenada nos reservatórios e a demanda. Para fins de melhor entendimento, os modelos testados no capítulo 4 mostrarão a matriz desenho M em sua forma aberta, ou seja, com suas colunas associadas ao coeficiente de regressão β correspondente no modelo.

Trabalhando-se com uma matriz desenho de *rank* completo, a estimação de mínimos quadrados, através da equação 3.4, será dada por $\hat{\beta} = (M'M)^{-1}M'PLD$.

Sejam os valores estimados do preço *spot* PLD dados por 3.6, $\widehat{PLD} = M\hat{\beta}$, e a estimação de ε , ou *resíduos*, dados por $\hat{\varepsilon} = PLD - \widehat{PLD}$. Como apresentado através da equação 3.9, a qualidade da regressão nos modelos apresentados será dada através da medição do *coeficiente de determinação* R^2 . Pelo fator R^2 é possível se medir o quanto da variação do preço *spot* da energia PLD pode ser explicada através dos preditores da matriz desenho M .

Na prática, quanto mais próximo o valor de R^2 estiver de "1", mais aderente é a estimativa. Se o R^2 estiver próximo de "0", os preditores possuem baixa influência sobre o PLD , e a conclusão é que eles não contribuem significativamente para a explicação do preço *spot*.

Na observação da fórmula de R^2 na equação 3.9 verifica-se que o denominador $\sum_{i=1}^n (PLD_i - \overline{PLD})^2$ possui a informação da variância do preço *spot*, enquanto o numerador $\sum_{i=1}^n \hat{\varepsilon}_i^2$ nos dá a informação do resíduo quadrático do modelo de preço. Desse último naturalmente espera-se que seja menor que a variância do próprio preço, pois através dessa divisão se está comparando o resíduo de um modelo que procura aderir da melhor forma possível ao preço *spot*, com a variância do próprio preço. Se essa divisão for próxima a "1" significa que não existe necessidade de um modelo pois a própria média dos preços já oferece uma resposta semelhante à resposta do modelo regressivo.

3.2.2 Normalidade do termo de erro

Como já introduzido na seção 3.1.4, o modelo assume a hipótese de distribuição probabilística do termo de erro como $\varepsilon \sim N_n(0, \sigma^2 I)$, onde n é o número de observações e I é a matriz identidade. A assunção da distribuição normal para o termo de erro é uma condição necessária para a condução dos testes de hipóteses que verificam a representatividade dos preditores nos modelos e pode ser avaliada através da investigação da base de dados. No presente estudo, o primeiro teste para verificação da hipótese de normalidade é a construção do gráfico do termo de erro contra o fator resposta, com o objetivo de se verificar o padrão visual de distribuição e a presença de *outliers* mais evidentes.

Em uma distribuição normal, o esperado para o gráfico $\hat{\varepsilon} \times \widehat{PLD}$ é ter um padrão de distribuição razoavelmente uniforme para os resíduos $\hat{\varepsilon}$ com relação ao eixo das abscissas dos valores de resposta estimados \widehat{PLD} , o que significa que conforme a variação do valor estimado da energia *spot* \widehat{PLD} o padrão de erro medido $\hat{\varepsilon}$ não tem padrões ou tendências. Um exemplo

de tendência é a verificação de que para valores mais altos de \widehat{PLD} (abscissas) o *range* de variação dos resíduos $\hat{\varepsilon}$ (ordenadas) ficam maiores, o que significa na prática que o modelo fica mais impreciso para valores mais altos de \widehat{PLD} . O que se observa é que quanto mais desigual a distribuição do gráfico, menos a distribuição do termo de erro ε se aproxima do que seria esperado para uma distribuição normal.

Na figura 3.1 abaixo são verificados dois casos:

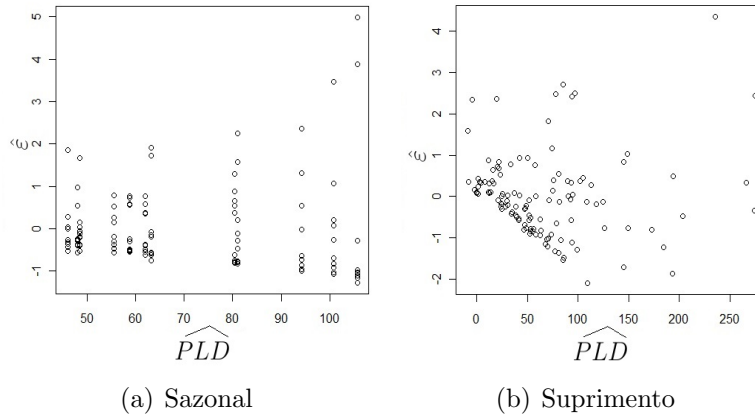


Figura 3.1: Resíduos x PLD de resposta

O gráfico (a) Sazonal oferece a distribuição para o modelo sazonal, no qual se verifica que para menores valores de \widehat{PLD} a distribuição dos resíduos está em um *range* menor do que para valores maiores de \widehat{PLD} . Isso é uma característica que evidencia uma tendência de distribuição desigual dos resíduos $\hat{\varepsilon}$, pois quanto maior o \widehat{PLD} menos aderente ele é à resposta original PLD , lembrando que $\hat{\varepsilon} = PLD - \widehat{PLD}$.

Para o gráfico (b) Suprimento, que verifica o modelo de suprimentos, apesar de se notar a existência de um afunilamento para valores mais baixos de \widehat{PLD} com relação aos resíduos menores que zero, não se nota um claro padrão de distribuição diferenciado do range quando comparado ao gráfico anterior. Essa distribuição afunilada para valores mais baixos ou próximos de zero de \widehat{PLD} é relativamente esperada pois, sendo PLD um preço, ele nunca será menor do que zero, logo existe uma tendência do valor estimado \widehat{PLD} ser inferior ao valor medido PLD , ou seja, o resíduo $\hat{\varepsilon} = PLD - \widehat{PLD}$ é preferencialmente positivo nessa região do gráfico. Os modelos apresentados nos gráficos 3.1 serão estudados no capítulo 4.

Nos casos que a variância do termo de erro não constante é mais acentuada pode ser necessária a utilização de outras abordagens de métodos dos mínimos quadrados ou transformações para equalização do termo de erro. O ideal é sempre evitar que o termo de erro dependa da magnitude da variável resposta.

Nesse trabalho, outro teste de normalidade que é conduzido é a verificação pelo gráfico Q-Q, que verifica o quanto a distribuição do resíduo $\hat{\varepsilon}$ se distancia do que seria esperado para uma distribuição normal, através da verificação de como é a distribuição dos seus quantis quando comparado aos quantis de uma distribuição normal. Nesse gráfico, no "eixo y" existe a distribuição amostral do resíduo, transformado para uma distribuição *standard* e em ordem crescente, e no "eixo x" está o quantil esperado e ordenado para uma distribuição *standard*

normal. Caso a distribuição dos pontos fiquem muito próximos a reta de suporte, a hipótese de que se está trabalhando com uma distribuição normal para o termo de erro é mais palpável. Essa hipótese pode ser duvidosa quando a distribuição desvia muito da reta de suporte.

Na figura 3.2 abaixo são mostrados dois exemplos de gráficos Q-Q:

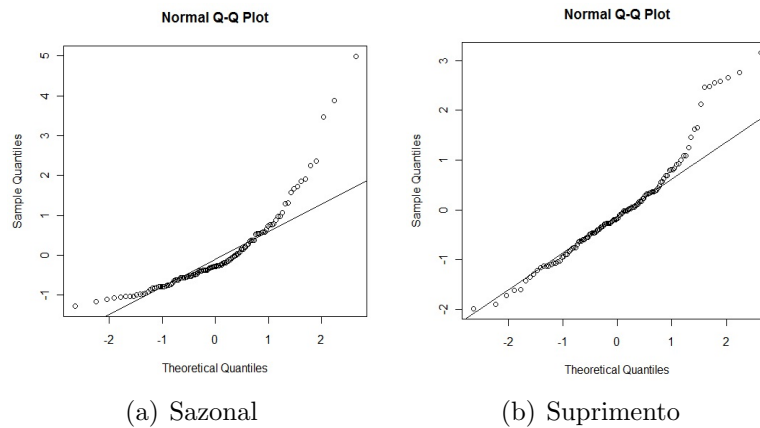


Figura 3.2: Gráfico Q-Q

O gráfico Q-Q (a) representa a distribuição com relação a parâmetros sazonais. Nesse gráfico se verifica que os pontos se distribuem de forma não conforme o que seria esperado para uma distribuição normal, dado o desvio dos pontos da reta suporte. O que se conclui com um gráfico exemplificativo desse tipo é que a distribuição dos resíduos está muito diferente do que seria esperado para uma distribuição normal, com tendências de assimetria que tornam uma hipótese de normalidade, nesse caso, forçada.

O gráfico Q-Q (b) compara a distribuição dos resíduos do modelo de suprimentos com o que seria esperado para uma distribuição normal. Nesse se verifica que os pontos em grande parte se apoiam sobre a reta suporte, o que representa uma distribuição que se aproxima de uma distribuição normal. O que se nota também é que para quantis mais altos a distribuição dos resíduos se distancia da reta suporte, característica essa que está presente normalmente em amostras com *outliers* ou curtose altas, o que de fato é um problema constante em estudos de precificação de energia elétrica. Esses gráficos serão melhor explorados no capítulo 4.

O terceiro teste utilizado no presente estudo para a verificação da normalidade é o teste de normalidade de Shapiro-Wilk. Neste teste segue-se o mesmo princípio do gráfico Q-Q, no qual verifica-se o quanto o resíduo se aproxima de uma distribuição normal, porém neste caso, é feito um teste de hipótese no qual verifica-se se a amostra se desvia muito em módulo de uma distribuição normal. No caso do resultado do *p-value* do teste ser menor do que o nível de significância escolhido (nível alpha), a hipótese nula pode ser rejeitada, oferecendo boas bases para afirmar que a distribuição do termo de erro não é normal. Para mais informações sobre os testes de normalidade aplicados é oferecida a referência de S. S. Shapiro, M. B. Wilk e H. J. Chen (1968) [20].

3.2.3 Verificação dos preditores significativos

Para cada modelo verificado, a matriz desenho M é um conjunto histórico de preditores. Como exemplo, como será verificado no capítulo 4 onde são apresentados os modelos em termos quantitativos, a matriz desenho M para o modelo de suprimento de energia apresenta, além de outros preditores, uma coluna com o volume histórico de água nos reservatórios, uma coluna para a quantidade de energia gerada pelas usinas nucleares e uma coluna com o volume de energia que entrou no sistema a partir da ENA.

Na análise de regressão pode ocorrer que um fator específico, ou mesmo um subgrupo de fatores dentro do conjunto preditores, estar prejudicando a aderência do modelo à variável resposta PLD . Para tornar possível a verificação da influência de um preditor específico na resposta, nesse estudo, foi utilizado o procedimento descrito na seção 3.1.4 o *Teste da Razão de Verossimilhança*. Neste teste, a hipótese nula verifica se um parâmetro de regressão (ou um conjunto de parâmetros), representado pelo vetor β , pode ser considerado "0", e caso hipótese se confirme, a conclusão é que o preditor associado ao β testado não influencia particularmente na variável resposta "PLD". Dessa forma, a adequação de um parâmetro de regressão, ou um subconjunto de parâmetros, pode ser avaliado através da análise da hipótese estatística:

$$H_0 : \beta_{investigado} = \mathbf{C}\beta = 0 \quad \text{ou} \quad H_1 : \mathbf{C}\beta \neq 0$$

onde \mathbf{C} é uma matriz $p \times (r + 1)$, sendo p o número de análises que serão verificadas simultaneamente e as colunas representam os parâmetros β que serão verificados (incluindo o intercepto). Nessa matriz, é colocado o valor "1" na posição do parâmetro de regressão β que está sendo verificado, e "0" nas outras posições. Seguindo os procedimentos da equação 3.12, a hipótese H_0 pode ser rejeitada, a um nível de confiança α , caso a desigualdade se confirme.

Na prática, ao rejeitar H_0 assume-se que o preditor (ou o conjunto de preditores) exerce influência na variável resposta PLD , pois o parâmetro de regressão β associado a esse preditor não pode ser considerado 0, a um nível de confiança α . Dessa forma a conclusão é que caso esse preditor seja retirado do modelo regressivo, informações importantes para a explicação do PLD serão perdidas.

3.3 Passos do método empregado

Com o objetivo de se resumir o método empregado no estudo, os seguintes passos serão seguidos:

1. Definição do modelo de regressão linear como em 3.14;
2. Estimação dos valores para os coeficientes da regressão através de 3.4;
3. Avaliação da adequação do modelo através da medida do fator R^2 , utilizando 3.9;
4. Rodada de testes para verificação da hipótese de normalidade do termo de erro ε e investigação da presença de *outliers*;

5. Finalmente, avaliação da influência dos coeficientes de regressão na resposta, com o intuito de se verificar quais são os preditores mais importantes utilizando os testes de hipóteses definidos.

Esses passos serão dados para cada conjunto de preditores, conforme descrito no capítulo 4 a seguir.

Capítulo 4

Análise dos Modelos

Este capítulo apresenta a aplicação da método apresentado no capítulo 3 na verificação das principais componentes que afetam o preço spot da energia elétrica para o subsistema Sudeste/Centro-Oeste. O primeiro modelo apresentado procura analisar qual a influência da sazonalidade no preço *spot*, para isso primeiramente é verificado qual é o comportamento do preço ao longo dos meses do ano e em seguida qual é o comportamento do preço ao longo dos meses clusterizados nas estações do ano.

Em seguida, no segundo modelo, são estudados como os aspectos relacionados ao fornecimento e produção de energia influenciam nos preços. Nesse modelo os parâmetros são inicialmente separados em dois grupos, no primeiro os parâmetros que possuem uma relação direta com o fornecimento de energia, como a ENA (Energia Natural Afluente) que é a energia obtida da vazão natural situada a montante das usinas hidrelétricas de referência e a energia armazenada nos reservatórios de água das usinas hidrelétricas. O segundo grupo de preditores são os que tem relação direta com o fornecimento de energia, ou seja, quanto foi turbinado nas usinas hidrelétricas e quanto foi gerado pelas usinas termoeletricas a combustível e nucleares.

No terceiro modelo estudado foram analisado os efeitos da demanda de energia no PLD. Os preditores representativos da demanda escolhidos para análise foram a Carga de Energia, a Carga de Demanda e o Intercâmbio de Energia entre o subsistema Sudeste/Centro-Oeste e os outros subsistemas. A Carga de Energia é a energia que foi de fato utilizada no subsistema, medida na saída das usinas controladas pelo ONS. A Carga de Demanda é uma tentativa de medição da influência dos "picos de demanda" nos preços de energia. Como será explicado a seguir, os picos de demanda são de difícil medição principalmente pela forma como a energia é caracterizada no sistema brasileiro. O Intercâmbio de Energia procura verificar se existe alguma relação entre os períodos que o sistema mais importa ou exporta energia com relação aos preços que são praticados do subsistema.

Após a análise desses três modelos é criado um modelo único que verifica a relação entre os preditores já analisados. Essa seção tem o objetivo de compor um único modelo com os melhores componentes dos modelos estudados preliminarmente, tendo em vista criar um modelo que ofereça uma melhor aderência aos resultados históricos.

4.1 Origem dos Dados

A origem dos dados utilizados para a análise são de duas fontes principais. A base de dados dos preços *PLD* é uma compilação dos preços semanais históricos mantidos pela CCEE [4] do período compreendido entre Janeiro/2003 a Fevereiro/2013 para o submercado Sudeste/Centro-Oeste, dividido nos três patamares de consumo diário, conforme definido na seção 2.1. Para tornar essa base de dados mensal, alinhado com os dados oferecidos pelo ONS como descrito a seguir, foi feita a média entre os três patamares carga de cada semana e a partir disso a média das semanas para definir um preço único para um determinado mês.

A base de dados mensal com o histórico de operação foram todos extraídos do site do ONS [17]. Os dados extraídos e utilizados para cada modelo serão apresentados nas seções correspondentes e são também do período compreendido entre Janeiro/2003 a Fevereiro/2013. Por esses dados já estarem em base mensal não foi necessário fazer nenhuma transformação, sendo utilizado, dessa forma, os dados originais oferecidos pelo ONS.

O histórico de mais de 10 anos de dados, praticamente todo o histórico disponível desde que o PLD foi criado, possibilitou embasar estatisticamente esse trabalho de forma suficiente para verificação da evolução e a influência dos preditores nos preços da energia elétrica no submercado Sudeste/Centro-Oeste.

4.2 Características Gerais do Preço Spot

Na abordagem estatística seguida neste trabalho, é importante ser feita uma análise de características estatísticas elementares do preço *spot*, para se conseguir identificar os padrões gerais do seu comportamento. Uma referência para a definição dessas estatísticas pode ser encontrada em Hill e Lewicki (2006) [9]:

Medida	Valor	Medida	Valor
Valor Min (R\$)	4.00	Valor Max (R\$)	502.45
Média (R\$)	70.62	Mediana (R\$)	36.44
Volatilidade (a.a.)	152.15%	Variança (a.a.)	231.52%
1° Quartil (R\$)	18.59	3° Quartil (R\$)	95.86
Assimetria	2.59	Curtose	8.41

Na Tabela I, feita a partir da base de dados definidas na seção 4.1, pode-se observar uma grande diferença entre a média e a mediana nos valores estudados. Essa diferença sugere que existe uma preponderância de valores mais baixos para os preços spot, verificados pela baixa

mediana, com saltos esporádicos de grande magnitude nos preços elevando a média de uma forma geral.

O *boxplot* 4.1 abaixo procura evidenciar como é a distribuição dos preços. A leitura desse gráfico é feito por base na linha mais escura dentro da caixa, representando a mediana que conforme a Tabela I vale R\$ 36,44. A linha fina da caixa abaixo da linha da mediana é o primeiro quartil e representa que 25% dos valores estão abaixo dela, a linha mais abaixo representa todos os valores que estão abaixo de $3/2$ do primeiro quartil. A linha acima da linha da mediana é o terceiro quartil e representa o valor que possui 25% das amostras maiores do que ele, e da mesma forma a linha acima desse representa todos os valores que estão acima de $3/2$ do terceiro quartil. As bolas acima desse gráfico são uma sugestão do que pode ser considerado *outlier*, ou seja, que estão acima do nível fixado anteriormente. A partir desse gráfico é possível verificar que os preços do PLD se concentram em preços relativamente baixos e que os preços mais altos são melhor distribuídos.

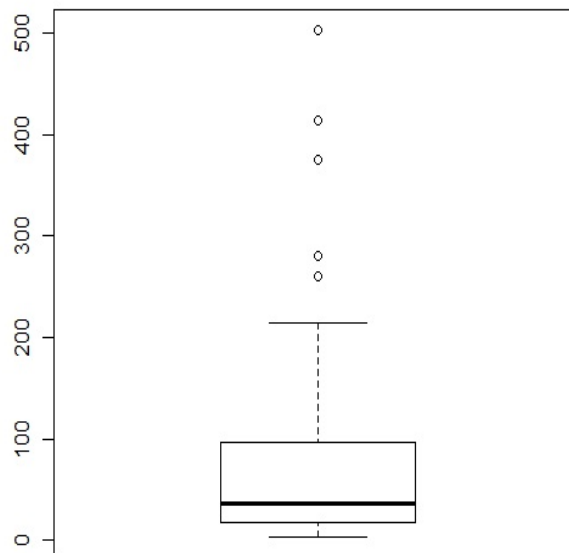


Figura 4.1: Boxplot do PLD histórico

No gráfico 4.1 é interessante observar também que os preços podem chegar a valores altíssimos, como por exemplo a R\$ 502,45 em Janeiro de 2008, como pode ser observado também através do gráfico abaixo 4.2. Como será observado a seguir esse ponto, entre outros, de fato serão caracterizados como *outliers* pois pioram o resultado final de alguns modelos.

A assimetria de 2.59 da Tabela I sugere uma cauda pesada para a direita da média na curva de distribuição de probabilidades, o que é outra característica da presença de saltos altistas nos preços. Essa tendência da curva de distribuição das probabilidades pode ser visualizada na figura 4.3.

A curtose alta de 8.41, quando comparada à distribuição normal, é característica de uma distribuição de probabilidades com um pico mais agudo ao redor do valor médio e caldas mais pesadas para os valores mais distantes da média. Na figura 4.3 é apresentado um histograma de densidade de probabilidade, no qual nota-se a concentração ao redor da média, e um segundo

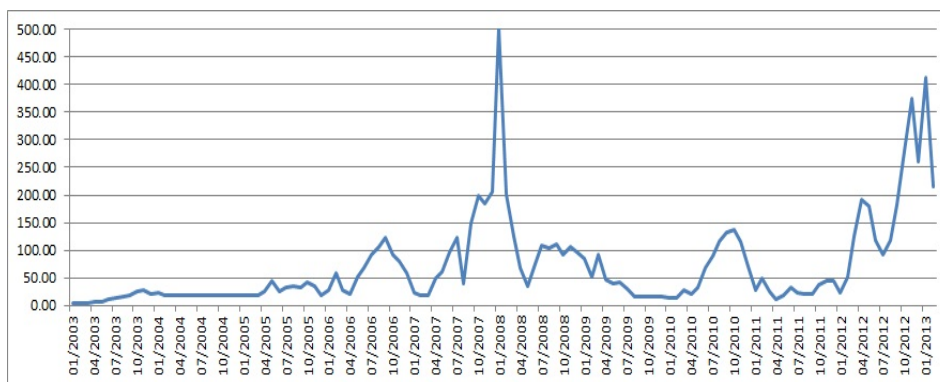


Figura 4.2: PLD histórico

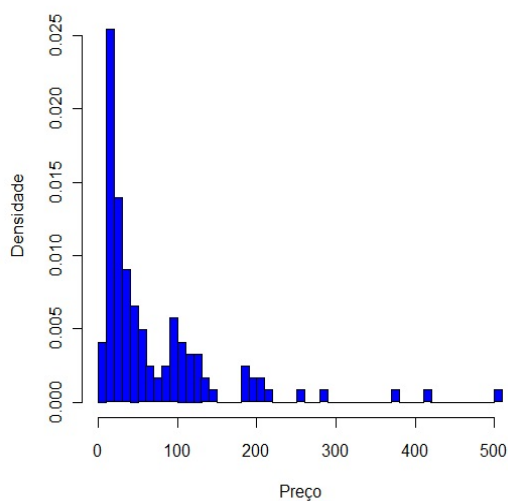


Figura 4.3: Densidade PLD

pico de densidade ao redor do preço spot de R\$ 100,00. Esses dados foram retirados da base de dados da CCEE, conforme explicitado na seção 4.1.

Uma característica constantemente observada em estudos nessa área de precificação é a alta volatilidade dos preços da energia elétrica. Em seus estudos Higgs, H. e Worthington, A.C. 2006 [10] sugerem que a alta volatilidade é causada pela componente de reversão à média e a frequência de picos de preços, onde o comportamento, reafirmado por Pilopovic D. 1998 [19], é principalmente guiado pelas características de consumo, armazenamento e clima das regiões analisadas. A alta volatilidade de 152.15% a.a., encontrada na base de dados do PLD deste estudo, é relacionada à influência dos custos de produção e pelas características de armazenagem de energia, como será visto a seguir.

Uma análise interessante é a da correlação entre o PLD e as variáveis que serão analisadas nos modelos das seções a seguir. A partir do diagrama abaixo é possível ter uma sensibilidade de quais variáveis são mais correlatadas com o preço da energia ou se influem positivamente ou negativamente nos preços. O significado de cada uma das variáveis está explicitado nas tabelas dos modelos, e podem ser verificadas nas Tabelas IV e VII a seguir.

Em uma análise primária, esse diagrama nos evidencia algumas tendências interessantes. A partir da correlação negativa que apresentam, faz sentido acreditar que a ENA e a Energia

Diagrama III : Correlações Entre Preditores e PLD			
Variável	Correlação	Variável	Correlação
ENA	-0.18	Nuclear	0.23
Energia Armazenada	-0.44	Carga de Energia	0.43
Hidrelétrico	0.28	Carga de Demanda	0.42
Termoelétrico	0.67	Intercâmbio de Energia	0.34

Armazena influenciam negativamente no preço, ou seja, quanto mais água no sistema, seja a partir de precipitação, seja a partir de armazenamento, maior é a tendência de baixa no valor do PLD. Essa situação é relativamente esperada pois as hidrelétricas possuem um custo marginal de operação menor por não precisar pagar pelo combustível, como as termoelétricas, e dessa forma com mais água no sistema a tendência é o aumento da produção de energia hidrelétrica e conseqüentemente ocasionando a diminuição do valor do PLD.

Ainda no Diagrama III, é possível verificar que, como seria de se esperar, as Cargas de Energia e de Demanda possuem uma correlação positiva com o PLD, apesar de, conforme esclarecido por Mayo 2009 [16], a demanda por energia ser inelástica e o preço não poder ser repassado ao consumidor final de forma eficiente por ser um mercado com tarifas reguladas no ACR (definido na seção 2.1). Nota-se também pelo diagrama a forte correlação entre o fator Termoelétrico e o preço, o que indica que essa variável é de grande importância para a verificação da tendência do preço *spot*.

4.3 Resposta a Parâmetros Sazonais

Na presente análise, seguindo Julia, L., Schwartz, E. 2001 em [15] e Higgs, H., Worthington, A.C 2006 in [10], o primeiro conjunto de parâmetros que será investigado são os relacionados a aspectos sazonais. Em suas apurações, a componente sazonal forma um *drift* determinístico ao modelo estocástico do preço *spot* de energia e essa seção visa verificar se esse comportamento determinístico também é válido para o subsistema Sudeste/Centro-Oeste brasileiro. Dessa forma esta componente do preço tem o objetivo de verificar a existência de regularidades no tempo, e se for o caso, a acuracidade dessas componentes na predição dos preços.

4.3.1 Originação dos Dados - Parâmetros Sazonais

Os parâmetros sazonais são estabelecidos através de variáveis *dummy*, que como explicados abaixo, coloca o valor "1" ao preditor caso o mês de referência do preço *spot* corresponder ao mês vigente e "0" para todos os outros meses. O seguinte exemplo ilustra como fica a equação de regressão para março de 2004:

$$PLD(Mar_{2004}) = \beta_0 + \beta_{fev} * 0 + \beta_{mar} * 1 + \beta_{abr} * 0 + \dots + \beta_{dez} * 0 + \varepsilon_{Mar_{2004}}$$

Dessa forma, consegue-se separar como o parâmetro de regressão β_{mar} influência os preços ao longo dos meses de março dos anos analisados.

Pela figura 4.2 apresentada anteriormente é possível observar que não há sinais claros de ciclos sazonais nos preços, como é possível verificar em países temperados que possuem esses ciclos de forma mais evidente, como Europa e Estados Unidos. A análise da seção a seguir procura oferecer um argumento estatístico para essa análise sazonal.

4.3.2 Modelo - Parâmetros Sazonais

Com o objetivo de avaliar se existe de fato uma diferença nos preços ao longo dos meses, utilizando-se a equação 3.14 da seção 3.2.1, o seguinte modelo é analisado:

$$PLD(t) = \beta_0 + \sum_{i=2}^{12} \beta_i M_{it} + \varepsilon \quad (4.1)$$

onde a variável dummy, representando o preço de referência no mês, é dado por:

$$M_{it} = \begin{cases} 1, & \text{se a data } t \text{ se refere aos mês } i \\ 0, & \text{se não} \end{cases}$$

O modelo mensal se inicia no mês $i = 2$ pela necessidade da matriz desenho "M" ser de rank completo. Dessa forma todas as variações de Janeiro são refletidas no parâmetro intercepto β_0 . O modelo resulta na estimação do parâmetro de regressão β_s , onde $s = 2, \dots, 12$, e do termo de erro ε que melhor explica a resposta da variável resposta $PLD(t)$ pelo método dos mínimos quadrados, conforme será explicitado na seção seguinte.

A análise de sazonalidade através de variáveis *dummy* é de fato a que mais se observa na literatura por ser de muito fácil interpretação. Existem porém outras formas de análise de sazonalidade, utilizando-se ainda a regressão linear. Como exemplo, pode-se citar a análise por funções cossenoidais, conforme a seguinte equação:

$$PLD(t) = \beta_0 + \gamma \cos \left((t + \tau) * \frac{2\pi}{365} \right) + \varepsilon$$

onde, nesse caso, o Método dos Mínimos Quadrados estima o intercepto β_0 , a magnitude γ e o fator de periodicidade τ . Contudo, no presente trabalho será verificado apenas o modelo sazonal de variáveis *dummy*.

4.3.3 Resultados - Parâmetros Sazonais

Os principais resultados da aplicação do método introduzido na seção 3.2.1 quando utilizado para estimar as variáveis da equação 4.1 são resumidos na Tabela II a seguir. Como dados de entrada foram utilizados os preços da CCEE, conforme citado na seção 4.1.

Tabela II : Variação Sazonal Mensal		
Parâmetro	Estimação (R\$)	p-value
β_0	105.65	5.33e-05
β_2	-42.37	0.236
β_3	-57.61	0.116
β_4	-59.60	0.105
β_5	-57.09	0.120
β_6	-50.01	0.172
β_7	-43.67	0.233
β_8	-46.77	0.202
β_9	-25.18	0.491
β_{10}	-11.42	0.754
β_{11}	-4.80	0.895
β_{12}	-24.59	0.501
Desvio Padrão do Resíduo		83.32
R^2		0.06288
Estatística-F		0.6709
p-value		0.7632

Pelo coeficiente de determinação R^2 , conforme calculado em 3.9, próximo a "0" é observado que a regressão linear baseada em parâmetros sazonais mensais não ofereceu bons resultados. A estatística-F apresentada na tabela é o teste de hipótese verificado em 3.12 para parâmetros nulos, o que reforça, pelo alto p-value, o argumento que não existe evidência estatística para rejeitar a hipótese H_0 conforme o teste de hipótese definido na seção 3.2.3. A conclusão é que os β_s podem ser considerados "0" para altos níveis de significância e o modelo sazonal não se adapta bem à variável resposta PLD.

Outra forma de se analisar a influência sazonal é através da clusterização dos dados nas diferentes estações do ano. Entende-se por clusterização a união de dados estatísticos em classes, ou *clusters*, com base em determinadas características dos dados. No modelo a seguir, foi utilizado o mesmo procedimento de 4.1 porém fazendo a clusterização a partir do agrupamento dos dados de preços nos meses do ano das quatro estações verão, outono, inverno e primavera. Dessa forma, os preços de janeiro, fevereiro e março foram reunidos no preditor "verão", abril, maio e junho foram reunidos no preditor "outono", julho, agosto e setembro foram reunidos no preditor "inverno" e outubro, novembro e dezembro foram clusterizados no preditor "primavera".

No modelo de regressão é excluído o parâmetro intercepto, para se manter a matriz desenho M com *rank* completo, conforme pontuado na seção 3.1.1. Os resultados são resumidos na seguinte tabela:

Tabela III : Efeito Sazonal			
Parâmetro	Estimação (R\$)	Desvio padrão	p-value
β_{verao}	73.08	14.45	1.56e-06
β_{outono}	50.08	14.92	0.00106
$\beta_{inverno}$	67.11	14.92	1.62e-05
$\beta_{primavera}$	92.05	14.92	9.93e-09
Desvio padrão do resíduo			81.71
R^2			0.4464
Estatística-F			23.79
p-value			1.923e-14

Apesar do modelo não oferecer um bom resultado, na Tabela III é possível verificar um coeficiente R^2 mais alto que o modelo anterior, que verificou os meses tratados individualmente. Conduzindo os testes de normalidade do termo de erro descritos na seção 3.2.2, se verifica pelo gráfico Q-Q 4.4 abaixo dos resíduos do modelo, que os valores desviam do que seria esperado pois os quantis amostrais representados pelos círculos não se apoiam sobre a reta suporte principalmente nos quantis mais altos da distribuição normal, o que dificulta o argumento de se confirmar a hipótese de que os resíduos são distribuídos conforme uma curva normal. Um teste de normalidade de Shapiro-Wilk apresenta um p-value de $3.477e - 13$, colaborando com o argumento para se rejeitar a hipótese de normalidade do termo de erro.

Dada a verificação de que distribuição dos resíduos não é normal, não é possível utilizar os testes da razão de verossimilhança, conforme definidos na seção 3.1.4, o que impossibilita a análise do modelo na busca de se encontrar as estações do ano (os preditores do modelo) mais representativas para o PLD. Dessa forma, os meses do ano quando vistos individualmente ou mesmo clusterizados nas estações, não apresentam um modelo aderente à realidade, o que leva à conclusão que o PLD não pode ser vistos através de um viés unicamente sazonal.

4.4 Resposta a Parâmetros de Suprimento

O sistema de suprimento de energia no mercado brasileiro é basicamente hidrotérmico. Uma vez que o sistema funciona por ordem de mérito de custo marginal de operação, isto é, os métodos produtivos de energia mais baratos são escalonados para serem consumidos anteriormente aos demais, até um limite que não comprometa o risco de suprimento futuro através do despacho hidráulico, de fato seria de se esperar que as variáveis que possuem uma relação direta com o método produtivo possuem altas probabilidades de influenciarem os preços.

Segundo Mayo (2009) [16] as principais variáveis que influenciam o Custo Marginal de Operação (CMO), que é a base de formação do PLD, são:

1. Nível atual dos reservatórios das usinas
2. Afluência do passado recente

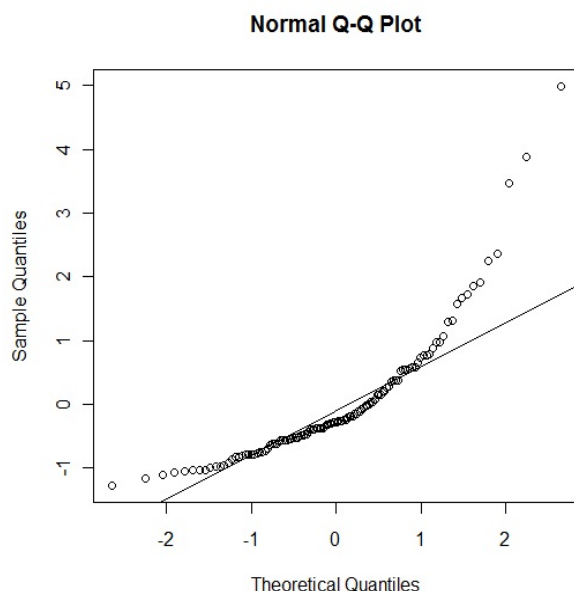


Figura 4.4: Gráfico Q-Q - Modelo Sazonal

3. A carga de consumo prevista
4. Curva de Aversão ao Risco
5. Expansão do sistema com relação à geração, transmissão e distribuição
6. Limitações nas conexões inter-regionais
7. Disponibilidade do sistema de geração e transmissão
8. Custo dos combustíveis na geração térmica
9. Função custo de déficit
10. Taxa de desconto

Desses itens, serão analisados com diferentes graus de precisão os itens 1, 2, 5, 6, 7 e 8. Com relação aos itens 1, 2 e 6 esses dados fazem parte das métricas utilizadas pelo ONS para o controle de despacho no sistema e são quantificadas e disponibilizadas. O item 5 pode ser medido através da evolução das fontes geradoras e sua previsão futura depende de fatores econométricos. Para o item 8, dada a falta de dados históricos para a análise, onde seriam necessários os preços históricos de compra de combustíveis nas usinas e em que proporção foram utilizados, é mais difícil a sua mensuração direta. Sabe-se porém que quanto maior a utilização do sistema termelétrico, mais caro é o custo marginal de operação, o que é um bom indício do custo do combustível propriamente dito. O custo está diretamente ligado ao volume de energia gerado pelas termelétricas e é a partir dessa perspectiva que esse custo relacionado às termelétricas será relacionado.

Nesta seção serão analisados como os itens 1, 2, 5 e 8 influenciam o PLD através dos valores históricos desses elementos como preditores, ou seja, os elementos da matriz desenho M

conforme descritos na seção 3.14. Os itens 6 e 7 serão verificados na seção seguinte, que tratará da análise dos Parâmetros de Demanda.

4.4.1 Originação dos Dados - Parâmetros de Suprimento

Inicialmente, o modelo será testado pela suposição que os preços serão guiados por cinco componentes de suprimentos, essas componentes serão definidas as seguir:

Tabela IV : Parâmetros de Suprimento		
Componente (Abreviação)	Descrição	Unidade de medida
ENA (E)	Afluência natural avaliada como energia turbinada no ponto de vista da usina hidrelétrica	MW médio
Energia Armazenada (S)	Água armazenada nos reservatórios das usinas hidroelétricas	GWh
Hidrelétrico (H)	Total de energia hidrelétrica produzida no subsistema	GWh
Termoelétrico (T)	Total de energia termoelétrica comum produzida no subsistema	GWh
Nuclear (N)	Total de energia termonuclear produzida no subsistema	GWh

Essas componentes podem ser divididas em dois grupos. O primeiro grupo, composto por "ENA" e "Energia Armazenada", são as componentes diretamente ligadas ao suprimento para produção de energia, no sentido de que são meios de produção utilizados pelas usinas hidrelétricas. Elas foram escolhidas pois existe uma preponderância de fontes de energia hidrelétrica no Brasil e dessa forma o teste pode ser importante para a verificação de como essas componentes estão ligadas ao preço *spot*.

O segundo grupo, composto por "Hidrelétrico", "Termoelétrico" e "Nuclear", representam os principais meios de produção de energia no subsistema. No sistema brasileiro é sabido, pelos próprios inputs utilizados para o cálculo do PLD, que o preço *spot* é reflexo dos Custos Marginais de Produção de energia elétrica, e dessa forma é de se esperar que essas componentes estejam entre as mais importantes no sentido de influência no comportamento do preço *spot*.

É importante notar que dada a dificuldade de se apurar qual o custo relacionado à produção do setor termoeletrico, no sentido de precificar-se o seu real custo de produção para cada uma das usinas, a participação do volume de energia produzido pelas termoeletricas é de grande valia para a apuração do quanto essa variável influencia o preço do PLD, pois no fim da linha o CMO será guiado pelas usinas termoeletricas mais caras que estão sendo despachadas.

Os dados foram obtidos da base dados do site do ONS disponível em [17], e foram oferecidas em base mensal conforme descrito na seção 4.1. As unidades de medidas escolhidas para as componentes são as usualmente utilizadas pelos participantes do sistema. Uma característica da análise por regressão linear é que as diferenças entre as unidades de medidas dos preditores são equalizadas pelos parâmetros da regressão, por exemplo, se o preditor Termoeletrico M_T , como apresentado abaixo, é medido em GWh , o parâmetro β_T correspondente será medido em $R\$/GWh$ e na multiplicação entre o preditor e o seu parâmetro é equalizada a unidade de medida do PLD em $R\%$.

4.4.2 O Modelo - Parâmetros de Suprimento

A partir do modelo base dado pela equação 3.14, a equação de regressão linear baseada em suprimentos será dada por:

$$PLD = \beta_0 + \beta_E M_E + \beta_S M_S + \beta_H M_H + \beta_T M_T + \beta_N M_N + \varepsilon \quad (4.2)$$

onde as variáveis são apresentadas com a legenda abreviada apresentada na Tabela IV e o parâmetro β_0 é o intercepto da regressão linear.

Nessa equação é importante verificar que não existe dependência do tempo, somente componentes que podem ser observadas no sistema sem relações diretas a aspectos temporais. Dessa forma, o objetivo principal não é a verificação de como as variáveis dos parâmetros de suprimento estão correlacionadas em aspectos temporais com o PLD e sim verificar qual é a influência do parâmetro como uma variável da função resposta PLD.

4.4.3 Resultados - Parâmetros de Suprimentos

Com o objetivo de avaliar o modelo 4.2, é seguido o procedimento definido na seção 3.2.1. Os resultados da estimação por mínimos quadrados são resumidos na tabela a seguir:

Tabela V : Efeitos do Suprimento			
Parâmetro	Estimação	Desvio Padrão	p-value
β_0	-52.0	4.9e+01	0.29
β_E	-4.0e-04	3.4e-04	0.23
β_S	-9.7e-04	3.1e-04	0.00203
β_H	1.1e-02	3.4e-03	0.00135
β_T	4.3e-02	6.5e-03	1.64e-09
β_N	7.5e-03	1.9e-02	0.69
Desvio padrão do resíduo			57.46
R^2			0.5301
Estatística-F			26.17
p-value			<2.2e-16

Pode-se observar que as estimações dos parâmetros são de diferentes ordens de grandeza, e em primeira instância isso faz parecer que, por exemplo, o parâmetro $\beta_T = 4.3e - 02$ é 100 vezes mais representativo no modelo que o parâmetro $\beta_E = -4.0e - 04$. Essas diferenças de magnitude dos parâmetros são geralmente compensadas pelos valores dos preditores, por exemplo, o valor médio do preditor Termoelétrico é 1.340 GWh enquanto o valor médio da ENA é 34.841 MW médio. Essa propriedade da regressão linear é útil quando se está trabalhando com preditores de diferentes magnitudes ou unidades de medidas.

Pelo termo R^2 da Tabela V, calculado como em 3.9, a regressão linear baseada em parâmetros de suprimento apresenta um melhor resultado quando comparada à análise de efeitos sazonais da seção 4.3, mas de qualquer forma ainda não apresentando um bom resultado. Com o objetivo de verificar-se a hipótese de normalidade do termo de erro ε , primeiramente foi conduzido o teste de normalidade de Shapiro-Wilk, conforme definido na seção 3.2.2, o qual apresentou um p-value de $1.281e - 12$, o que mostra que a partir desse teste de hipótese de distribuição normal poderia ser rejeitada.

O gráfico Q-Q apresentou uma aderência muito boa para os quantis mais baixos, mas sugere uma distribuição com caldas mais pesadas em sua curva de densidade do que seria esperado para uma curva normal, como pode ser observado na figura abaixo 4.5 (Modelo Inicial). A simetria de 1.51 na análise do termo de erro reforça essa ideia. Esses resultados também podem ser causados por *outliers* na base dados, o que se mostrou verdadeiro na condução de análise dos dados efetuados. Após esta análise se verificou que o gráfico Q-Q fornece evidências suficientes para não se rejeitar a hipótese de normalidade dos dados, apesar da tendência de caudas pesadas como podem ser verificados no gráfico abaixo 4.6 (Modelo Final).

Na primeira análise, os *p-values* apresentados na tabela acima para o Intecepto (β_0), para a ENA (β_E) e para a Nuclear (β_N) sugerem que esses parâmetros da regressão poderiam ser descartados. Após a retirada de um *outlier*, um salto nos preços de 145.2% em Janeiro/2008, esses parâmetros de regressão se apresentaram sem significância pelo teste de hipótese verificado através da aplicação de 3.12, resultando em um *p-value* de 0.11. Dessa forma a hipótese H_0 , de não influência desses parâmetros na regressão, não pode ser rejeitado para altos níveis de

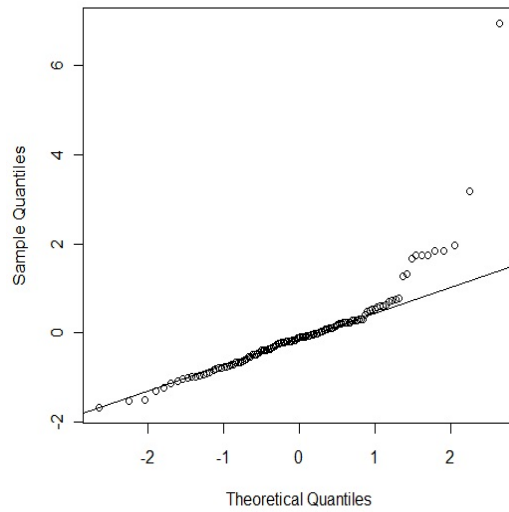


Figura 4.5: Gráfico Q-Q - Modelo de Suprimentos Inicial

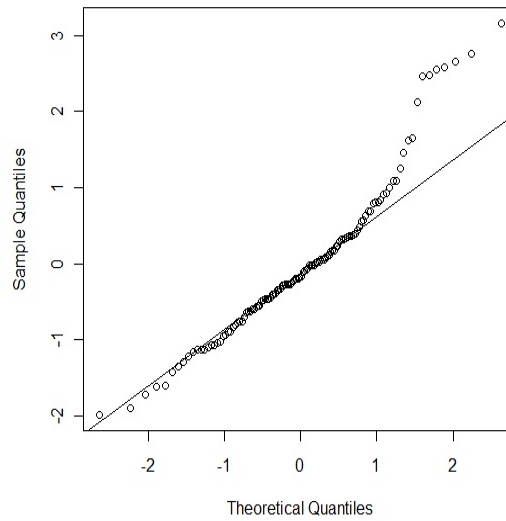


Figura 4.6: Gráfico Q-Q - Modelo de Suprimentos Final

significância.

Os principais resultados do modelo de regressão linear sem os parâmetros rejeitados são apresentados na Tabela VI:

Tabela VI : Efeitos de Suprimento 2			
Parametro	Estimação	Desvio Padrão	p-value
β_S	-9.8e-04	2.2e-04	1.97e-05
β_H	7.0e-03	1.7e-03	8.15e-05
β_T	4.4e-02	4.9e-03	9.50e-15
Desvio padrão do resíduo			44.7
R^2			0.7986
Estatística-F			156
p-value			<2.2e-16

Pelo fator R^2 de 0.7986 é possível verificar que o PLD possui uma parcela significativa do seu comportamento guiado por fatores de suprimentos. Os p -values dos parâmetros oferecem evidências estatísticas suficientes para concluir que os três parâmetros são relevantes para os preços. Para se verificar a qualidade do modelo de regressão linear é apresentada na figura 4.7 a distribuição dos Resíduos contra o \widehat{PLD} , apresentando uma boa distribuição dos pontos, sem anomalias significativas.

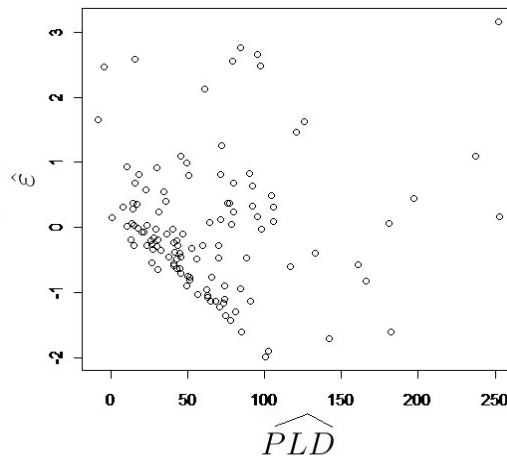


Figura 4.7: Suprimentos : Resíduos x PLD

Dessa forma, os parâmetros que são mais representativos para o modelo de regressão linear baseados em parâmetros de Suprimentos são a Energia Armazenada, a geração de energia Hidrelétrica e a geração de energia Termoelétrica. Uma observação interessante, e não intuitiva, é que a ENA atual aparentemente não influenciam os preços. A análise dos dados evidencia uma alta correlação entre a ENA e a Energia Armazenada, como esperado, mas essa correlação oferece melhores resultados quando a série ENA é atrasada em três meses, verificando uma correlação de 0.74. Dessa forma, estatisticamente, a série ENA não traz novas informações para o modelo quando a Energia Armazenada já é conhecida, porém seria útil se o objetivo fosse a predição da série de Energia Armazenada. O gráfico 4.8 apresenta de forma visual a defasagem entre a Energia Armazenada e ENA para zero, um e três meses de defasagem. Nota-se uma forte relação no formato das curvas entre a Energia Armazenada e a ENA defasada em 3 meses.

4.5 A Resposta a Parâmetros de Demanda

Como apontado por Mayo R. (2009) em [16], um fator representativo na análise de risco do preço *spot* de energia elétrica é a volatilidade causada pelos rápidos saltos ocasionais nas demandas por energia. No mercado de energia brasileiro, os reflexos desses rápidos desbalanceamentos entre a demanda e o suprimento são difíceis de se evidenciar quando se analisam os preços, principalmente devido a forma de medição semanal do PLD e a inelasticidade provocada pelo mercado regulado de energia. Dessa forma, os modelos a seguir verificam como as mudanças

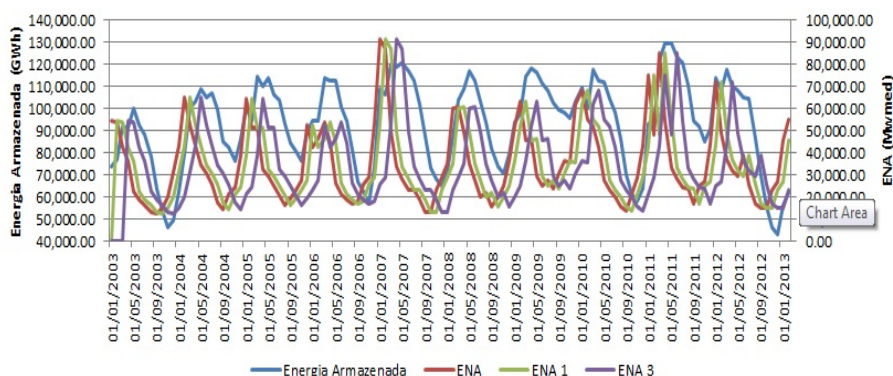


Figura 4.8: Relação Energia Armazenada x ENA

volumétricas na demanda alteram os preços em uma base mensal e não se concentrarão na medição da influência dos lapsos de curto prazo entre o suprimento e a demanda.

4.5.1 Originação de Dados - Parâmetros de Demanda

Com o objetivo de se medir os efeitos da demanda nos preços, a análise é feita a partir de três preditores:

Tabela VII : Parâmetros de Demanda		
Componente (Abreviação)	Descrição	Unidade de Medida
Carga de Energia (CE)	Demanda requerida pelo sistema suprido em base mensal	GWh
Carga de Demanda (CD)	Mais alta média de demanda medida em intervalos de 60 segundos no mês	MWh/h
Intercâmbio de Energia (IE)	A energia total intercambiada entre o subsistema Sudeste/Centro-Oeste e os outros subsistemas no mês	GWh

Os dados foram obtidos da mesma base de dados das variáveis dos preditores de suprimento [17]. A componente "Carga de Energia" é a demanda verificada no subsistema através da medição da soma da energia produzida em todas as geradoras de energia controladas pelo ONS, completada pelo preditor "Intercâmbio de Energia", que avalia a energia intercambiada entre os

subsistemas, do ponto de vista do subsistema Sudeste/Centro-Oeste. Nesse preditor os valores positivos representam energia exportada e valores negativos significam energia importada.

O preditor "Carga de Demanda" é uma tentativa de verificação da existência de picos de demanda refletidos nos preços, mas, como explicado anteriormente, esse fator não traz muita informação pois o preço *spot*, na forma como é calculado no mercado brasileiro, não pode ser de fato encarado como uma medida instantânea do preço da energia.

4.5.2 O Modelo - Parâmetros de Demanda

Na mesma linha de (3.14), o modelo de regressão linear é dado por:

$$PLD = \beta_0 + \beta_{CE}M_{CE} + \beta_{CD}M_{CD} + \beta_{IE}M_{IE} + \varepsilon \quad (4.3)$$

onde as variáveis são apresentadas com a legenda abreviada apresentada na Tabela VII e o parâmetro β_0 é o intercepto da regressão linear.

Todas as demais suposições do modelo de suprimento também valem para o modelo de demanda.

4.5.3 Resultados - Parâmetros de Demanda

O Método dos Mínimos Quadrados aplicados ao modelo de regressão linear, conforme apresentado na seção 3.2.1, é resumido na seguinte tabela:

Tabela VIII : Efeito da Demanda			
Parâmetro	Estimação	Desvio Padrão	p-value
β_0	-353.1	6.8e+01	8.5e-07
β_{CE}	1.0e-02	7.6e-03	0.17
β_{CD}	4.0e-03	4.9e-03	0.42
β_{IE}	2.8e-02	5.4e-03	1.2e-06
Desvio padrão do resíduo			67.62
R^2			0.3378
Estatística-F			20.07
p-value			1.408e-10

Conforme esperado, pelo *p-value* dos estimadores β_{CD} verifica-se que esses parâmetros não afetam o comportamento dos preços significativamente. O fator R^2 não apresentou um bom resultado (0.3378). Na mesma linha da análise de regressão a preditores de suprimento, o mesmo *outlier* afetou a regressão e, após a sua remoção, a suposição de normalidade pode ser aceita pela observação do gráfico Q-Q, apesar da persistência de uma calda pesada levemente acima do que seria esperado para uma distribuição normal.

O gráfico 4.9 apresenta duas figuras, a primeira 4.9(a) mostrando a distribuição entre os resíduos $\hat{\epsilon}$ contra o \widehat{PLD} , onde não é verificada nenhuma tendência que comprometa a distribuição dos resíduos e o segundo gráfico 4.9(b) que apresenta o gráfico Q-Q no qual verifica-se a calda pesada à direita, apesar da boa aderência à curva normal para os quantis mais baixos.

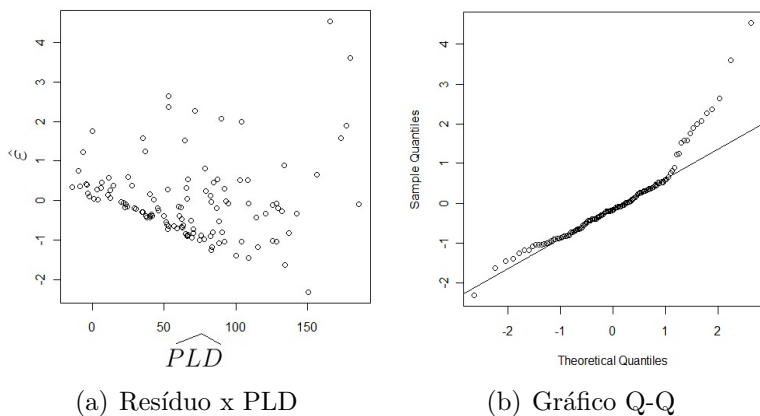


Figura 4.9: Modelo Demanda

Aplicando-se os testes de hipóteses para verificação da possibilidade de remoção dos preditores menos significativos, conforme metodologia da seção 3.2.3, evidenciou-se a possibilidade de remoção do preditor M_{CD} do modelo de regressão. Dessa forma os principais resultados do modelo de regressão, após a retirada dos *outliers* e dos preditores menos significativos, são apresentados na tabela a seguir:

Tabela IX : Efeitos da Demanda 2			
Parâmetro	Estimação	Desvio Padrão	p-value
β_0	-320.2	50.0	3.2e-09
β_{CE}	0.02	2.2e-03	4.07e-11
β_{IE}	2.6e-02	4.5e-03	3.42e-08
Desvio padrão do resíduo			56.15
R^2			0.4065
Estatística-F			40.41
p-value			4.276e-14

Pelo fator R^2 no valor de 0.4065 verifica-se que somente o efeito da demanda não é suficiente para oferecer um modelo de regressão linear que explique o comportamento do PLD, apesar de não apresentar um valor baixo o suficiente a ponto de se concluir que o preço *spot* não é influenciado em nada por efeitos de demanda. Um ponto interessante é que entre os modelos testados, esse foi o único que além dos preditores observados, também possui um forte representação do intercepto β_0 para compor a equação de regressão.

4.6 Modelo Final

Como conclusão dos resultados, a presente análise tem o objetivo de se verificar a adequação de um modelo que combine os melhores preditores dos modelos estudados. A regressão linear oferece liberdade suficiente para conseguir integrar em um modelo os preditores das diferentes classes que foram analisadas individualmente, ou seja, através dos melhores preditores sazonais, de suprimento e de demanda foi construído um modelo único afim de se avaliar se o resultado seria aderente ao PLD observado. Dessa forma, o modelo final possui a seguinte equação:

$$PLD = \beta_{verao}M_{verao} + \beta_{outono}M_{outono} + \beta_{inverno}M_{inverno} + \beta_{primavera}M_{primavera} + \beta_T M_T + \beta_S M_S + \beta_{CE} M_{CE} + \beta_{IE} M_{IE} + \varepsilon \quad (4.4)$$

Como se verifica, nessa equação foram colocados os quatro preditores sazonais M_{verao} , M_{outono} , $M_{inverno}$ e $M_{primavera}$, que, apesar de não apresentarem um resultado bom quando analisados individualmente, em conjunto com outros preditores podem contribuir com o resultado final. Conforme verificado na seção 4.3.3, uma característica desses quatro preditores é que por serem variáveis *dummy* dos meses do ano, no caso de se colocar o intercepto β_0 a matriz desenho M perderia a característica de ser de *rank* completo, pois o preditor do intercepto (uma coluna de 1s) seria combinação linear dos preditores sazonais. Dessa forma, o intercepto β_0 , presente no modelo final de Demanda, não pode constar no Modelo Final.

Os preditores do Modelo de Suprimento presentes no modelo final são os mesmos presentes no modelo da Tabela VI, ou seja, os preditores Termoelétrico (M_T - total de energia termoelétrica produzida) e Energia Armazenada (M_S - nível dos reservatórios das hidrelétricas). Verificou-se que o preditor hidrelétrico não era necessário no Modelo Final, como esclarecido a seguir. A contribuição do Modelo de Demanda se deu pelos preditores presentes na Tabela IX, sendo eles o preditor Carga de Energia (M_{CE} - demanda requerida pelo subsistema Sudeste/Centro-Oeste) e Intercâmbio de Energia (M_{IE} - energia intercambiada entre o subsistema Sudeste/Centro-Oeste e os outros subsistemas).

O primeiro passo da simulação mostrou que, quando as componentes mais importantes de suprimento e de demanda são combinadas, a componente β_H apresenta um alto *p-value*, e então em teoria poderia ser descartada. Em uma análise primária, isso pode ser explicado pela alta correlação entre os preditores M_H e M_{EC} , que verifica que em termos estatísticos essas duas componentes agregam ao modelo praticamente as mesmas informações. Devido à alta importância da energia hidrelétrica ao sistema, quase toda oscilação de demanda de energia é absorvida pelas usinas hidrelétricas, e dessa forma, do ponto de vista do preço *spot*, um modelo mais simples pode ser encontrado pela eliminação da componente Hidrelétrica do modelo. Isso é bem representado pela evolução temporal das duas curvas conforme a figura 4.10 a seguir:

Os principais resultados do modelo final são apresentados na Tabela X:

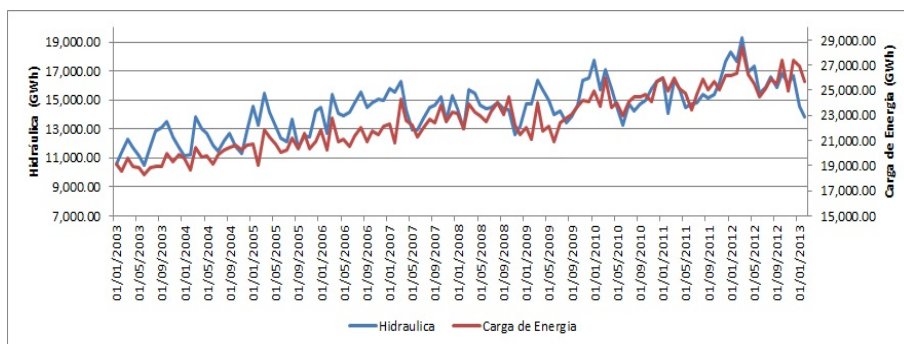


Figura 4.10: Hidráulica e Carga de Demanda

Tabela X : Modelo Final			
Parâmetro	Estimação	Desvio Padrão	p-value
β_T	2.8e-02	6.4e-03	3.4e-05
β_S	-1.5e-03	3.7e-04	1.4e-04
β_{CE}	1.2e-02	2.4e-03	2.1e-06
β_{IE}	1.2e-02	3.8e-03	2.5e-03
β_{verao}	-114.5	45.3	0.01
β_{outono}	-93.1	46	0.05
$\beta_{inverno}$	-113.9	44.2	0.01
$\beta_{primavera}$	-134.1	43.6	2.6e-03
Desvio padrão do resíduo			42.6
R^2			0.825
Estatística-F			66.53
p-value			< 2.2e-16

O modelo final apresenta o mais alto $R^2 = 0.825$ entre todos os modelos estudados e esse fator representa uma alta adequação à variável resposta PLD. O que se nota nesse modelo é que apesar das componentes Sazonais e de Demanda não apresentarem bons resultados quando estudadas em modelos isolados, quando em conjunto ao modelo de Suprimento, essas componentes auxiliam o atingimento do R^2 máximo. Nesse modelo todos os p-values possuem uma significância igual ou inferior a 5%, o que é um indício de que não consegue-se um modelo mais aderente com a eliminação dos preditores menos significativos.

Capítulo 5

Introdução à Predição do PLD

Uma vez definido o modelo que oferece uma resposta estimada aderente à realidade, conforme a seção 4.6, é possível utilizar esse modelo para se fazer predição de valores. O presente estudo não teve o objetivo de fazer a predição do PLD propriamente dito e sim de verificar quais componentes influenciam o preço *spot* de energia para o submercado Sudeste/Centro-Oeste, indicando dessa forma, quais componentes precisam ser estimadas para se ter um bom modelo preditivo. De qualquer forma, até para indicar a direção dos próximos trabalhos, será dada uma pequena introdução teórica de uma proposta de ferramenta para construção de um modelo de predição do PLD.

Dado um vetor de preditores selecionados, ou seja, valores definidos para a quantidade de energia termelétrica produzida (M_{T0}), valores de energia armazenados nos reservatórios das hidrelétrica (M_{S0}), valores de demanda por energia (M_{CE0}), quantidade de energia intercambiada entre os outros submercados e o submercado Centro-oeste/Sudeste (M_{IE0}) já definidos e a estação do ano correspondente para análise (M_{verao0} , $M_{outono0}$, $M_{inverno0}$ ou $M_{primavera0}$) a seguir será introduzido uma forma de resolução para o problema de predição.

Inicialmente seja:

$$M_0 = [M_{T0}, M_{S0}, M_{CE0}, M_{IE0}, M_{verao0}, M_{outono0}, M_{inverno0}, M_{primavera0}]$$

o vetor dos preditores observados M_0 .

A partir da equação 3.14 e sabendo-se que $\varepsilon \sim N_n(0, \sigma^2 I)$, e logo $E[\varepsilon] = 0$, tem-se que:

$$\begin{aligned} E(PLD_0|M_0) = M'_0\beta = M_{T0}\beta_T + M_{S0}\beta_S + M_{CE0}\beta_{CE} + M_{IE0}\beta_{IE} + \\ + M_{verao0}\beta_{verao} + M_{outono0}\beta_{outono} + M_{inverno0}\beta_{inverno} + M_{primavera0}\beta_{primavera} \end{aligned} \quad (5.1)$$

Nesse caso, segundo Johnson e Wichern 2009 [14] a estimação de mínimos quadrados é dada por $\widehat{PLD}_0 = M'_0\hat{\beta}$, com a variância mínima dada por $Var(M'_0\hat{\beta}) = M'_0(M'M)^{-1}M_0s^2$. Nesse caso o intervalo de confiança de $E(PLD_0|M_0)$ ao nível de confiança α será dado por:

$$M'_0\hat{\beta} \pm t_{n-r-1}\left(\frac{\alpha}{2}\right)\sqrt{(M'_0(M'M)^{-1}M_0)s^2} \quad (5.2)$$

onde n é o número de amostras do histórico observável que parametrizaram o modelo de MMQ, r é o número de preditores, M é o matriz desenho conforme definido na seção 3.2.1, $t_{n-r-1}(\frac{\alpha}{2})$ é o percentil superior ao nível $\frac{\alpha}{2}$ de uma distribuição *t-student* com $n - r - 1$ graus de liberdade.

Por esse intervalo, encontra-se qual o valor esperado de PLD_0 no plano dado pelas respostas projetadas \widehat{PLD} com uma distribuição a um nível α de confiança. Ou seja, dado um modelo já estabelecido definido pelos valores históricos de preditores (M) e respostas (PLD), encontra-se qual o valor esperado (valor médio) para a resposta PLD_0 dado os valores de preditores, já conhecidos a priori, M_0 no envoltório de distribuição da estimação \widehat{PLD} .

A figura 5 faz uma representação do que seria a predição do valor de PLD em função da quantidade produzida de energia termoelétrica (M_{T0}) e da energia armazenada (M_{S0}). Neste caso foi feito somente um *backtesting* com observações já feitas do PLD , comparando-se com o que seria esperado para predição do valor $E(PLD_0|M_0)$. Para efeitos visuais foi utilizado no gráfico dois dos principais componentes que influenciam o PLD, apesar do algoritmo utilizado para predição envolver todos os componentes do Modelo Final.

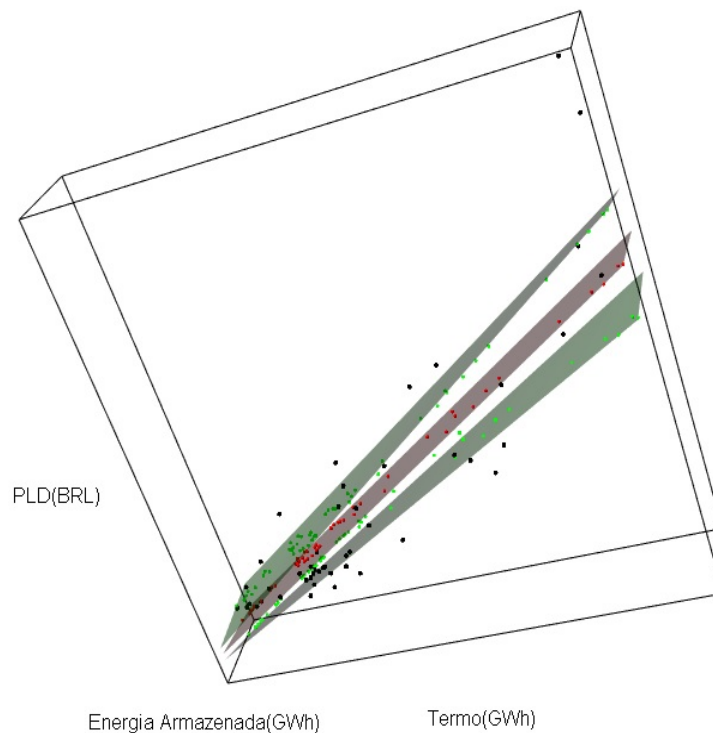


Figura 5.1: Predição do Valor Médio de PLD

Na figura nota-se a diferenciação pela cor dos pontos. Os pontos negros são as observações já realizadas do PLD_0 (eixo z) em função da energia armazenada M_{S0} (eixo y) e a geração de termelétricas M_{T0} (eixo x). Os pontos vermelhos são os pontos dados pelo valor esperado $E(PLD_0|M_0)$ de predição do PLD_0 dado a observação do vetor de preditores M_0 . Fixado os coeficientes de regressão $\hat{\beta}$, essas observações se encontrarão no plano vermelho formado pela combinação linear de todas as observações possíveis dos preditores M_i conforme descrito na seção 3.1.2. Dessa forma, cada um dos pontos negros possui uma projeção ortogonal no plano vermelho de predições, que corresponde aos pontos vermelhos correspondentes. Obviamente,

caso fosse feito um estudo de predição real, no qual os valores de PLD não fossem conhecidos, não existiriam os pontos negros.

Os pontos verdes são as representações dos intervalos de confiança ao nível de 99%, conforme descrito pela equação 5.2, para cada um dos valores médios previstos por $M'_0\hat{\beta}$. Esses pontos verdes não formam exatamente um hiperplano, como o hiperplano vermelho, pois dependem dos valores observados M_0 em termos de norma, conforme o trecho da equação $(M'_0(M'M)^{-1}M_0)$, porém foi incluído um plano verde com os valores máximos de distância ao plano vermelho somente para efeitos visuais. Dessa forma, para esse exemplo, os intervalos de confiança para os valores médios esperados $M'_0\hat{\beta}$ a um nível de confiança de 99% se encontrarão entre os dois planos verdes.

Esse exemplo foi somente uma representação da aplicação da teoria de predição para o caso do presente estudo. Um estudo mais aprofundado sobre a variância dos preditores M_i seria necessário para a condução de previsões do preço PLD para utilização em posicionamento de mercados ou o estudo de precificação de derivativos de energia elétrica, por exemplo.

Capítulo 6

Conclusão

O preço *spot* de energia elétrica, onde quer que seja avaliado, sempre possui uma alta volatilidade quando comparada a outros ativos como ações ou mesmo outras commodities. A volatilidade histórica anualizada do preço *spot* de energia pode-se alterar significativamente entre os países, como excedendo 900% a.a. no mercado Australiano, aproximando a 300% a.a. em Cinergy (Baseada em Ohio-EUA) e 152% a.a. no presente estudo no subsistema Sudeste/Centro-Oeste brasileiro. Na análise comparativa entre os mercados de outros países com o mercado brasileiro observa-se que as características da malha energética, os padrões de consumos, o nível de integração e a própria forma como a energia *spot* é caracterizada em cada sistema podem influenciar significativamente a volatilidade dos preços.

Além da alta volatilidade, a função de distribuição de probabilidade leptocúrtica do preço *spot*, muito característica em outros mercados, também foi observado para o mercado brasileiro. Essa distribuição de caudas pesadas é característica de processos com saltos, como observa-se na maioria dos estudos que envolvem o preço *spot* de energia elétrica. Como explicado, essa tendência se evidencia pela diferença entre o preço médio do PLD (R\$70,62) e a mediana (R\$36,44) indicando a presença de saltos altistas para os preços, além da própria curtose de 8.41, valor esse característico de variáveis aleatórias com distribuição com caudas pesadas.

Observou-se nos dois modelos de preço *spot* de energia elétrica em mercados estrangeiros apresentados no capítulo 2 uma forte componente sazonal, comportamento esse que pode ser verificado através das análises feitas nos Diagramas I e II desse capítulo, o que propicia a esses modelos a definição de uma componente determinística para o modelo de série temporal que guie os preços de forma bastante aderente durante os meses do ano. Quando os Diagramas I e II são comparados à Tabela II do subsistema Sudeste/Centro-Oeste nota-se que os p-values dos Diagramas são em geral muito mais baixos. Dessa forma, o que se verifica é que a componente sazonal não apresenta uma boa aderência quando vista no mês a mês utilizando-se variáveis *dummy* para o subsistema Sudeste/Centro-Oeste brasileiro. Os altos p-values apresentados na Tabela II praticamente propiciam a exclusão de todos os parâmetros β do modelo de regressão, a exceção do intercepto β_0 .

Nessa linha, em termos de sazonalidade o melhor resultado apurado foi a partir da clusteração dos preços *spot* em séries nas diferentes estações do ano, que em termos de aderência

à sazonalidade apresentou um coeficiente de determinação de $R^2 = 0.4462$, resultado esse que particularmente não pode ser considerado um bom resultado quando considerado individualmente. Dessa forma, para o submercado estudado, um modelo estocástico com apenas uma parcela determinística guiada por componentes sazonais não parece ser aderente à realidade e não oferecerá uma solução plausível para o problema de modelagem do preço *spot*.

O modelo que apresentou a mais alta influência no comportamento do PLD, quando analisado individualmente, foi o de preditores de Suprimentos, especialmente quando agrupados os preditores Energia Armazenada, produção Hidrelétrica e produção Termoelétrica. O modelo final com componentes de suprimentos apresentou um coeficiente de determinação de $R^2 = 0.7986$, o que por si só já pode ser considerado um modelo aderente ao PLD. Entre essas componentes, destaca-se a forte relação do PLD com a Energia Armazenada e a produção Termoelétrica.

Um resultado interessante verificado é a baixa influência da ENA atual nos preços *spot*, resultado esse explicado, em parte, pela alta correlação entre essa série e a Energia Armazenada. Verificou-se que quando a série histórica da ENA é defasada em alguns meses, a Energia Armazenada se correlaciona em muito com essa, o que indica que não há necessidade de se manter essa variável no modelo quando a Energia Armazenada já é conhecida. Esse resultado conflita com o costume do mercado em basear-se na ENA futura para fazer as suas previsões sobre o PLD, muito embasado na própria forma de cálculo ex-ante do PLD pelos algoritmos do ONS e CCEE que, valendo-se de previsões de ENA futuras, fazem o controle do despacho das usinas e conseqüentemente definem o CMO. Uma forma de tornar o mercado menos dependente da imprevisibilidade da ENA futura talvez seja através do aprofundamento dos estudos de como a ENA influencia a Energia Armazenada nos reservatórios das hidrelétricas e como os sistemas do ONS e CCEE fazem essa relação.

Em todos os modelos revelou-se a presença de saltos nos preços, com uma constante presença de valores altos desviando da média de forma relativamente rápida e voltando a distribuições mais próximas aos valores médios. Essa ocorrência se evidencia principalmente quando se analisa os gráficos Q-Q dos resíduos de cada um dos modelos. Essa tendência é de fato presente em muitos outros estudos que tratam da precificação da energia elétrica, como verificado nos artigos de Lucia e Schwartz 2001 [15] e Higgs e Worthington 2006 [10], e essa característica se evidencia não só para análise dos preços em função do tempo, como também se observa o mesmo comportamento quando o PLD está em função de outras variáveis, como preditores de Suprimento ou de Demanda.

O Modelo Final mostrou que os preditores de Demanda, através da Carga de Energia e Intercâmbio de energia, trouxeram ao modelo informação suficiente para se poder abstrair a componente Hidrelétrica β_H do modelo. Apesar das componentes Sazonais não apresentarem bons resultados quando estudadas individualmente, elas estão presentes no modelo final, ajudando a alcançar a boa aderência ao PLD medido pelo coeficiente de determinação $R^2 = 0.825$.

Conforme observado no presente estudo, o PLD para o subsistema SE/CO é mais influenciado pela energia armazenada que de fato a ENA medida no mês de referência. De qualquer

forma existe uma relação aparente entre a ENA defasada em alguns meses e a energia armazenada conforme é demonstrado na figura 4.8, tornado possível a condução de estudos para prever com certa segurança o nível de armazenamento de água nos reservatórios com base na ENA de alguns meses atrás. No caso do presente estudo encontrou-se uma maior correlação com a ENA defasada em 3 meses. A vantagem desses estudos é que uma importante variável para a predição do PLD já poderia ser identificada com certa margem de confiabilidade.

Foi apresentado no capítulo 5 uma pequena introdução exemplificativa de como seria a aplicação do Modelo Final para predição do PLD. Um estudo completo sobre a previsão de preços de energia elétrica deverá focar na qualidade de entendimento dos preditores do modelo. Exemplos de estudos sobre os preditores poderiam recair na verificação da relação entre a ENA e a energia armazenada nos submercados pelos reservatórios das hidrelétricas. É bem conhecido no mercado de energia elétrica brasileiro a influência que a ENA produz no PLD, sendo o assunto de previsão dessa componente muito bem explorado, inclusive para fins comerciais. As dificuldades de previsão dessa componente vão desde a incerteza de estimação de condições meteorológicas futuras, ou seja, a necessidade de dizer com precisão quando, quanto e em qual localidade irá chover, até a intensidade da irradiação solar nessas localidades, a presença de ventos fortes e a umidade do ar por exemplo.

Para um estudo completo será necessário também a avaliação de quanto dessa energia provinda de precipitação se torna realmente ENA para o sistema, sendo influenciado por fatores como absorção e encharcamento do solo e da calha dos rios, velocidade de escoamento da água e a presença de sistemas de irrigação interligados, por exemplo. Após tudo isso, seria necessário ainda fazer um estudo para verificar-se o quanto dessa energia realmente entra nos modelos dos sistemas do ONS, pois por mais real e bem detalhado que um modelo de entrada de energia possa ser, o ONS faz os seus cálculos para o gerenciamento do Sistema Interligado Nacional através de medições próprias que no fim da linha são as adotadas como oficiais e servem de base para o cálculo do PLD pela CCEE.

A componente que mais influencia o PLD é a quantidade de energia termelétrica gerada M_T . Isso era de fato esperado por termos no Custo Marginal de Operação a base de cálculo do PLD, que normalmente terá como preço de referência a geração de uma termelétrica, já que é a forma de energia mais cara do sistema elétrico brasileiro. O grande problema da predição dessa componente é a alta volatilidade que ela apresenta, de 112,4% ao ano, com uma clara tendência de aumento da utilização dessa fonte de energia nos últimos anos, como demonstra a figura 6.1. Apesar da alta variabilidade, as previsões M_{T0} para o curto prazo podem apresentar boa performance e pode-se inclusive tirar-se conclusões com respeito à periodicidade de utilização dessas fontes de energia.

Algumas medidas do governo federal vem contribuindo para o aumento da volatilidade do PLD, principalmente com respeito a incorporação novas componentes na sua fórmula de cálculo. No dia 08 de março de 2013, através do Diário Oficial da União [3], o governo anunciou o início de um processo que tem o objetivo de incorporar parte do custo dos mecanismos de aversão ao risco que antecipam o despacho de termelétricas ao PLD, tendo em vista a preservação dos

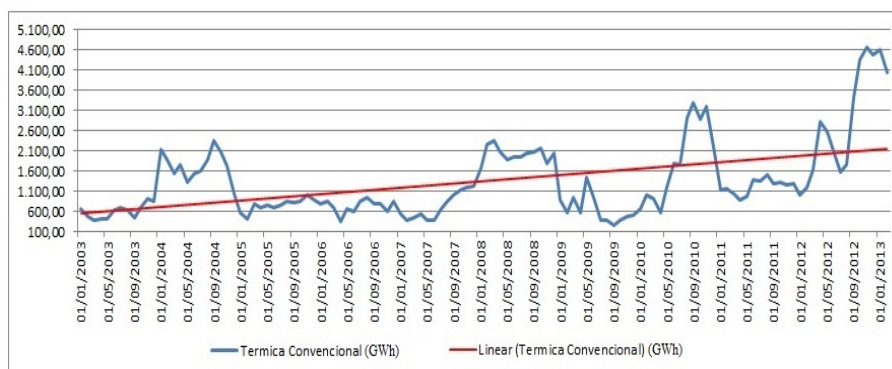


Figura 6.1: Utilização de Energia Termelétricas

reservatórios das hidrelétricas e um rateio mais justo dos custos associados ao instrumento.

Esses encargos eram primeiramente rateados entre todos os agentes do mercado via Encargo de Serviços do Sistema (ESS) por segurança energética e a partir dessa data os custos das termelétricas que são despachadas fora da ordem de mérito são em parte custeados pelos compradores do Mercado de Curto Prazo (MCP). Essa incorporação, além de aumentar o PLD em si, faz aumentar também a imprevisibilidade dos preços, pois incorpora-se ao PLD um valor que a priori não se conhece, uma vez que a escolha das usinas que gerarão a energia fora da ordem de mérito não seguem a mesma metodologia de escolha racional do "o mais barato gera energia primeiro", como é utilizado no CMO.

Verifica-se também que a ANEEL e o Governo Federal estão claramente investindo em novas formas de produção de energia elétrica, com o intuito de aumentar e diversificar a matriz energética brasileira. Do ponto de vista do aumento da oferta de energia para o mercado como um todo essa situação é bastante positiva, porém uma vez que as fontes óbvias de energia já estão cada vez mais difícil de serem encontradas, pois em grande parte já estão sendo exploradas, existe a necessidade de implantação de formas de geração de energia alternativas. Dessa forma, a implantação de usinas hidrelétricas sem reservatórios, ou a fio-d'água como são conhecidas, de grande porte na região norte e a implantação de plantas eólicas na região nordeste colocam um maior grau de imprevisibilidade na oferta de energia, dado que são formas de geração de energia com menores variáveis de controle pelo ONS.

Essa tendência de introdução de novas componentes no SIN pode resultar, como consequência, em um aumento da volatilidade do PLD para os submercados. A maior imprevisibilidade a ser inserida pode porém, em parte, ser compensada pelo melhoramento da infraestrutura de interligação entre os subsistemas, principalmente com relação às novas linhas de transmissão que estão sendo construídas ligando os subsistemas Norte e Nordeste com o Sudeste e Centro-oeste. O desafio de controle dos subsistemas interligados será cada vez mais complexo para o ONS, e o sucesso desse controle influenciará fortemente o comportamento do PLD nos próximos anos.

Com relação ao desenvolvimento do mercado de energia elétrica, o primeiro desafio para o mercado brasileiro é o desenvolvimento de mercados de bolsa eficientes. De fato, se mostra que os mercados de energia futura do mercado brasileiro estão cada vez mais líquidos, o que

é uma vantagem para todas as partes. Existem iniciativas, até por parte da própria CCEE, em desenvolver bolsas de energia para operar em mercados de entrega de energia no curto prazo ou mesmo no mercado *real-time*, como existe em muitos mercados estrangeiros conforme apresentado no capítulo 2. Se um dia chegarmos nessa realidade poderemos deixar de utilizar o *proxy* de preço *spot* que é o PLD e termos um preço *spot* de negociação de energia elétrica de fato. O desenvolvimento de *clearings* e mercados financeiros de energia elétrica também são uma clara tendência, e estudos estão sendo feitos nesse sentido. Porém, muitas dessas iniciativas esbarram na própria legislação e de como o PLD é contabilizado, o que indica que mudanças mais profundas precisarão ser feitas até o Brasil possuir um mercado de bolsa de energia estabilizado, eficiente e suficientemente líquido.

A metodologia utilizada neste estudo pode ser utilizada para a precificação de derivativos de energia elétrica em uma perspectiva diferente da qual é atualmente feita, isto é, modelando-se o preço *spot* como um processo estocástico guiado por uma componente sazonal determinística. O modelo apresentado no presente estudo é de fato aderente ao PLD, e os estudos para definição dos preditores M_0 é que caracterizarão a qualidade ou não da previsão do preço *spot*. O que se observa é que dado um histórico relativamente grande e confiável para os preditores, é possível fazer-se pesquisa para avaliação e previsão dessas componentes com certa qualidade. Essa qualidade porém fica muito comprometida se "a regra do jogo" de contabilização dos preços for constantemente alterada como observou-se recentemente. O mercado como um todo, e a própria pesquisa acadêmica, fica comprometida com essas alterações pois o que se encara como premissa em um momento pode-se não ser verificado no longo prazo, com claras penalidades para as previsões feitas.

Bibliografia

- [1] Black, F. e Scholes, M.: The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, Vol. 81, pp. 637 – 659. 1973.
- [2] CCEE, Regras de Negociação. Disponível em <http://www.ccee.org.br/>, 2010.
- [3] CCEE, Nota explicativa: Alteração do Critério de Rateio do Despacho Térmico por Segurança Energética. Disponível em: <http://www.ccee.org.br/>, 2013.
- [4] CCEE base de dados: Disponível em <http://www.ccee.org.br>, Consultada em fevereiro de 2013
- [5] Cox J. e Ross S.: The valuation of options for alternative stochastic process, *Journal of Financial Economics*, 3, 145-166.
- [6] Clewlow L. e Strickland C.: *Energy Derivatives, Pricing and Risk Management*. Lacima Publications – London, England 2000.
- [7] Deng S.: *Stochastic Models of Energy Commodity Prices and Their Applications: Mean-reversion with Jumps and Spikes*. Program on Workable Energy Regulation of the University of California Energy Institute, 2003.
- [8] Harrison M. e Kreps D.: Martingales and arbitrage in multiperiod securities markets, *Journal of Economy Theory*, 20, 381-408.
- [9] Hill T. e Lewicki P.: *Statistics: methods and applications*, StatSoft, Estados Unidos 2006.
- [10] Higgs H. e Worthington A.C.: *Stochastic price modelling of high volatility, mean-reverting, spike-prone commodities: The Australian wholesale electricity market*, University of Wollongong, Working Papers Series, 2006.
- [11] Hull J. e White A.: Pricing interest-rate derivative securities, *Review of Financial Studies*, 3, 572-592 1990.
- [12] de Jong C.: *The nature of power spikes: A regime-switching approach*, Rotterdam School of Management, Erasmus University Rotterdam, Working Paper, 2005.
- [13] Jaillet P. e Tompaidis S.: *Modeling energy prices and pricing and hedging derivatives securities*, 1998.

- [14] Johnson R., Wichern D.: Applied Multivariate Statistical Analysis, Prentice-Hall Inc., 2002.
- [15] Lucia J. e Schwartz E.: Electricity prices and power derivatives: Evidence from the Nordic Power Exchange, Review of Derivatives Reserch, 5: 5-50, 2001
- [16] Mayo R.: Risk management and electricity derivatives, Editora Synergia, Rio de Janeiro, 2009.
- [17] ONS data base. Disponível em: <http://www.ons.org.br/historico>.
- [18] Olson Pemberton Jr T. : Modelando o preço spot de energia elétrica no Brasil : Um modelo estocástico com reversão à média, mudança de regime markoviano e difusão com saltos, Dissertação apresentada ao Programa de Mestrado Profissional de Economia da Faculdade Ibmec São Paulo, 2006.
- [19] Pilipovic D.: Valuing and managing energy derivatives, MacGraw - Hill, New York, 1998.
- [20] Shapiro S. S., Wilk M. B., e Chen H. J.: A Comparative Study of Various Tests for Normality, Journal of the American Statistical Association, Volume 63, Issue 324, 1968.
- [21] TSO and Nord Pool: A powerful market, Transission System Operator companies and Nord Pool ASA, 1999
- [22] Taylor J., VanDoren P.: California's Electricity Crisis, Topics in Regulatory Economics and Policy Series Volume 42, Springer, 2002
- [23] Villaplana P., Escribano A., Pena : Modeling Electricity Prices: International Evidence. Universidad Carlos III de Madrid, Working Paper 02-27, Economics Series 08, 2002.