

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ADMINISTRAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO

MODELO DE FORMAÇÃO DE PREÇOS DE COMMODITIES AGRÍCOLAS
APLICADO AO MERCADO DE AÇÚCAR E ÁLCOOL

Leonel Molero Pereira

Orientador: Prof. Dr. José Roberto Securato

SÃO PAULO

2009

Profª. Dra. Suely Vilela
REITORA DA UNIVERSIDADE DE SÃO PAULO

Prof. Dr. Carlos Roberto Azzoni
DIRETOR DA FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE

Prof. Dr. Isak Kruglianskas
CHEFE DO DEPARTAMENTO DE ADMINISTRAÇÃO

Prof. Dr. Lindolfo Galvão Albuquerque
COORDENADOR DO PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO

LEONEL MOLERO PEREIRA

**MODELO DE FORMAÇÃO DE PREÇOS DE COMMODITIES AGRÍCOLAS
APLICADO AO MERCADO DE AÇÚCAR E ÁLCOOL**

Tese apresentada ao Departamento de Administração da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo como requisito para obtenção do título de Doutor em Administração

Orientador: Prof. Dr. José Roberto Securato

SÃO PAULO

2009

FICHA CATALOGRÁFICA

Elaborada pela Seção de Processamento Técnico do SBD/FEA/USP

Pereira, Leonel Molero

Modelo de formação de preços de commodities agrícolas aplicado ao mercado de açúcar e álcool / Leonel Molero Pereira. – São Paulo, 2009.

209 p.

Tese (Doutorado) – Universidade de São Paulo, 2009

Bibliografia.

1. Preço 2. Açúcar 3. Álcool como combustível 4. Mercado futuro
I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. II. Título.

CDD – 338.52

*À Mariana,
minha inspiração.*

Agradecimentos

Ao Prof. Dr. José Roberto Securato, meu orientador nos cursos de mestrado e doutorado, pela orientação de vida, pelos conhecimentos e didática nos cursos de pós-graduação, pelo apoio e oportunidades oferecidas durante toda a minha carreira acadêmica, pelos conhecimentos quantitativos indispensáveis em Teoria de Finanças e pela amizade.

À Prof. Dra. Celma de Oliveira Ribeiro, membro da banca na qualificação e defesa, pela orientação desse trabalho, pelas críticas construtivas e oportunas, sem as quais o objetivo central da pesquisa não seria alcançado, pelo rigor e dedicação acadêmica, pela doçura e compreensão em todos os momentos.

Ao Prof. Dr. José Roberto Savoia, membro da banca, por aceitar prontamente o convite, pelo dinamismo no desenvolvimento do Departamento de Finanças da FEA e pelas oportunidades oferecidas durante o doutorado.

Ao Prof. Dr. André Luiz Oda, membro da banca de qualificação e defesa, pelas críticas e contribuições ao trabalho, pelos conhecimentos e didática em Teoria de Finanças, pelos votos de confiança na minha carreira acadêmica e pelo companheirismo no desenvolvimento desse e de outros projetos.

Ao Prof. Dr. Ricardo Humberto Rocha, membro da banca, por aceitar prontamente o convite, pelos conhecimentos sobre o Mercado de Derivativos, pelo dinamismo e energia durante o curso de pós-graduação, pelos conselhos e oportunidades na minha carreira como docente.

Aos amigos que participaram diretamente na elaboração desse trabalho: Eduardo Morato Mello, Bert Leatherman e Ana Tereza Pinto de Oliveira, pela revisão.

Em especial, a minha querida esposa, Mariana Esteves Tafner, que participou comigo em cada minuto, privando-se de momentos mais agradáveis.

À minha família, meu pai, Antonio Leonel Pereira, pelo apoio incondicional ao meu desenvolvimento acadêmico, minha mãe, Pilar Molero Pereira, minhas irmãs, Dea, Lúcia e Selma, pela motivação e vibrações positivas, meus sogros, Antônio Carlos e Isabel, e cunhado, Rodrigo, também pelo apoio e motivação constantes.

Por fim, ao pessoal da Secretaria, Comissão de Pós-Graduação e Biblioteca da FEA.

RESUMO

PEREIRA, Leonel Molero. **Modelo de formação de preços de *commodities* agrícolas aplicado ao mercado de açúcar e álcool**. 2009. 209 p. Tese (Doutorado) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2009.

O problema estudado nesta tese foi a formação de preços de *commodities* agrícolas relacionadas com a produção de bioenergia. A possibilidade de substituição de combustíveis fósseis, derivados do petróleo, por alternativas renováveis, como o etanol proveniente da cana-de-açúcar, inseriu um novo contexto no mercado de *commodities*. O objetivo principal desta tese foi propor um modelo de formação de preços que levasse em consideração o conceito de *commodities* agrícolas como componentes da matriz energética e aplicá-lo ao mercado brasileiro de açúcar e álcool. Para a especificação do modelo, foram elaboradas premissas que têm base na interdependência de preços com o petróleo, na Teoria de Estocagem, na sazonalidade das safras e na volatilidade do mercado. A volatilidade foi considerada no estudo porque, em períodos de turbulência econômica, os investidores buscam o mercado de *commodities* para proteger o valor real do capital. Para elaborar e testar o modelo, a pesquisa foi dividida em quatro partes interrelacionadas. A primeira consistiu na análise de outros modelos da literatura e dos processos estocásticos descritos pelas variáveis que compõem os preços das *commodities*. A segunda parte consistiu na elaboração das premissas, dispostas na forma de hipóteses, que foram testadas com dados do mercado futuro de açúcar e álcool da BM&FBOVESPA e com preços do mercado agrícola divulgados pelo CEPEA. A amostra analisada compreendeu o período de 2 de janeiro de 2002 a 30 de junho de 2008. Nesta etapa foram utilizados, entre outros métodos, testes estatísticos de significância de coeficientes de regressões multivariadas pelo Método dos Mínimos Quadrados. Os resultados dos testes indicaram que o açúcar Granger causa os preços do etanol, confirmaram a presença de *backwardation* forte nas séries de preços, confirmaram também que os preços do petróleo antecipam informações sobre a tendência dos preços do açúcar, com defasagem de um mês. Além desses resultados, confirmou-se a presença de sazonalidade e, verificou-se, de forma não conclusiva, uma relação positiva entre a volatilidade dos mercados e os preços do açúcar. Na terceira parte da pesquisa, o modelo foi especificado em um sistema de três equações na forma de espaço de estado, cujos parâmetros foram estimados por meio do filtro de Kalman. Na quarta e última parte da pesquisa, foram

geradas séries de previsão n passos à frente utilizando os parâmetros estimados e os resultados confrontados com os preços observados do açúcar no mercado à vista. Medidas de erros de previsão foram calculadas e comparadas com as de um outro modelo na literatura, adequado ao mercado de brasileiro açúcar – o modelo de dois fatores. Verificou-se que o modelo proposto é estatisticamente superior, em termos de previsão, ao modelo de dois fatores, no nível de 1% de significância. Verificou-se também que, quanto maior o horizonte de previsão, maior é o ganho de informação relativo do modelo proposto. A redução percentual de erros foi superior a 10%, quando analisada a previsão de três meses à frente. Portanto, foi possível concluir que o modelo, que incorpora a interdependência do petróleo na formação de preços de *commodities* relacionadas à produção de biocombustíveis, é melhor, em termos de previsão, do que um outro modelo sugerido na literatura que não leva essa premissa em consideração, quando aplicado ao mercado brasileiro de açúcar. Os resultados da pesquisa podem ter aplicação pragmática em Administração de Empresas do setor sucroalcooleiro e na formação de preços no mercado de derivativos de *commodities* agrícolas.

Palavras-chave: açúcar, etanol, bioenergia, formação de preços, biocombustíveis, *commodities* agrícolas, mercado futuro de açúcar.

ABSTRACT

PEREIRA, Leonel Molero. **Agricultural commodity pricing model applied to the sugar and ethanol markets.** 2009. 209 p. Thesis (Doctoral) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2009.

The problem addressed by this thesis is the formation of prices of agricultural commodities related to bioenergy production. The possibility of substituting renewable alternatives, such as ethanol derived from sugar cane, for fossil fuels derived from petroleum has brought a new dimension to commodity markets. The primary objective of this thesis is to propose a pricing model which takes into account the concept of agricultural commodities as components of the energy matrix and to apply it to the Brazilian sugar and ethanol markets. The bases of the premises used to build this model were oil-price interdependence, inventory theory, harvest seasonality, and market volatility. Volatility was considered in the study because, in times of economic turbulence, investors turn to the commodities market to protect the real value of their capital. Four interrelated research tasks were undertaken to develop and test the model. The first step consisted of a literature analysis of other models and the stochastic processes to which variables that influence commodity prices are subject. The second part consisted of the elaboration of presumptions in the form of hypotheses which were tested using data from the BM&FBOVEPSA sugar and ethanol futures markets and agricultural prices published by CEPEA, with the test period being from January 2, 2002, to June 30, 2008. This step employed, among other methods, tests of statistical significance with multivariate regressions using the method of least squares. The test results indicated that sugar Granger causes ethanol prices, that strong backwardation exists in the price series, and that petroleum prices are predictive of sugar price patterns, with a one-month lag. In addition, the results confirmed the influence of seasonality and pointed, albeit somewhat inconclusively, to a positive relationship between market volatility and sugar prices. In the third step of the study, a model was defined using a system of three equations in state space form with parameters estimated using Kalman filtering. The fourth and last stage of the research generated series projections n steps forward using the estimated parameters, and the results were compared to sugar prices observed in the market. Measures of predictive error were calculated and compared with those of another model cited in literature as applicable to the Brazilian sugar market – the two-factor model. The proposed model proved itself to be statistically superior, in terms of predictiveness, to the two-factor model, at the level of 1% significance. Moreover,

its predictive superiority rose as the period of time analyzed increased. The error reduction in percentage terms was greater than 10% over a forward-looking period of three months. Therefore, it is possible to conclude, in the context of the Brazilian sugar market, that this model, which incorporates petroleum interdependence in the formation of prices of commodities related to biofuel production, is better, in terms of predictive power, than another model cited in literature that does not take this premise into consideration. The results of this research can be applied in the management of sugar-alcohol companies and in the formation of derivatives prices in the agricultural commodities market.

Keywords: *sugar, ethanol, bioenergy, pricing, biofuels, agricultural commodities, sugar futures market.*

SUMÁRIO

1	PROBLEMA DE PESQUISA.....	7
1.1	Introdução.....	7
1.2	A Situação Problema e Objetivos de Pesquisa.....	8
1.3	Justificativa da Pesquisa.....	10
1.4	Metodologia de Pesquisa.....	11
1.5	Questão da Originalidade.....	12
2	FUNDAMENTAÇÃO TEÓRICA.....	15
2.1	Introdução.....	15
2.2	Conceitos Fundamentais do Mercado de <i>Commodities</i>	16
2.2.1	Definição de Commodity.....	16
2.2.2	Mercado à Vista, Termo e Futuro.....	18
2.2.3	<i>Hedge</i> , Especulação e Arbitragem.....	23
2.2.4	Principais Mercados Futuros.....	24
2.2.5	Mercados de Açúcar e Alcool.....	26
2.2.6	Mercados de Petróleo.....	27
2.3	Teoria da Estocagem.....	29
2.3.1	Princípios da Teoria da Estocagem.....	29
2.3.2	Custo de Carregamento.....	31
2.3.3	Contango e Backwardation.....	33
2.3.4	Retorno de Conveniência.....	36
2.4	Processos Estocásticos dos Preços de <i>Commodities</i>	38
2.4.1	A Trajetória dos Preços no Tempo.....	38
2.4.2	A Variável Aleatória.....	39
2.4.3	Processo de Wiener.....	40
2.4.4	Lema de Itô.....	44
2.4.5	Reversão à Média.....	47
2.5	Modelagem de Preços de <i>Commodities</i>	50
2.5.1	Modelo com Base no CAPM.....	50
2.5.2	Modelos de Um, Dois e Três Fatores.....	52
2.5.3	Modelo de Curto e Longo Prazo.....	56
2.5.4	Sazonalidade.....	58

3	O MODELO PROPOSTO	61
3.1	Evolução dos Modelos Anteriores	61
3.2	Teoria da Paridade de Duas <i>Commodities</i> de Walras	61
3.3	Relação do Preço do Petróleo com Outras <i>Commodities</i>	63
3.4	Premissas do Modelo	65
3.5	Especificação do Modelo	67
3.5.1	Condição de Não Arbitragem do Preço Futuro	68
3.5.2	Reversão à Média do Retorno de Conveniência	69
3.5.3	Movimento Browniano do Preço à Vista	71
3.5.4	Relação do Petróleo com o Preço da <i>Commodity</i>	73
3.5.5	Relação da Volatilidade do Mercado com o Preço da <i>Commodity</i>	75
3.5.6	Componente Sazonal.....	77
3.5.7	Equações na Forma de Espaço de Estado	79
4	METODOLOGIA DE PESQUISA	83
4.1	Hipóteses de Pesquisa	85
4.2	Tratamento Estatístico e Análise dos Dados	88
4.2.1	Parte 1 – Verificação das Premissas do Modelo	88
4.2.2	Teste de Raiz Unitária e Cointegração	96
4.2.3	Volatilidade Condicional GARCH.....	98
4.2.4	Parte 2 – Estimação do Modelo com o Filtro de Kalman	102
4.2.5	Filtro de Kalman.....	103
4.2.6	Parte 3 – Testes de Capacidade de Previsão dos Modelos.....	109
4.3	Definição Teórica e Operacional das Variáveis	113
4.4	População, Amostragem e Coleta de Dados	121
4.5	Limitações da Pesquisa	123
4.5.1	Não Identificação de Variáveis Cointegradas	123
4.5.2	Causalidade Reversa e Endogeneidade	124
4.5.3	Variáveis Omitidas e Redundantes	125
4.5.4	Seletividade da Amostra.....	126
4.5.5	Erro de Mensuração das Variáveis	127
5	RESULTADOS DA PESQUISA	129
5.1	Estatísticas Descritivas	130
5.2	Estimação da Volatilidade do Mercado	133
5.2.1	Autocorrelação dos Resíduos Quadráticos.....	133

5.2.2	Resultados do Modelo GARCH.....	135
5.3	Resultados dos Testes de Raiz Unitária e Cointegração	137
5.4	Parte 1 – Resultados dos Testes das Premissas	141
5.4.1	Testes Relativos à Pergunta 1.....	141
5.4.2	Testes Relativos à Pergunta 2.....	143
5.4.3	Testes Relativos às Perguntas 3, 4 e 5.....	147
5.5	Parte 2 – Resultados do Filtro de Kalman.....	155
5.5.1	Séries Filtradas e Resultados da Previsão n Passos à Frente.....	158
5.6	Parte 3 – Resultados do Testes de Capacidade de Previsão	160
5.6.1	Medidas de Erro de Previsão.....	160
5.6.2	Resultado do Teste de Previsão Granger e Newbold	162
5.7	Resumo dos Resultados da Pesquisa	164
6	CONCLUSÕES.....	167
	Referências Bibliográficas	172
	Glossário de Variáveis	177
	Apêndice I – Teste de Normalidade.....	181
	Apêndice II – Teste de Causalidade de Granger	182
	Apêndice III – Matriz de Covariância Robusta de Newey-West	184
	Apêndice IV – Saídas do Modelo GARCH	185
	Apêndice V – Saídas do Teste de Raiz Unitária	188
	Apêndice VI – Saídas do Teste de Cointegração	194
	Apêndice VII – Saídas da Regressão Multivariada.....	195
	Apêndice VIII – Saídas do Filtro de Kalman	198

LISTA DE TABELAS

Tabela 1 - Resumo estatístico da BM&FBOVESPA do futuro de açúcar ref. 20/3/2008.....	22
Tabela 2 – Estatística descritiva das variáveis brutas – dados diários.....	130
Tabela 3 – Estatística descritiva das variáveis brutas – dados mensais.....	131
Tabela 4 – Autocorrelação dos resíduos quadráticos – dados diários.....	133
Tabela 5 – Autocorrelação dos resíduos quadráticos – dados mensais.....	134
Tabela 6 – Coeficientes do processo GARCH – dados diários.....	135
Tabela 7 – Coeficientes do processo GARCH – dados mensais.....	136
Tabela 8 – Coeficientes do processo GARCH sem processo AR – dados mensais.....	136
Tabela 9 – Estatísticas dos testes ADF de raiz unitária – dados diários.....	138
Tabela 10 – Estatísticas dos testes ADF de raiz unitária – dados mensais.....	138
Tabela 11 – Estatísticas do teste de cointegração – dados diários.....	139
Tabela 12 – Estatísticas do teste de cointegração – dados mensais.....	139
Tabela 13 – Estatística do teste ADF para etanol e açúcar à vista.....	140
Tabela 14 – Estatísticas do teste de cointegração entre etanol e açúcar à vista.....	141
Tabela 15 – Teste de causalidade de Granger para etanol e açúcar – dados semanais.....	142
Tabela 16 – Teste t unicaudal de média do <i>backwardation</i> – dados diários.....	144
Tabela 17 – Teste t unicaudal de média do <i>backwardation</i> – dados mensais.....	146
Tabela 18 – Correlograma serial cruzado entre petróleo e açúcar – dados diários.....	148
Tabela 19 – Correlograma serial cruzado entre petróleo e açúcar – dados mensais.....	149
Tabela 20 – Regressão linear múltipla MQO – dados diários.....	150
Tabela 21 – Regressão linear múltipla MQO – dados mensais.....	151
Tabela 22 – Estatísticas dos resíduos da regressão – teste de normalidade.....	152
Tabela 23 – Autocorrelação dos resíduos da regressão – dados diários.....	153
Tabela 24 – Autocorrelação dos resíduos da regressão – dados mensais.....	154
Tabela 25 – Resultados do filtro de Kalman.....	156
Tabela 26 – Medidas de erro de previsão – dados diários.....	161
Tabela 27 – Medidas de erro de previsão – dados mensais.....	161
Tabela 28 – Medidas de erro de previsão – dados mensais.....	163

LISTA DE QUADROS

Quadro 1 - Indicadores de preços divulgados pelo CEPEA em 20/3/2008.....	21
Quadro 2 - Bolsas de negociação de <i>commodities</i>	25
Quadro 3 - Especificações do álcool anidro.....	115
Quadro 4 - Variáveis analisadas no estudo estatístico.....	120
Quadro 5 - Origem e periodicidade das séries de dados coletadas.....	122

LISTA DE ILUSTRAÇÕES

Ilustração 1 - Componentes sazonais seno e cosseno para período de doze meses.....	59
Ilustração 2 - Esquema metodológico da pesquisa científica.....	85
Ilustração 3 - Retornos diários do Índice Dow Jones no período de jan./2002 a jun./2008.	99
Ilustração 4 - Gráficos das variáveis – dados diários	132
Ilustração 5 - Gráficos do etanol e açúcar a vista.....	132
Ilustração 6 - Volatilidade GARCH do Índice Dow Jones.....	137
Ilustração 7 - Gráfico dos preços à vista do açúcar observados e filtrados – dados diários.....	158
Ilustração 8 - Gráfico dos preços à vista do açúcar observados e filtrados – dados mensais.....	158
Ilustração 9 - Previsão n passos à frente do preço à vista do açúcar – dados diários.....	159
Ilustração 10 - Previsão n passos à frente do preço à vista do açúcar – dados mensais.....	159

1 PROBLEMA DE PESQUISA

*Little people discuss other people.
Average people discuss events.
Big people discuss ideas.*

(Rudolf Emil Kalman, 1985)

1.1 Introdução

A humanidade passa por um processo de evolução constante, que consiste em superar as provações impostas por ela mesma. A ciência permitiu ao homem conter a evolução de doenças, produzir nutrientes em larga escala, construir moradias seguras e confortáveis, criar transportes rápidos e sistemas de comunicação instantânea, recursos que ampliaram a longevidade e a qualidade da vida no planeta.

Como consequência desse processo evolutivo, uma parcela da população passou a ter acesso aos benefícios da vida moderna, e milhões de indivíduos passaram a exercer um novo padrão de consumo. Para abastecer esse neocontingente, houve um aumento na demanda por infraestrutura, produtos, serviços e, conseqüentemente, por **insumos e energia**.

Com o passar dos anos, o aumento no consumo sem um recíproco aumento na produção provocou redução nos níveis globais dos estoques de matérias-primas e alta generalizada nos preços das *commodities*. Além disso, a provável escassez de alguns recursos não renováveis, como o petróleo, levou à discussão sobre a adoção de substitutos. O petróleo ainda consiste em uma fonte de energia relativamente de baixo custo, mas, à medida que aumenta a dificuldade de exploração e produção de combustíveis de origem fóssil, surge a necessidade de propor alternativas não esgotáveis.

Algumas outras fontes alternativas de energia foram exploradas com sucesso, como é o caso da energia nuclear, hidrelétrica, eólica ou solar. Porém, devido à restrição dos recursos naturais ou tecnológicos, essas fontes talvez não sejam suficientes para substituir toda a energia fornecida pelo petróleo e para atender o aumento na demanda. Nesse contexto, a

fotossíntese das plantas surge como alternativa renovável de captação de energia que pode ser armazenada na forma de **biocombustíveis** ou **agrocombustíveis**.

A partir do momento em que as oleaginosas se tornaram economicamente viáveis para serem utilizadas como biodiesel – ou a cana-de-açúcar, para substituir a gasolina na forma de etanol – surge uma nova interpretação para as condições da formação dos preços das *commodities* agrícolas que, antes destinadas apenas à alimentação, se tornaram componentes relevantes da **matriz energética**. É dentro desse contexto que foi desenvolvida a presente pesquisa sobre a formação dos preços de *commodities*, levando em consideração o fato de que os produtos agrícolas passaram a fazer parte da matriz energética e sugerindo uma provável relação de interdependência na formação de preços das *commodities* com o petróleo.

Essa nova interpretação levou à discussão, no presente trabalho, dos modelos anteriores de formação de preços de *commodities* disponíveis na bibliografia e à sugestão de um novo modelo, mais adequado ao contexto atual.

1.2 A Situação Problema e Objetivos de Pesquisa

A formação de preços de ativos é um tema de pesquisa recorrente nos estudos em Administração de Empresas. Os estudos sobre a trajetória dos preços de ativos, em sua maioria, têm o intuito de tentar antecipar informações sobre a formação dos preços, permitindo que os participantes do mercado possam posicionar-se estrategicamente, com o objetivo de maximizarem o valor esperado do fluxo de caixa em relação aos preços do ativo. A previsão dos preços de um determinado ativo depende do estudo de processos estocásticos e de variáveis exógenas, que pode resultar na elaboração de um modelo que descreva a trajetória dos preços e que seja ao mesmo tempo parcimonioso.

Quando estudos sobre formação de preços são aplicados ao mercado de *commodities*, o pesquisador enfrenta um contexto um pouco diferente dos estudos aplicados ao mercado de ativos financeiros. Isso ocorre porque o mercado de *commodities* possui características peculiares, que fazem com que o pesquisador tenha de, além de estudar os processos tradicionais de formação de preços, levar em consideração conceitos característicos do mercado de *commodities*, como a teoria da estocagem, a teoria de paridade de preços, os

preços no mercado de derivativos de *commodities* e as distorções de observação do preço à vista das *commodities* físicas, dentre outras características desse mercado.

A formação dos preços à vista das *commodities* depende da localidade onde estas são negociadas, envolve entrega física e incorpora custos de frete e armazenagem. Os preços à vista de *commodities* publicamente divulgados são médias dos preços negociados em localidades diferentes, podendo haver distorções na observação dessa variável. Por isso, a presente pesquisa utiliza técnicas estatísticas que permitem a estimação dos parâmetros do modelo proposto que descreva a trajetória da variável preço à vista da *commodity*, tratando-a como **variável não observável**.

O **objetivo** da pesquisa é propor um **modelo de formação de preços de *commodities*** que seja aplicável ao mercado de *commodities* agrícolas, levando em consideração, além dos conceitos propostos na literatura, a premissa relevante do atual contexto no qual as *commodities* agrícolas estão inseridas, como **componentes da matriz energética**. O modelo proposto foi aplicado e testado para o mercado brasileiro de açúcar, por ser esta a principal *commodity* agrícola brasileira dentro do conceito de energia renovável.

Para atingir o objetivo principal, a pesquisa foi dividida em quatro partes inter-relacionadas:

- a) **componentes do modelo:** essa seção foi elaborada por meio de pesquisa de fontes de origem secundária, com o intuito de tentar identificar todas as variáveis que possam integrar o modelo pelo estudo de: modelos sugeridos anteriormente por outros autores, características dos mercados à vista e de derivativos de *commodities*, fenômenos decorrentes da teoria da estocagem – como o **retorno de conveniência** e o **backwardation**, processos estocásticos aos quais os preços de *commodities* possam estar sujeitos e processos de estimação de parâmetros adequados a modelos dessa natureza;
- b) **premissas do modelo:** o modelo proposto foi elaborado a partir de premissas que têm base no conceito de *commodity*, como componentes da matriz energética, da influência da volatilidade do mercado na formação dos preços, sazonalidade, dentre outras. As premissas foram verificadas por meio de testes estatísticos de hipóteses, antes da efetiva implementação no modelo;

- c) **especificação do modelo:** nesta parte, o modelo de formação de preços de *commodities* foi especificado em um sistema de equações na forma de espaço de estado, identificando isoladamente os processos estocásticos de cada uma das variáveis que compõem o modelo;
- d) **aplicação do modelo:** o modelo foi aplicado ao mercado de açúcar no Brasil, cujos parâmetros foram estimados por meio do filtro de Kalman, levando em consideração premissas previamente testadas. Além disso, um outro modelo disponível na literatura foi implementado e comparado em termos de capacidade de previsão. Discutidos os resultados, conclusões foram inferidas a partir deles.

1.3 Justificativa da Pesquisa

Os preços das *commodities* têm sido tema relevante de discussão em âmbito mundial, tanto no ambiente acadêmico, quanto nos ambientes político e empresarial. A redução dos estoques globais de *commodities*, provocada pelo crescimento econômico de países como China e Índia, e a volatilidade nos preços do barril do petróleo têm feito com que as fontes de energia renováveis provenientes de *commodities* agrícolas sejam tópico relevante no atual cenário econômico.

A volatilidade nos preços das *commodities*, os biocombustíveis e a possibilidade de falta de alimentos, principalmente para os países pobres, geraram debates em organismos internacionais, como a ONU, Organização das Nações Unidas. Os defensores do uso de biocombustíveis afirmam que os produtos de origem vegetal podem sequestrar carbono da atmosfera e reduzir os efeitos do aquecimento global. Por outro lado, os críticos dos biocombustíveis afirmam que o uso de produtos agrícolas como fonte de energia pode tornar os alimentos menos acessíveis para os países pobres. Além disso, afirmam que a ampliação da área plantada com destino à produção de biocombustíveis pode acelerar o desmatamento de florestas tropicais.

Os defensores da cana-de-açúcar, representados no Brasil pela União da Indústria de Cana-de-açúcar (UNICA, 2008), afirmam que o potencial energético dessa cultura permite, com apenas 1% da área agricultável no Brasil, substituir 50% das necessidades de combustíveis

destinados a veículos leves. Em estudos recentes, a Empresa de Pesquisa Energética (EPE, 2008) apresentou a cana-de-açúcar como a segunda fonte de energia primária no Brasil, medida em termos de toneladas equivalentes de petróleo, superando até mesmo a participação da energia hidráulica. Os produtos derivados da cana, entre eles o etanol e o bagaço, ampliaram, de 14,5% em 2006 para 16% em 2007, sua participação na matriz energética brasileira, segundo dados preliminares da EPE (2008).

Além do uso como biocombustível, as *commodities* têm sido utilizadas, no mercado financeiro, como instrumento de proteção de valor do capital. Em situações de crise e inflação, os investidores buscam alternativas que permitam a manutenção do valor real do capital, dentre elas o mercado futuro de *commodities*. Uma das características das *commodities*, segundo Geman (2005:1), é a manutenção do valor relativo do capital em qualquer moeda. A turbulência provocada pela volatilidade no mercado financeiro leva a uma migração dos fluxos de liquidez para esses mercados.

Dentro desse novo cenário mundial, surge a necessidade do desenvolvimento de um **modelo** de formação de preços de *commodities* que incorpore, além das premissas sugeridas nos trabalhos anteriores, componentes que passaram a interferir no processo de formação dos preços desses ativos.

1.4 Metodologia de Pesquisa

O presente trabalho possui características que Cooper e Schindler (2003:48) definem como **movimento duplo do pensamento científico**, integrando pensamentos indutivos com dedutivos por meio de testes de hipóteses. As hipóteses de pesquisas são apresentadas detalhadamente no quarto capítulo, dedicado exclusivamente à explanação da metodologia adotada na pesquisa.

O pensamento dedutivo consiste em chegar a uma conclusão a partir da verificação de premissas previamente elaboradas. Uma dedução é válida desde que seja impossível que a conclusão seja falsa se as premissas forem verdadeiras. O pensamento indutivo permite ao pesquisador fazer inferências e generalizações sobre o todo a partir da observação de

particularidades. O pensamento científico, segundo Cooper e Schindler (2003:46), é definido da seguinte forma:

Os estilos de pensamento associados ao método científico são geralmente mencionados como meios importantes para assegurar a verdade, embora essa verdade possa não ser permanente nem duradoura.

Em relação à abordagem epistemológica, o método adotado é empírico-analítico. Segundo Martins (1994:26), abordagens empírico-analíticas:

(...) são abordagens que apresentam em comum a utilização de técnicas de coleta, tratamento e análise de dados marcadamente quantitativas. Privilegiam estudos práticos. Suas propostas têm caráter técnico, restaurador e incrementalista. Têm forte preocupação com a relação causal entre as variáveis. A validação da prova científica é buscada através de teste dos instrumentos, graus de significância e sistematização das definições operacionais.

1.5 Questão da Originalidade

Um tema de pesquisa, segundo Martins (1994:20) deve ser **original, importante e viável**. Segundo o autor, em um tema original, existe a possibilidade de encontrar novos resultados, é um trabalho de pesquisa que demonstra avanço na área de estudo a que se dedica.

A abordagem feita neste trabalho considera as *commodities* agrícolas como componente da matriz energética e utiliza esse conceito como premissa fundamental do modelo de formação de preços. Essa é considerada a principal contribuição deste estudo ao conhecimento científico.

Não foi encontrada na bibliografia, tanto nacional como internacional, trabalhos que considerassem explicitamente esse tipo de conceito na formação de preços de *commodities*. A vantagem que este estudo tem em relação aos anteriores é o fato de os biocombustíveis, apesar de serem explorados há muitos anos como fonte alternativa de energia, terem se tornado apenas recentemente tema internacionalmente relevante na substituição dos derivados de petróleo e da relação deste com a formação global nos preços de *commodities*.

As principais pesquisas desenvolvidas sobre esse assunto estão predominantemente concentradas, em termos relativos de importância por quantidade de citações, em

pesquisadores relacionados com o Professor Eduardo Schwartz, na Universidade da Califórnia. Além desse núcleo de pesquisas, a Professora Hélyette Geman, da cadeira de Finanças da Universidade de Paris, e os pesquisadores relacionados a ela também se tornaram uma importante fonte de trabalhos acadêmicos sobre o assunto.

No Brasil, os principais estudos acadêmicos de formação de preços de *commodities* são gerados por pesquisadores da ESALQ-USP, Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo. Alguns outros trabalhos relacionados ao tema foram escritos por acadêmicos de Economia na PUC-RJ, Pontifícia Universidade Católica do Rio de Janeiro, mas apenas questionando os processos de estimação aplicados aos modelos preexistentes na literatura. Não foram encontrados trabalhos nacionais com abordagens similares a esta.

Além da questão sobre a matriz energética, não foram encontrados trabalhos que considerassem, como variável de entrada do modelo, a volatilidade do mercado de ações para identificar o efeito das turbulências do mercado financeiro na trajetória de preços de *commodities*.

O presente estudo, além disso, apresenta o modelo em forma de equações de espaço de estado, utilizando o processo de estimação de parâmetros do filtro de Kalman. O filtro de Kalman, ferramenta já conhecida na estimação de modelos de *commodities*, é mais abrangente em termos de resultados do que os modelos autorregressivos mais difundidos, como os modelos vetoriais e ARIMA.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Introdução

Os modelos de formação de preços de *commodities* envolvem múltiplos conceitos teóricos. É difícil, para o pesquisador, estudar esse mercado de forma específica, abordando apenas um ou outro conceito isoladamente, sem correr o risco de ser omissivo. Por isso, tentou-se aqui tratar o tema com uma abrangência mínima necessária, desenvolvendo os conceitos imprescindíveis com maior profundidade.

A fundamentação teórica foi organizada em quatro grupos complementares. Apesar de serem tratados separadamente, os conceitos teóricos possuem igual importância no desenvolvimento do **modelo de formação de preços de *commodities***, objetivo central do presente estudo. O trabalho de revisão da teoria foi dividido conforme a seguinte estrutura:

- a) **conceitos fundamentais**: nessa parte procurou-se abordar as características dos mercados de *commodities* e os papéis de seus participantes;
- b) **teoria da estocagem**: foram abordadas as regras que regem a estrutura de preços peculiar aos mercados à vista e futuro de *commodities*;
- c) **processos estocásticos**: nessa etapa foram abordados os modelos que descrevem a trajetória dos preços no tempo;
- d) **modelagem dos preços**: nessa etapa foi apresentada a evolução dos modelos de formação de preços de *commodities*, desde os trabalhos seminais, até as recentes abordagens sobre o assunto.

2.2 Conceitos Fundamentais do Mercado de *Commodities*

2.2.1 Definição de *Commodity*

De acordo com Kaldor (1939:3), Copeland e Weston (1988:302) e Geman (2005:1), *commodity* pode ser definida como um ativo físico que possui características **padronizadas**, de ampla negociação em diversas localidades, que pode ser transportado e **armazenado** por um longo período de tempo.

Commodity ainda pode ser definida como um tipo de produto no qual não há diferenças qualitativas entre os mercados onde é negociado, ou seja, entre negócios de um mesmo produto em mercados diferentes, não existe **preferência**, em termos de qualidade, por parte dos compradores do produto. Copeland e Weston (1988:302), porém, afirmam que pode haver pequenas diferenças de qualidade entre cada lote negociado, essas diferenças devem estar dentro de limites aceitáveis e previamente especificados em contratos de fornecimento.

Ainda sobre a definição da terminologia – de acordo com Geman (2005:1), o termo *commodity* pode ser atribuído a um bem de consumo cuja escassez, na forma de exaustão na extração, ou na redução de estoques globais, causará um impacto no preço em âmbito mundial. A autora define ainda que, dadas as volatilidades das diferentes moedas no mundo, uma *commodity* terá o mesmo valor em termos relativos, nas diferentes moedas, podendo ser utilizada como referência de valor.

É comum que a terminologia *commodity* seja atribuída aos insumos, ou matérias-primas. Isso ocorre porque os insumos ainda não foram industrialmente transformados, o que facilita a padronização. Quando um produto sofre transformação na indústria, ganha características particulares que o distinguem, dificultando a padronização e a negociação em larga escala.

Copeland e Weston (1988:302) sugerem que *commodity* é um ativo cuja padronização permite a execução de maior número de negociações, o que favorece a liquidez. Em mercados organizados, como as bolsas de valores, é importante que as modalidades de ativos

negociadas permitam a presença de liquidez, que, nesse caso, pode ser entendida como a facilidade de entrar e sair de uma posição, comprada ou vendida, com rapidez e facilidade. Se um determinado agente do mercado está em uma posição comprada em determinado ativo, ele conseguirá, em um mercado líquido, sair da posição rapidamente, vendendo seu ativo. Uma condição necessária para a presença de liquidez em mercados organizados é a padronização, que permite agilidade nas negociações, partindo-se da premissa de que as contrapartes envolvidas na operação saibam previamente as características da mercadoria negociada.

A *commodity*, para permitir liquidez nas negociações em um mercado organizado, deve possuir simultaneamente, segundo Kaldor (1939:3), uma série de atributos, descritos a seguir:

- a) o ativo deve ser totalmente padronizado em suas características;
- b) deve ser um bem de demanda generalizada;
- c) deve ser um bem não perecível, ou seja, o ativo não deve perder valor com a mera passagem do tempo;
- d) o valor do ativo deve ser proporcional ao volume.

Kaldor (1939:3) salienta que há apenas duas classes de ativos que satisfazem a condição necessária para negociação em larga escala. A primeira classe são as *commodities*, negociadas em mercados organizados; a segunda classe de ativos são os financeiros, como títulos e ações, que possuem, em grau máximo, todos os atributos acima descritos.

As *commodities* podem ser negociadas em diferentes tipos de mercados: no mercado à vista, que envolve entrega física imediata da mercadoria, ou ser negociada como referência nos mercados derivativos. Na seção seguinte é feita uma breve descrição dos tipos de mercados onde as *commodities* são negociadas.

2.2.2 Mercado à Vista, Termo e Futuro

Um produto físico padronizado deve sempre ser negociado no mercado à vista, segundo Geman (2005:1). O **mercado à vista** é a negociação de compra e venda de um determinado ativo com entrega imediata. No caso dos ativos financeiros, esse processo ocorre de forma simples e ágil, pois eles são escriturais e a liquidação ocorre em curtíssimos espaços de tempo.

Com *commodities*, a negociação à vista é um pouco menos trivial. A liquidação de uma operação à vista com *commodities* envolve a entrega física do produto, documentação para transporte, contratação de frete, logística na estocagem, seguros e, no caso de exportação, procedimentos aduaneiros. Além disso, para que a entrega física da mercadoria ocorra, é necessário que o produtor tenha disponibilidade de estoque, no volume, no local e na qualidade exigidas pelo comprador. Por esses motivos, na negociação à vista no mercado de *commodities*, sempre haverá uma defasagem mínima entre o momento de negociação e a entrega efetiva da mercadoria.

Geman (2005:3) explica que os custos de frete, de logística e de seguros, podem ser assumidos pelo comprador ou pelo vendedor. A venda é classificada como FOB (*Free on Board*) se o comprador arcar com as despesas de frete e seguro. Se o custo da entrega for responsabilidade do vendedor, é classificada como CIF (*Cost, Insurance and Freight*). Essas terminologias, comuns no comércio internacional, também são utilizadas no mercado doméstico de *commodities* físicas.

A entrega da *commodity*, como explicado anteriormente, depende da disponibilidade dos estoques e de uma relação de fornecimento entre o produtor vendedor e a indústria compradora da mercadoria. O fornecimento de *commodity* é a etapa inicial na cadeia produtiva e a escassez de estoques pode provocar a interrupção desse processo.

Durante a negociação, o comprador da indústria sempre deseja pagar o menor preço possível, para ampliar suas margens, em contrapartida, depende do fornecimento contínuo para manter sua atividade. O produtor vendedor, por outro lado, deseja obter o maior preço possível por sua mercadoria. Se o mercado fornecedor em questão não for monopolista e se a *commodity* for armazenável, o produtor pode aguardar para obter o melhor preço de venda, retendo o

fornecimento. Essa estratégia será eficiente apenas se o incremento nos preços for superior ao custo marginal de armazenagem. Pindyck e Rubinfeld (1994:337) explicam que o poder monopolista ocorre apenas quando o vendedor tem a habilidade de incrementar os preços acima de seu custo marginal. Em um mercado competitivo de *commodity*, isso não ocorrerá se o comprador tiver a opção de comprar de outros fornecedores. Em consequência do descompasso nas necessidades entre compradores e vendedores, o mercado à vista de *commodities* favorece a presença de intermediários, ou *tradings*, que facilitam a negociação e a entrega física, entre a indústria e os produtores de *commodities*, evitando a interrupção no fornecimento.

As *commodities* negociáveis podem ser provenientes do **agronegócio**, como é o caso da soja, milho, algodão, açúcar e álcool, ou provenientes da atividade de **extração**, como no caso do petróleo, gás, minério de ferro, metais não ferrosos e minerais não metálicos. O que diferencia o mercado extrativista, das *commodities* provenientes do agronegócio, é o fato de estas possuírem **sazonalidade**, devido aos períodos de safra e entressafra. A sazonalidade na produção das *commodities* agrícolas dificulta a adequação temporal do fornecimento do agronegócio às necessidades da demanda na indústria.

Marques e Mello (1999:7) sugerem que essa relação de interdependência entre os fornecedores e compradores de *commodities* físicas favorece a adoção de **contratos de fornecimento** entre os membros da cadeia produtiva, o que representa uma vantagem não só para o comprador, pois mantém a continuidade no fornecimento, mas também para o vendedor, pois garante a venda da mercadoria.

Os contratos de fornecimento são comuns em negociações com *commodities* de agronegócio. Eles estabelecem a data, as características da mercadoria e o volume negociado. No contrato também pode ser estipulado antecipadamente o preço que será pago pela mercadoria na entrega futura, definido como **preço a termo**. O contrato de fornecimento é o primeiro passo para a introdução dos participantes do mercado de *commodities* nos contratos derivativos em sua modalidade mais simples, que é o **contrato a termo**.

Um contrato a termo, segundo Hull (2006:3), “é um acordo de compra ou venda de um ativo em uma data determinada por um preço determinado”. Comparado ao contrato *spot*, a diferença está no prazo de entrega do ativo físico, que ocorre em data futura.

O contrato a termo é tipicamente negociado no **mercado de balcão**, também conhecido como OTC (*Over the Counter*). Nesse mercado, os contratos a termo são negociados diretamente entre as partes, as operações não são padronizadas e as contrapartes podem assumir o risco de crédito da operação ou, alternativamente, contratar um seguro de crédito ou os serviços de uma câmara de liquidação.

Em um contrato a termo há duas partes envolvidas: o comprador, que assume o compromisso de comprar o ativo físico, com entrega em data futura e pagar por ele um preço preestabelecido; e o vendedor, que se compromete em entregar a mercadoria na data estabelecida.

No contrato a termo, as contrapartes assumem o **risco de liquidação**, que é dividido em liquidação física e liquidação financeira. A liquidação física consiste na entrega da mercadoria e a financeira, no pagamento pela mercadoria entregue. Em ambos os casos, pode ocorrer inadimplência, tanto na liquidação física, quanto na financeira.

No contrato a termo, a liquidação pode ser física ou por **ajuste por diferença**. Na liquidação do tipo ajuste por diferença, não há entrega física da mercadoria, apenas liquidação financeira, na qual as contrapartes acordam em pagar a diferença entre o preço a termo e o preço de mercado.

Algumas instituições divulgam o preço de mercado que serve como referência para as liquidações de contratos por diferença. No Brasil, entre as instituições mais referenciadas está o **CEPEA**, Centro de Estudos Avançados em Economia Aplicada – entidade criada por docentes da ESALQ-USP.

O Quadro 1, a seguir, mostra algumas das mercadorias cujos preços são divulgados pelo CEPEA:

Quadro 1 - Indicadores de preços divulgados pelo CEPEA em 20/3/2008

Mercadoria	Volume	Valor de referência em R\$	Periodicidade
Açúcar Cristal	Saco de 50kg	27,68	Diária
Álcool Anidro	Litro	0,83386	Semanal
Algodão	Arroba	47,60613	Diária
Café Arábica	Saca de 60kg	242,73	Diária
Milho	Saca de 60kg	26,89	Diária
Soja	Saca de 60kg	42,04	Diária

FONTE: CEPEA (2008).

Mais padronizado do que o mercado a termo, o **mercado futuro** está presente em mercados organizados de bolsas. Os contratos futuros possuem, em sua maioria, um maior volume de negociação do que o contrato a termo, portanto são mais líquidos, mais transparentes para as partes envolvidas e publicamente divulgados, servindo como referência na formação de preços.

O **contrato futuro** é bastante similar ao termo. É também um acordo de compra e venda de determinado ativo, com liquidação em data preestabelecida e a preço predeterminado. A liquidação do contrato futuro também pode ser física ou por diferença. Conforme Hull (2006:6), nas negociações do mercado futuro, as contrapartes não se conhecem, são anônimas, porém as operações são publicamente divulgadas. Os contratos futuros são tipicamente negociados em bolsa. No Brasil, a bolsa de maior expressão na negociação de contratos futuros é a **BM&FBOVESPA, Bolsa de Valores, Mercadorias e Futuros**. Schouchana e Miceli (2004:10) definem as características do contrato futuro da seguinte maneira:

Nos contratos futuros, constam especificações de qualidade dos produtos negociados, cotação, variação mínima de apregoação, oscilação máxima diária, unidade de negociação, meses de vencimento, data de vencimento, local de formação do preço e de entrega da mercadoria, período e procedimentos de entrega e retirada da mercadoria, liquidação financeira, arbitramento, ativos aceitos como margens de garantia e custos operacionais. Os contratos futuros são padronizados, de modo que, no pregão, sejam negociados o preço e a quantidade de contratos, uma vez que todos se referem ao mesmo produto, mesmo local de entrega e mesma quantidade por contrato.

Os contratos futuros são padronizados nas bolsas em termos de: prazos, volumes, características do ativo, cotação em bolsa, ajustes diários e margens de garantia. A câmara de liquidação, para garantir a efetiva liquidação das operações, exige depósitos de garantia dos participantes do mercado. No caso de qualquer uma das partes não honrar com suas obrigações em relação ao contrato futuro, a câmara de liquidação pode, arbitrariamente, executar as garantias depositadas e liquidar a posição do participante inadimplente.

Para reduzir os valores de depósito de margem, as bolsas de valores adotaram o mecanismo de **ajustes diários**. Eles consistem nas liquidações financeiras das posições pelo método de ajuste por diferença, ou seja, a câmara de liquidação, com base nos **preços de ajuste** divulgados pela bolsa, realiza o fluxo financeiro das posições dos participantes, em relação às oscilações de preços dos contratos futuros de um dia para o outro.

De acordo com Schouchana e Miceli (2004:10), com a utilização do mecanismo de ajuste diário, a margem de garantia requerida para cobrir a liquidação financeira de um dia fica muito menor do que se fosse exigida para cobrir o equivalente ao período integral do contrato.

A título de ilustração, na Tabela 1, a seguir, estão os preços de ajuste dos contratos futuros de açúcar. Na coluna à esquerda da tabela estão os meses de vencimento dos contratos. A segunda coluna apresenta o volume de contratos em aberto, nos quais os participantes ficaram posicionados de um dia para o outro. Em seguida, o número de negócios ocorridos no dia, os contratos negociados, o volume financeiro em reais e, na última coluna à direita, o preço de ajuste divulgado no resumo estatístico da BM&FBOVESPA do pregão referente ao dia 20/3/2008.

Tabela 1 - Resumo estatístico da BM&FBOVESPA do futuro de açúcar ref. 20/3/2008

Vencimento	Contratos em aberto	Número de negócios	Contratos negociados	Volume financeiro (R\$)	Preço de ajuste (US\$)
Abr./2008	687	8	71	477.709	14,52
Jul./2008	150	1	10	66.439	14,50
Set./2008	150	2	2	13.756	14,70

FONTE: BM&FBOVESPA (2008).

Nos mercados à vista, termo e futuro, os participantes podem atuar com diferentes intuitos. Na seção seguinte, é feita uma breve explanação dos papéis dos participantes em relação a sua atuação nos mercados de derivativos de *commodities*.

2.2.3 *Hedge, Especulação e Arbitragem*

O mercado de derivativos de *commodities* permite que seus participantes atuem com objetivos distintos. Os produtores de mercadoria física podem utilizar o mercado para proteger-se contra oscilações do preço de venda, enquanto a indústria compradora pode utilizar o mercado para fixar o preço de seus insumos. Outros participantes, especialistas na formação de preços, podem utilizar o mercado com o intuito de fazer lucros com as oscilações nos preços.

Os participantes que atuam no mercado de derivativos com o intuito de reduzir seus riscos, procurando uma **proteção** contra a volatilidade dos preços, são denominados *hedgers*. Johnson (1960:140) descreve o *hedger* como “(...) um investidor na *commodity* real que deseja um seguro contra os riscos de preço”. Normalmente é o produtor do ativo físico, ou o comprador da indústria de transformação.

O *hedger* pode proteger o valor de seu estoque de *commodity* realizando, simultaneamente uma venda na mesma quantidade em contratos futuros. Se as variações no preço da *commodity* a vista forem iguais as do contrato futuro, ou seja, se os movimentos de preço forem paralelos, uma perda com os preços à vista pode ser compensada por um ganho de igual montante nos contratos futuros.

Um outro tipo de participante do mercado de derivativos é o **especulador**. Normalmente, ele não é o produtor do bem físico nem deseja a entrega física da mercadoria, apenas o ajuste financeiro por diferença. O especulador é um especialista na formação de preços de *commodities* e atua no mercado com o objetivo de fazer lucro com as oscilações de preços. Seu papel é importante porque proporciona **liquidez** ao mercado futuro, permitindo que outros participantes possam assumir posições inversas em relação às dos especuladores.

O papel do especulador, ainda segundo Johnson (1960:140), é o de assumir o risco transferido pelo *hedger* com o intuito de auferir algum tipo de ganho. O mercado futuro permite a

transferência de risco de um grupo para outro. O *hedger* estaria disposto, nesse caso, a pagar um prêmio pelo risco transferido ao especulador.

Kaldor (1939:1), em trabalho seminal, analisa o papel do especulador na formação de preços no mercado de ativos físicos. Define a especulação como “(...) a compra (ou venda) de bens com o intuito de revender (recomprar) em uma data posterior (...)”, ação motivada pela expectativa de uma mudança nos preços. Para obter ganhos anormais, o especulador, segundo Kaldor (1939:2), deve ter uma previsão da demanda mais precisa do que o investidor regular. Porém afirma o autor que, em um mundo de previsões perfeitas, não poderia haver ganho especulativo.

O terceiro participante do mercado de derivativos é o **arbitrador** – aquele que atua no mercado de derivativos com o intuito de fazer lucro certo sem assumir riscos, simplesmente aproveitando a diferença de preços de um mesmo ativo que é negociado em mercados diferentes. O arbitrador compra e vende simultaneamente, não fica com o ativo e, conseqüentemente, não precisa dispor de recursos financeiros, seu papel no mercado é atuar como regulador de preços e manter o **equilíbrio** dos mercados.

Segundo Working (1948:5), a arbitragem permite que o preço futuro esteja em equilíbrio com o preço à vista mais uma diferença, decorrente do custo de carregamento e da taxa de juros livre de risco. O custo de carregamento e a diferença entre os preços futuros e o mercado à vista serão abordados posteriormente, ainda neste capítulo.

No mercado futuro, é importante salientar também o papel do **intermediário financeiro**, que viabiliza a execução da operação em bolsa, normalmente na figura das **CTVM, Corretoras de Títulos e Valores Mobiliários**. Esses participantes podem atuar como formadores de opinião para seus clientes, recebendo comissão pelas transações, em forma de corretagem.

2.2.4 Principais Mercados Futuros

A estrutura moderna de negociação em mercados futuros, de acordo com Geman (2005:9), surgiu nos Estados Unidos, na segunda metade do século XIX. Em 1848 foi fundada a *Chicago Board of Trade* (CBOT, 2008), que desde o início operava *commodities* agrícolas.

Atualmente, existem diversas bolsas de negociação de *commodities* ao redor do mundo e mais de 600 milhões de contratos futuros são negociados anualmente. No Quadro 2, a seguir, estão listadas as principais bolsas de negociação de contratos futuros de *commodities*:

Quadro 2 - Bolsas de negociação de *commodities*

Commodities Negociadas	Bolsas	Siglas
Açúcar, etanol, boi gordo, café arábica, milho, soja	Bolsa de Valores, Mercadoria e Futuros	BM&FBOVESPA
Açúcar, cacau, café e algodão	New York Mercantile Exchange	NYMEX
Açúcar, cacau, café robusta e batata	London International Financial Futures Exchange	LIFFE
Açúcar, cacau e café	Coffe, Sugar and Cocoa Exchange	CSCE
Petróleo, gasolina e gás propano	New York Mercantile Exchange	NYMEX
	International Petrol Exchange (Londres)	IPE
	Hong Kong Futures Exchange	HKEX
	Philadelphia Board of Trade	PBOT
Trigo, milho, soja, prata e ouro	Chicago Board of Trade	CBOT
Toucinho de porco, celulose, ouro e clima	Chicago Mercantile Exchange	CME
Algodão e arroz	Chicago Rice and Cotton Exchange	CRCE
Trigo, milho e suco de laranja	Minneapolis Grain Exchange	MGE
	New York Cotton Exchange	NYCE
	New York Mercantile Exchange	NYMEX
Metais	New York Mercantile Exchange	NYMEX
	London Metal Exchange	LME
	Commodity Exchange	COMEX
Eletricidade	Nordic Power Exchange	NORDPOOL
	European Energy Exchange	EEX
	Amsterdam Power Exchange	APX
	Paris Power Exchange	POWERNEXT

FONTE: Adaptado de Geman (2005:12).

O modelo de formação de preços de *commodity* proposto neste estudo é aplicado ao mercado de açúcar e álcool, por essa razão, na seção seguinte, são apresentadas as características dos contratos futuros dessas *commodities* e as bolsas onde são negociadas.

2.2.5 Mercados de Açúcar e Álcool

Os principais mercados de negociação de contratos futuros de açúcar estão localizados em Nova York, na **NYMEX**, e em Londres, na **LIFFE**. Os contratos futuros de açúcar possuem liquidez também em São Paulo, na **BM&FBOVESPA**. As características dos contratos futuros negociados nas bolsas são as seguintes:

- a) **NYMEX**: o contrato de açúcar n. 11, de acordo com a NYMEX (2008), foi o primeiro contrato de *commodity* agrícola negociado em bolsa, desde 1936. O tamanho do contrato é de 112 mil libras-peso, o equivalente a 50 toneladas, a cotação é feita em centavos de dólar por libra-peso. Atualmente o preço está em torno de 13 centavos de dólar por libra-peso. O tipo de liquidação é financeiro, através de ajuste por diferença, sem entrega física. Os vencimentos dos contratos ocorrem nos meses de março, maio, julho e outubro;
- b) **LIFFE**: são negociados dois tipos de contrato, conforme especificações da LIFFE (2008): o contrato de açúcar bruto e o açúcar branco – *raw sugar* e *white sugar*. O açúcar cristal branco n. 407, o mais negociado, possui as mesmas características do contrato da NYMEX, com 50 toneladas por contrato e cotado em centavos de dólar por libra-peso, com vencimentos nos mesmo meses da NYMEX. A diferença é que o contrato da LIFFE permite a entrega física FOB, em locais designados pela bolsa e dentro das especificações por ela determinadas;
- c) **BM&FBOVESPA**: o contrato futuro de açúcar negociado conforme as regras da BM&FBOVESPA (2008) é o açúcar cristal especial com entrega no município de São Paulo. O tamanho do contrato é de 270 sacas de 50 quilos e cotado em centavos de dólar por saca. Os vencimentos ocorrem nos meses de fevereiro, abril, julho, setembro e novembro. A liquidação do contrato pode ser financeira ou física. A entrega física deve ser solicitada previamente pelo vendedor, realizada em locais determinados pela

bolsa, no Estado de São Paulo, e dentro das características especificadas no contrato. Para isso, a bolsa exige do vendedor um certificado de Análise e Classificação, válido por 30 dias.

O contrato futuro de álcool anidro carburante, ou simplesmente **etanol**, é negociado na BM&FBOVESPA, cotado em reais por metro cúbico. O tamanho do contrato é de 30 metros cúbicos, com vencimentos em todos os meses do ano.

O município de Paulínia é o local de referência para a formação de preços do contrato futuro de álcool. No vencimento do contrato, a BM&FBOVESPA permite a liquidação com entrega física, mediante solicitação pelo vendedor e respeitando as especificações determinadas pela bolsa.

O contrato de etanol ainda possui uma tímida negociação em bolsa, com um volume médio de 150 contratos por semana, de acordo com dados da BM&FBOVESPA (2008).

Conforme visto anteriormente, os preços à vista, que servem como referência para os ajustes financeiros nos vencimentos dos contratos futuros de açúcar da BM&FBOVESPA, são divulgados em base diária pelo CEPEA (2008). Os preços do etanol são divulgados, também pelo CEPEA, porém em base semanal. Os preços do açúcar, em saca de 50 quilos, são coletados das negociações sem frete entre usinas e compradores, referentes a negócios efetuados nas regiões de Piracicaba, Ribeirão Preto, Jaú e Assis.

Devido à importância do etanol como fonte renovável de energia e como alternativa para substituição de combustíveis fósseis, foram levadas em consideração, neste estudo, as cotações do petróleo, como variável de entrada no modelo de formação de preços proposto. Por essa razão, na seção seguinte, são apresentados os principais mercados futuros de negociação de petróleo.

2.2.6 Mercados de Petróleo

O petróleo é um tipo de *commodity* não padronizada, por isso os principais contratos futuros no mundo utilizam, segundo Geman (2005:201), um pequeno número de tipos de petróleo

como referência de produto físico para as negociações dos contratos futuros. As principais bolsas onde os contratos futuros de petróleo são negociados são a **NYMEX**, em Nova York, e a **IPE**, em Londres. O primeiro vencimento do contrato futuro de petróleo negociado em Nova York é referência mundial de formação de preços dessa *commodity*.

As características dos contratos futuros de petróleo negociados em bolsa são as seguintes:

- a) **NYMEX:** o futuro de petróleo tipo WTI (*West Texas Intermediate*) negociado na NYMEX é o contrato futuro sobre *commodity* mais negociado no mundo. Quase trezentos mil contratos de petróleo são negociados diariamente. O tamanho do contrato é de mil barris de petróleo, podendo ser liquidado no vencimento por entrega física, cujo local de referência é Cushing, no estado de Oklahoma. O contrato, também denominado de *light, sweet crude oil*, possui baixo teor de enxofre, portanto, pode ter um refino mais fácil para obtenção de derivados, (NYMEX, 2008);
- b) **IPE:** a *International Petroleum Exchange*, em Londres, é a bolsa de negociação de petróleo tipo *Brent*, também conhecido como *Brent blend* ou *Brent* de Londres. Mas desde 2005 tem sido negociado eletronicamente na ICE (*Intercontinental Exchange*). O petróleo tipo *Brent* não é tão leve quanto o WTI, possui mais enxofre, ideal para a produção de gasolina e outros derivados intermediários, como querosene e diesel. O preço do barril de *Brent* está em torno de um dólar abaixo do WTI. A liquidação do contrato futuro, segundo a ICE (2008), pode ser física ou financeira. A cotação do contrato é feita em dólares por barril e o lote de negociação é de mil barris de petróleo.

Nas seções anteriores foram apresentados os conceitos fundamentais sobre as características dos contratos de *commodities* e seus principais mercados de negociação. Nas seções seguintes serão apresentadas algumas das abordagens teóricas que discutem o processo de formação de preços entre os mercados à vista e o mercado futuro de *commodities*.

Os preços formados nos mercados à vista e futuro deveriam, supostamente, seguir trajetórias paralelas. Uma queda no preço à vista deveria ser acompanhada por uma queda proporcional no preço do contrato futuro, devido aos princípios de arbitragem que regem a formação de preços futuros.

No mercado de *commodities*, o processo de formação de preços futuros pode romper as regras de arbitragem, causando fenômenos denominados como *contango* e *backwardation*. Na seção seguinte será discutida uma das teorias que tentam explicar os fenômenos de formação de preços das *commodities* em relação aos contratos futuros, denominada **teoria da estocagem**.

2.3 Teoria da Estocagem

A Teoria da Estocagem procura explicar a diferença de preços entre os mercados à vista e futuro de *commodities*. Em certas ocasiões, o preço à vista da mercadoria pode ficar acima do preço no mercado futuro. Isso ocorre porque o detentor da mercadoria física possui o benefício de reter os estoques e, em busca de melhores preços, aguardar um melhor momento para realizar a venda. O detentor do estoque, ao tomar essa decisão, deve levar em consideração algumas variáveis relevantes, tais como as taxas de juros, a sazonalidade nos preços, o custo de armazenagem e o benefício esperado com o aumento do preço de venda.

Nesta seção serão apresentados os principais trabalhos na literatura que tratam da Teoria da Estocagem.

2.3.1 Princípios da Teoria da Estocagem

A teoria da estocagem foi abordada, inicialmente, nos trabalhos seminais de Kaldor (1939:3), Working (1949:1257), Telser (1958:235) e Brennan (1958:53).

A premissa dessa teoria é a de que o detentor da mercadoria física tem o benefício de reter o produto quando o preço de mercado estiver baixo e liberá-lo quando o preço à vista atingir patamares mais favoráveis. Essa estratégia somente é benéfica para o detentor da *commodity* se o benefício marginal do aumento de preços exceder o custo de estocagem.

Existe uma relação entre o aumento dos preços da *commodity* e o volume global de estoques. Quando os estoques de uma determinada *commodity* são excepcionalmente grandes, não há

vantagem em reter o produto, pois não haverá escassez no fornecimento e o custo de estocagem, nesse caso, será maior do que o benefício em termos de preço.

Por outro lado, se os estoques são moderados e há alguma probabilidade de escassez, a competitividade da demanda pelo produto pode render ao detentor do estoque físico algum tipo de retorno. Nesse caso, o benefício da estocagem será maior do que o custo de armazenagem.

Esse comportamento na formação dos preços à vista da *commodity* interfere na relação de preços com o mercado futuro. Em certas ocasiões, o preço à vista pode ficar acima do preço no mercado futuro, isso ocorre quando o preço se eleva, devido à escassez de estoques. Para quantificar a diferença entre esses preços, Working (1949:1258) sugere uma variável que definiu como *price of storage*.

Preço da estocagem, ou *price of storage*, é uma variável calculada pela diferença entre o preço futuro e o preço à vista, ou a diferença entre dois contratos futuros com vencimentos distintos. Pode ser um valor positivo ou negativo. O preço de estocagem negativo ocorre quando existe a probabilidade de escassez no fornecimento da *commodity*. O preço da estocagem, representado por $F_t - S_t$ na Equação 2.1, deve ser igual aos juros no período $S_t r$ mais o custo marginal de armazenagem, menos o ganho marginal de conveniência para uma unidade adicional no estoque, segundo Fama e French (1987:56):

$$F_t - S_t = S_t r + W_t - C_t \quad (2.1)$$

sendo:

S_t : preço do ativo à vista no tempo t ;

F_t : preço do contrato futuro no tempo t , com vencimento em T , posterior a t ;

r : taxa de juros livre de risco;

W_t : custo marginal de armazenagem no tempo t ;

C_t : valor do ganho marginal de conveniência no tempo t .

O ganho marginal pela conveniência, na equação acima, é um valor financeiro que mede o ganho obtido pelo detentor da mercadoria física com o aumento do preço da *commodity*, em um momento em que os estoques globais estão baixos. Nessa relação, o proprietário do estoque tem custos de taxa de juros livre de risco e de armazenagem, porém, tem um ganho de conveniência compensatório, devido aos melhores preços de venda.

Quanto maior o ganho de conveniência de uma determinada *commodity*, mantidos constantes as taxas de juros e custos de armazenagem, maior será o preço à vista, em relação ao preço do contrato futuro. Se o ganho de conveniência for maior do que os juros e o custo de armazenagem, $C_t > S_t r + W_t$, o preço à vista da *commodity* será maior do que o preço futuro, resultando no fenômeno denominado de *backwardation*. Por outro lado, se $C_t < S_t r + W_t$, o preço futuro será maior do que o à vista, resultando no fenômeno denominado *contango*. Os termos *contango* e *backwardation* são abordados com mais detalhes em seção posterior.

2.3.2 Custo de Carregamento

O custo de carregamento é a soma de todos os custos para manutenção dos estoques da mercadoria física, é uma variável que determina a diferença de preços entre o mercado à vista e o mercado futuro da *commodity*. O custo de carregamento é calculado a partir do custo de financiamento – representado pela taxa livre de risco, custo de armazenagem, custo de depreciação primária e retorno de conveniência. O custo de depreciação primária é a perda de valor do ativo com a mera passagem do tempo. Por exemplo, quando a mercadoria, com o tempo, se estraga no armazém no qual foi estocada.

Working (1948:1) define o custo de carregamento como uma medida da “(...) diferença, em um dado momento, entre os preços de uma *commodity* para duas diferentes datas de entrega”. O custo de carregamento deve ser uma variável que mede a relação entre os preços de mercadorias que tenham essencialmente a mesma qualidade, entregues no mesmo local, porém em diferentes datas.

Eventualmente, se os preços no mercado futuro estiverem excessivamente elevados, um agente do mercado pode, para obter ganho com a diferença de preços, comprar mercadoria física e simultaneamente vendê-la no mercado futuro. Nessa operação de arbitragem, o agente deve carregar a mercadoria física até o vencimento do contrato futuro, incorrendo no custo de carregamento. Essa operação é conhecida, segundo Geman (2005:37), como *cash and carry*. Se o inverso acontecer, ou seja, o preço futuro estiver relativamente baixo, o agente pode vender o ativo à vista e simultaneamente comprá-lo no mercado futuro, realizando a operação denominada *reverse cash and carry*. Quando o mercado está em **condição de não arbitragem**, o agente de mercado não tem possibilidade de obter lucros sem assumir riscos. Uma operação *reverse cash and carry* pode ser feita em quatro etapas simultâneas:

- i) o arbitrador toma o ativo à vista por arrendamento, pagando o retorno de conveniência δ_t ;
- ii) realiza a venda do ativo à vista pelo preço S_t ;
- iii) com os recursos provenientes da venda, aplica na taxa do ativo livre de risco r ;
- iv) para proteger-se contra oscilações do preço, adquire contratos futuros pelo preço F_t .

Dessa forma o resultado da operação, em condição de não arbitragem, é:

$$S_t e^{r(T-t)} = F_t e^{\delta_t(T-t)} \quad (2.2)$$

De acordo com a Equação 2.2 acima, a diferença entre o preço futuro e à vista da mercadoria é determinada pela taxa livre de risco, pelo retorno de conveniência e pelo prazo $T - t$ até o vencimento do contrato futuro.

Em alguns trabalhos disponíveis na literatura, o custo de carregamento foi generalizado, e a variável δ_t foi definida apenas como retorno de conveniência, ou seja, os custos de armazenagem e depreciação primária foram omitidos na equação. Nas seções posteriores as outras variáveis referentes à Teoria da Estocagem serão discutidas.

2.3.3 Contango e Backwardation

Working (1948:3) analisou as diferenças de preços à vista e futuro no mercado de trigo nos Estados Unidos, nas décadas de 1930 e 1940. Nos dados observados, identificou dois fenômenos, que definiu como *contango* e *backwardation*. O primeiro ocorre quando os preços à vista estão abaixo dos preços futuros; o segundo, quando o preço à vista da *commodity* está acima do preço futuro.

Backwardation é um valor, medido em termos financeiros, da diferença entre o preço à vista e o preço no mercado futuro de determinada *commodity*. Logo, ocorre se $S_t - F_t$ for um número positivo. É conveniente ressaltar que, mesmo na presença de *backwardation*, os preços à vista e futuro são intimamente relacionados. O preço à vista reflete todas as alterações ocorridas no preço futuro, ou seja, existe certo paralelismo na formação de preços.

Na literatura, autores como: Johnson (1960:40), Benirschka e Binkley (1995:515), Litzenger e Rabinowitz (1995:1517) e Frechette e Fackler (1999:761) justificaram a existência de *backwardation* nos preços das *commodities* com base nas seguintes teorias:

- a) a primeira teoria define o fenômeno como um prêmio pelo risco que o especulador merece por assumir o risco do *hedger*, que assume uma posição predominantemente vendida no mercado futuro, no caso o produtor que deseja garantir o preço de venda de sua *commodity*;
- b) a segunda teoria justifica existência de *backwardation* como resultante do retorno de conveniência;
- c) uma terceira teoria explica o fenômeno de *backwardation* como resultante de erros de medida ou variável não observável;
- d) e, ainda, uma quarta teoria procura defender o argumento de que o *backwardation* é causado pela distância dos estoques dos produtores em relação ao centro de consumo.

A seguir, essas teorias serão brevemente discutidas.

A primeira teoria tem base no trabalho de Johnson (1960:140), que se referiu à presença de *backwardation* no mercado de *commodities* como consequência das operações de *hedge*. Os produtores de *commodities*, que usualmente assumem posição vendida no mercado futuro, pagam um prêmio pelo risco assumido pelos especuladores, fazendo, dessa forma, com que os preços futuros fiquem abaixo dos preços à vista.

A segunda teoria sugere que a existência do *backwardation* pode ocorrer, segundo Working (1948:1), em decorrência de faltas momentâneas de estoque no mercado à vista, momento no qual o retorno de conveniência é máximo.

Frechette e Fackler (1999:761) questionaram a explicação do fenômeno de *backwardation* pelas faltas momentâneas de estoque, pois é difícil justificar a existência de *backwardation* em períodos de safra das *commodities* agrícolas, quando não há falta de estoque. Os autores sugerem uma terceira teoria, ao afirmar que o fenômeno é mais afetado por erros de medida e variáveis não observáveis. Os autores explicam que os preços à vista são calculados pela média de preços de *commodities* negociadas em localidades diferentes, o que causa distorções em relação aos preços futuros. Essas distorções de preços são causadas porque, quando a média do preço à vista é calculada, pode haver falta de estoque em alguma das localidades, fazendo com que o preço à vista médio seja mais elevado do que o preço futuro.

A quarta teoria, sugerida por Benirschka e Binkley (1995:515) em seu trabalho com o mercado de milho nos Estados Unidos, assume que a existência de *backwardation* nos preços é causada pela distância dos estoques dos grandes centros de consumo, onde o preço da *commodity* é formado. Os modelos desenvolvidos pelos autores mostram que a distribuição espacial dos estoques, e não seu nível, é determinante da diferença dos preços. Para eles, em um mercado geograficamente disperso, o custo de oportunidade de estocar uma *commodity* aumenta com a distância do mercado comprador, ou seja, áreas produtoras mais distantes do mercado consumidor têm mais vantagens em estocar a *commodity* e o *backwardation* também aumenta com a distância.

Não se descarta a hipótese do *backwardation* dos preços de *commodity* ser explicado por mais de uma das teorias acima mencionadas. E acredita-se que possam existir ainda outras variáveis, não identificadas na literatura, que também possam explicar o fenômeno.

- **Backwardation forte e fraco**

O *backwardation* pode ser classificado, de acordo com sua intensidade, em fraco ou forte. *Backwardation* forte ocorre quando o preço à vista é maior do que os preços futuros. No *backwardation* fraco, consideram-se os preços futuros descontados a valor presente pela taxa livre de risco. Litzenberger e Rabinowitz (1995:1518) definiram as seguintes variáveis para o cálculo do *backwardation* segundo a intensidade:

$$B_t^w = S_t - F_t e^{-r(T-t)} \quad (2.3)$$

$$B_t^s = S_t - F_t$$

Sendo:

B_t^w : *backwardation* fraco;

B_t^s : *backwardation* forte.

Litzenberger e Rabinowitz (1995:1520), em seu estudo empírico, verificaram presença de *backwardation* nos preços dos contratos futuros de petróleo do WTI na NYMEX, no período de fevereiro de 1984 a abril de 1992. Verificaram que, quanto mais longos os vencimentos dos contratos futuros, maiores os valores de *backwardation*, tanto para B_t^w , quanto para B_t^s . Identificaram que no contrato WTI de vencimento mais curto, houve 81,80% de presença de *backwardation* fraco e 71,09% de *backwardation* forte. Para o contrato futuro mais longo, os valores percentuais subiram para 93,88% e 76,91%, respectivamente, para *backwardation* fraco e forte.

De acordo com a Teoria da Estocagem, a presença de *backwardation* é resultante de uma taxa de retorno de conveniência relativamente elevada. Na seção seguinte são explicitadas as abordagens teóricas sobre essa variável.

2.3.4 Retorno de Conveniência

As teorias sobre o retorno de conveniência foram abordadas por Working (1948, 1949), Brennan (1958:53), Brennan e Schwartz (1985:139), Frechette e Fackler (1999:761) e mais recentemente por Geman (2005:24).

O retorno de conveniência pode ser definido como uma remuneração, medida em termos de taxa de juros no ativo, que o detentor da *commodity* física recebe pela possibilidade de escassez de estoques. Segundo Frechette e Fackler (1999:761), é um “(...) fluxo de benefícios não pecuniários (...)” que o dono da *commodity* física recebe e pode ser comparado como o retorno de dividendos de ativos do mercado financeiro. Retorno de conveniência e *backwardation* não são sinônimos. *Backwardation* é uma variável observável, verificada pela diferença de preços entre o mercado à vista e o futuro, enquanto o retorno de conveniência é um conceito econômico não observável.

De acordo com Geman (2005:24), o retorno de conveniência é um “(...) prêmio que faz jus ao proprietário do ativo físico, mas não ao proprietário do contrato futuro”. A autora define o termo como a opção de tempo embutida na *commodity* que o proprietário do ativo físico detém, que permite ao produtor vender a *commodity* quando os preços estiverem altos e segurar quando os preços estiverem baixos.

Brennan e Schwartz (1985:139) afirmam que “o retorno marginal de conveniência deve ser inversamente proporcional ao volume de estoque da *commodity*”. Porém, o volume de estoques pode ser tão grande que o retorno de conveniência seja, no limite, igual a zero.

O retorno de conveniência é expresso em termos de taxa δ_t , que será tratada, doravante no presente estudo, de acordo com a interpretação dada por Geman (2005:25), ou seja, líquida do custo de armazenagem e depreciação primária.

Dessa maneira, a formação de preços de ativos, em condição de não arbitragem, sugere que a diferença entre o preço à vista e o futuro é decorrente da taxa livre de risco menos o retorno de conveniência:

$$F_t = S_t e^{(r-\delta_i)(T-t)} \quad (2.4)$$

sendo:

S_t : preço do ativo à vista no tempo t ;

F_t : preço do contrato futuro no tempo t ;

T : vencimento do contrato futuro, posterior a t ;

r : taxa livre de risco;

δ_i : retorno de conveniência.

Neste caso específico, o retorno de conveniência é representado pela taxa de arrendamento conhecida e não aleatória, ou seja, constante no tempo. Nesse caso, a formação de preços dependeria do processo estocástico de apenas uma variável, o preço do ativo S_t , e o modelo de formação de preços de *commodity* seria similar aos modelos sugeridos por Black e Scholes (1973:639) e Merton (1973:160).

Partindo da premissa que a volatilidade do retorno de conveniência é igual a zero, a volatilidade do preço à vista passa ser igual à volatilidade do preço do contrato futuro:

$$\sigma_\delta = 0 \Rightarrow \sigma_S = \sigma_F \quad (2.5)$$

A volatilidade no mercado de *commodities* possui um comportamento um pouco distinto daquele do mercado de ativos financeiros. Os níveis de estoque, conforme Geman (2005:28), interferem na volatilidade dos preços das *commodities*. Algumas dessas características sobre volatilidade nesse mercado são apresentadas a seguir:

- a volatilidade de preços de uma *commodity* tende a ser inversamente relacionada ao volume global de estoques. Quando não há estoque excedente para amortecer a demanda, os preços à vista tendem a oscilar mais;

- o preço de uma *commodity* e sua volatilidade são positivamente correlacionados, e ambos são negativamente correlacionados aos níveis de estoque. Essa característica distingue o comportamento dos preços de *commodities* e de ações. No mercado de ações, há aumento de volatilidade quando os preços caem;
- a volatilidade dos preços futuros de *commodities* tende a reduzir-se com o aumento do prazo para vencimento do contrato. Esse efeito é explicado pelo fato de que as notícias que afetam os níveis globais de estoque, têm efeito imediato apenas nos contratos futuros de curto prazo. Os contratos de maturidades mais longas não sofrem tanto o efeito de notícias, pois o volume de produção tem tempo para ajustar-se antes do vencimento do contrato mais longo.

Na seção seguinte são apresentados alguns dos processos estocásticos presentes na literatura que descrevem a trajetória dos preços no mercado de *commodities*.

2.4 Processos Estocásticos dos Preços de *Commodities*

2.4.1 A Trajetória dos Preços no Tempo

Para determinar o preço de uma *commodity* (S_t), é preciso definir qual é o modelo que melhor descreve a trajetória de seus preços no tempo. Por essa razão, optou-se por adotar um modelo que tenha o tempo como variável independente, que descreva um processo estocástico. O modelo que se quer determinar é o seguinte:

$$S_{t-1} \xrightarrow{\text{Modelo}} S_t \quad (2.6)$$

Um processo estocástico, segundo Hull (2006:263), é suposto para uma variável cujos valores mudem com o tempo e possuam um componente aleatório. Um modelo que descreve o

processo estocástico de uma variável, como o preço de uma *commodity*, possui pelo menos dois componentes distintos: um determinístico e outro estocástico, ou seja, aleatório no tempo. Em verdade, o componente determinístico pode assumir valor zero e descrever um processo particular denominado movimento browniano.

Conforme Hull (2006:263), o preço de determinada *commodity* não é observado em tempo contínuo, ou seja, os preços são observados quando há negociação e nos intervalos em que ocorram os pregões, ou seja, em tempo discreto.

Por essa razão, o modelo que se deseja obter para descrever a trajetória de preços da *commodity* será inicialmente desenvolvido em tempo discreto e depois generalizado em tempo contínuo.

2.4.2 A Variável Aleatória

Para determinar um modelo de formação de preços, é preciso estudar as propriedades estatísticas da variável aleatória e a função conhecida que melhor descreva sua distribuição de probabilidade. Uma tendência natural e enviesada do pesquisador é transformar a variável, para fazer com que sua distribuição de probabilidade seja descrita por uma função conhecida.

Securato (1996:33) sugere a adoção da taxa de juros do retorno do ativo como variável objetivo para a análise. A taxa de retorno contínua R_t é obtida pela diferença dos logaritmos naturais dos preços da *commodity*, segundo a seguinte equação:

$$R_t = \ln(S_t) - \ln(S_{t-1}) \quad (2.7)$$

Uma das vantagens do retorno contínuo é que, em modelos de variáveis que seguem processos estocásticos, segundo Enders (2004:164), é possível extrair o componente de tendência da amostra ao tirar as diferenças dos valores observados. No retorno contínuo, tiram-se as diferenças dos logaritmos naturais, o que permite, além de eliminar a tendência dos preços, que a variável assuma valores dentro do intervalo $-\infty < R_t < \infty$ exigido para uma distribuição normal.

2.4.3 Processo de Wiener

O retorno contínuo dos preços de uma *commodity* pode seguir um processo estocástico conhecido como **processo de Wiener**. Segundo Hull (2006:265), esse processo é “utilizado em física para descrever o movimento de uma partícula que é sujeita a um grande número de pequenos choques moleculares e algumas vezes referido como movimento browniano”.

Partindo-se da premissa não comprovada de que a distribuição de probabilidade dos retornos seja descrita por uma função normal, $R_t \sim N(\mu, \sigma^2)$, pode-se padronizar uma dada observação, sendo $z = (R_t - \mu_t) / \sigma_t$ e, dessa forma, calcular o retorno, conforme a seguinte equação:

$$R_t = \mu_t + \sigma_t z \quad (2.8)$$

Séries que contêm tendência podem não reverter a um nível de longo prazo. Nota-se que uma série com tendência pode possuir componentes determinísticos μ_t e estocásticos σ_t , porém não são estacionárias.

Segundo Enders (2004:164), a diferença entre séries com tendência e séries estacionárias é que “choques aleatórios em uma série estacionária são necessariamente temporários; com o tempo, o efeito dos choques aleatórios irá se dissipar e a série reverterá para a média de longo prazo”. O autor sugere duas formas para tornar uma série estacionária: a diferenciação e a eliminação de tendência por regressão. No caso dos retornos contínuos da *commodity*, optou-se pela técnica da diferenciação do logaritmo dos preços.

Dado que o retorno contínuo pode ser calculado pela diferença dos logaritmos naturais, o logaritmo do preço à vista pode ser escrito em função do retorno:

$$\ln(S_t) = \ln(S_{t-1}) + R_t \quad (2.9)$$

O preço à vista de uma determinada *commodity* pode ser calculado aplicando-se o retorno contínuo ao preço imediatamente anterior, $S_t = S_{t-1} e^{R_t}$. É possível modelar o preço utilizando-se, no lugar do retorno, sua média e desvio-padrão, conforme a seguinte equação:

$$S_t = S_{t-1} e^{\mu_t + \sigma_t z} \quad (2.10)$$

Ao determinar o processo estocástico que descreve a trajetória do preço de uma *commodity* no passado, é possível realizar uma projeção um passo à frente, partindo da premissa de que esse processo modelado seja persistente no tempo. A previsão do valor de S_{t+1} pode ser feita da seguinte maneira:

$$S_{t+1} = S_{t-1} e^{R_t} e^{R_{t+1}} \quad (2.11)$$

Supondo que os retornos contínuos são independentes e identicamente distribuídos: $\text{covar}[R_{t-1}, R_t] = 0$ e $\text{Var}[R_{t-1}] = \text{Var}[R_t]$, tem-se a seguinte relação para as médias e variâncias:

$$\mu_t + \mu_{t+1} = 2\mu_t \quad (2.12)$$

$$\begin{aligned} \text{Var}[R_{t-1} + R_t] &= \text{Var}[R_{t-1}] + \text{Var}[R_t] + 2\text{covar}[R_{t-1}, R_t] \\ \text{Var}[R_{t-1} + R_t] &= 2\text{Var}[R_t], \quad i.i.d. \end{aligned} \quad (2.13)$$

Dado que o desvio-padrão do retorno é $\sigma_t = \sqrt{\text{Var}[R_t]}$, pode se propor a seguinte relação para a projeção um passo à frente do preço da *commodity*:

$$S_{t+1} = S_{t-1} e^{2\mu_t + \sqrt{2}\sigma_t z} \quad (2.14)$$

Para uma previsão Δt à frente, com $\Delta t \rightarrow 0$, tem-se a seguinte função:

$$S_{t+\Delta t} = S_t e^{\Delta t \mu_t + \sqrt{\Delta t} \sigma_t z} \quad (2.15)$$

Em tempo contínuo, para um intervalo de tempo infinitesimal dt , é possível preservar a relação dos preços da *commodity*:

$$S_{t+dt} = S_t e^{dt \mu_t + \sqrt{dt} \sigma_t z} \quad (2.16)$$

Sendo a equação que especifica a dinâmica dos retornos dessa *commodity* dada por:

$$\ln(S_{t+dt}) = \ln(S_t) + dt\mu_t + \sqrt{dt}\sigma_t z \quad (2.17)$$

$$R_t = dt\mu_t + \sqrt{dt}\sigma_t z \quad (2.18)$$

Ou simplesmente:

$$R_t = \mu dt + \sigma \sqrt{dt} z, \quad i.i.d. \quad (2.19)$$

O preço de um ativo, segundo Hull (2006:263), pode seguir um processo estocástico particular denominado processo de Markov ou markoviano, no qual apenas o preço de hoje é relevante para a previsão do preço futuro, os valores passados dos preços são irrelevantes.

Esse processo pode ser contestado quando aplicado a preços de *commodities* em geral. Geman (2005:61) sugere a existência de algum tipo de memória nos preços de longo prazo, por exemplo, ao se identificar sazonalidade nos preços desses produtos. Porém, podendo-se separar, ou modelar, o efeito sazonal, o processo markoviano poderá ser válido em determinados modelos de formação de preços de *commodity*.

Conforme visto anteriormente, aplicou-se a primeira diferença para eliminar a tendência nos preços, pode-se também aplicar a diferença, a fim de eliminar a tendência nos retornos, conforme segue:

$$\begin{aligned} R_t &= \mu_t + \sigma_t z \\ R_{t+1} &= \mu_{t+1} + \sigma_{t+1} z', \quad i.i.d. \\ \Delta R_{t+1} &= R_{t+1} - R_t \end{aligned} \quad (2.20)$$

A primeira diferença dos retornos, segundo Enders (2004:7), é igual à segunda diferença do logaritmo dos preços $\Delta R_t = \Delta^2 \ln(S_t)$, calculada da seguinte forma:

$$\Delta R_{t+1} = \mu_{t+1} - \mu_t + \sigma_{t+1} z' - \sigma_t z \quad (2.21)$$

Tendo a variância da diferença dos retornos dada por:

$$\begin{aligned} \text{Var}[R_{t+1} - R_t] &= \text{Var}[R_{t+1}] + \text{Var}[R_t] - 2\text{covar}[R_{t+1}, R_t] \\ \text{Var}[R_{t+1} - R_t] &= 2\text{Var}[R_t], \quad i.i.d. \end{aligned} \quad (2.22)$$

Dessa forma, elimina-se o componente determinístico, permanecendo apenas o componente estocástico:

$$\begin{aligned} \sqrt{2\text{Var}[R_t]} &= \sqrt{2}\sigma_t \\ \Delta R_{t+1} &= \sqrt{2}\sigma_t z \end{aligned} \quad (2.23)$$

Ou na forma contínua:

$$dR = \sqrt{dt}\sigma_t z \quad (2.24)$$

De forma genérica, é possível supor que a diferença dos retornos de um ativo é composta apenas de choques aleatórios e que dependam exclusivamente da volatilidade σ_t dos retornos. No limite com $\Delta t \rightarrow 0$, Geman (2005:59) supõe que:

$$R_t = \ln \frac{S_t}{S_{t-1}} \approx \frac{S_t - S_{t-1}}{S_{t-1}} \approx \frac{dS}{S_{t-1}} \quad (2.25)$$

Dado que o retorno é uma função de sua média e volatilidade, a equação que descreve a dinâmica do retorno de um ativo pode ser dada por:

$$\frac{dS}{S_t} = \mu dt + \sigma \sqrt{dt} z \quad (2.26)$$

O uso da taxa de retorno discreta na equação do movimento browniano dos preços de uma *commodity*, como na equação acima, pode ser questionável. Os retornos discretos possuem uma distribuição que, apesar de se assemelhar com uma distribuição normal, não possui as condições necessárias para que seja normal, pois o retorno discreto não pode ser menor do que -100%.

Essa condição de distribuição permitiu a derivação de equações fechadas para formação de preços de opções como em Black e Scholes (1973:639) e Merton (1973:160).

Na seção seguinte será feita uma breve explanação do Lema de Itô, que permite a determinação do processo de uma variável transformada.

2.4.4 Lema de Itô

Cochrane (2005:461) explica que o Lema de Itô (1951) permite que se obtenha uma equação que descreva a dinâmica de uma variável transformada (y_t) como função da variável original (S_t), ou seja, $y_t = f(S_t)$. No caso específico do preço de uma *commodity*, deseja-se determinar a função do logaritmo natural do preço no tempo t :

$$y_t = \ln(S_t) \quad (2.27)$$

A equação do movimento browniano do preço de uma determinada *commodity* pode ser representada da seguinte forma, conforme visto anteriormente:

$$dS = \mu S_t dt + \sigma S_t \sqrt{dt} z \quad (2.28)$$

A equação acima é definida por Cochrane (2005:459) como um modelo de difusão de preços de um ativo (S) através de um processo estocástico. O autor sugere uma representação mais genérica, para uma variável (x) qualquer, aleatória no tempo, no qual a simbologia (\cdot) significa uma função da variável x , cuja equação de difusão é representada por:

$$dx = \mu(\cdot)dt + \sigma(\cdot)dz_t \quad (2.29)$$

Na Equação 2.29 acima, o **componente estocástico** $\sqrt{dt}z$ foi representado, por simplicidade de notação, apenas por dz_t , com $\Delta t \rightarrow 0$. Cochrane (2005:460) utiliza z_t para denotar z em

função do tempo e determina que z_t é igual à soma de suas pequenas diferenças, sendo as esperanças de dz_t dadas por:

$$\begin{aligned} E_t(dz_t) &= 0 \\ E_t(dz_t^2) &= dt \end{aligned} \quad (2.30)$$

Cochrane (2005:460) sugere a construção de uma variável y que seja uma função de x , $y_t = f(x_t)$, com o intuito de explicar os procedimentos adotados no Lema de Itô.

Partindo da definição anterior de que $\sqrt{dt}z$ é representado por dz_t , para $\Delta t \rightarrow 0$, Cochrane (2005:460) sugere que $dz_t^2 = dt$, que os termos $dt dz_t$, dt^2 tendem a zero e que, aproximadamente, pequenas variações no valor da variável transformada y podem ser calculadas pela derivada parcial em função de x , conforme a seguinte equação:

$$\Delta y \approx \frac{dy}{dx} \Delta x \quad (2.31)$$

Na expansão de segunda ordem da série de Taylor para a variável y em função de x , tem-se:

$$dy = \frac{df(x)}{dx} dx + \frac{1}{2} \frac{d^2 f(x)}{dx^2} dx^2 \quad (2.32)$$

Expandindo o segundo termo da equação, com base na equação de difusão estocástica de dx , tem-se:

$$dx^2 = [\mu_x dt + \sigma_x dz_t]^2 = \mu_x^2 dt^2 + \sigma_x^2 dz_t^2 + 2\mu_x \sigma_x dt dz_t \quad (2.33)$$

Agora, segundo Cochrane (2005:460), aplica-se a regra definida no Lema de Itô de que $dt^2 = 0$ e que $dz_t^2 = dt$, tem-se, portanto, que $dx^2 = \sigma_x^2 dt$. Substituindo-se os valores de dx e dx^2 na equação de dy :

$$dy = \frac{df(x)}{dx}(\mu_x dt + \sigma_x dz_t) + \frac{1}{2} \frac{d^2 f(x)}{dx^2} \sigma_x^2 dt \quad (2.34)$$

$$dy = \left[\frac{df(x)}{dx} \mu_x + \frac{1}{2} \frac{d^2 f(x)}{dx^2} \sigma_x^2 \right] dt + \frac{df(x)}{dx} \sigma_x dz_t \quad (2.35)$$

Logo, a equação do Lema de Itô, segundo Cochrane (2005:461), que determina o modelo de difusão estocástica da variável transformada y , em função da variável original x , é representada por:

$$dy = \left[\frac{df(x)}{dx} \mu_x(.) + \frac{1}{2} \frac{d^2 f(x)}{dx^2} \sigma_x^2(.) \right] dt + \frac{df(x)}{dx} \sigma_x(.) dz_t \quad (2.36)$$

A peculiaridade dessa equação está na presença de um segundo termo no componente determinístico.

Definida a equação do Lema de Itô, é possível voltar ao problema inicial, no qual se desejava determinar a equação de dispersão estocástica do logaritmo natural do preço de uma *commodity* (S_t), utilizando-se uma variável transformada definida pela função $y_t = \ln(S_t)$, da qual se podem calcular as seguintes derivadas parciais:

$$\frac{\partial y_t}{\partial S_t} = \frac{1}{S_t}, \quad \frac{\partial^2 y}{\partial S^2} = -\frac{1}{S_t^2} \quad (2.37)$$

A equação de dispersão da variável y deverá ser igual a:

$$dy = \left[\frac{1}{S_t} \mu_x S_t + \frac{1}{2} \left(-\frac{1}{S_t^2} \right) \sigma_x^2 S_t^2 \right] dt + \frac{1}{S_t} \sigma_x S_t dz_t \quad (2.38)$$

Ou simplesmente:

$$dy = \left[\mu_x - \frac{1}{2} \sigma_x^2 \right] dt + \sigma_x dz_t \quad (2.39)$$

A equação acima é de grande utilidade para a determinação de modelos de formação de preços de derivativos, como é o caso dos trabalhos de Black e Scholes (1973:639) e Merton (1973:160), que definem equações fechadas para formação de preços de opções. Esses modelos descrevem processos estocásticos em movimento browniano para preços de ativos financeiros, que podem ser aplicados também aos preços de derivativos de *commodity*.

Além do movimento browniano, explicado nesta seção, o preço de *commodities* podem seguir um processo de reversão à média, apresentado na seção seguinte.

2.4.5 Reversão à Média

O modelo analisado na seção anterior utilizou uma variável aleatória transformada: as diferenças dos logaritmos do preço da *commodity* S_t . O objetivo dessa transformação foi tornar a série **estacionária**. Enders (2004:319) afirma, porém, que é possível estimar modelos lineares estacionários utilizando variáveis não estacionárias e define essas variáveis como cointegradas.

Hamilton (1994:571) define que uma variável aleatória y_t , que sofre um processo estocástico, é denominada **cointegrada** quando cada uma das séries do modelo, tomadas individualmente, seja não estacionária com pelo menos uma raiz unitária. Ao mesmo tempo, pelo menos uma combinação linear das variáveis deve ser estacionária.

Hamilton (1994:572) explica que, quando há cointegração, mesmo que alguns eventos possam causar mudanças permanentes na variável em questão y_t , existe algum tipo de relação de equilíbrio de longo-prazo (*long-run*), unindo as variáveis novamente.

A principal característica de variáveis cointegradas, segundo Enders (2004:328) é que a “(...) trajetória é influenciada por quaisquer desvios extensos do equilíbrio de longo prazo”. Mesmo que o sistema como um todo retorne para o equilíbrio de longo prazo, pelo menos uma das variáveis continuará respondendo ao efeito do desequilíbrio ocorrido. Em outras palavras: a dinâmica de **curto prazo** (*short-run*) deve ser influenciada pelos desvios ocorridos nas relações de **longo prazo** (*long-run*).

O processo de retorno ao equilíbrio de longo prazo é definido por Enders (2004:329) como **correção do erro**. Esse processo, característico de variáveis cointegradas, é mencionado por Cochrane (2005:459) como processo Ornstein-Uhlenbeck de reversão à média.

Cochrane (2005:459) explica que o processo de reversão à média começa a partir de **um processo autorregressivo** de ordem 1, $AR(1)$. Supondo que o preço da *commodity* (S_t) cointegre com ele mesmo, descrevendo um processo $AR(1)$, conforme a seguinte equação:

$$S_t = \beta_0 + \beta_1 S_{t-1} + \varepsilon_t \quad (2.40)$$

sendo:

ε_t : ruído branco, componente de erro estocástico do processo $AR(1)$;

β_0 e β_1 : coeficientes da regressão.

A média do preço da *commodity* é dada por $E[S_t] = \beta_0 + \beta_1 S_{t-1}$, ou seja, o preço pode ser calculado em relação a sua média:

$$S_t = \mu_t + \varepsilon_t \quad (2.41)$$

Partindo da equação acima, é possível definir uma nova variável: o **desvio em relação à média** $D_t = S_t - \mu_t$, similar ao ruído branco ε_t . O valor esperado do desvio D_t , no limite, é igual a zero, $E[D_t] = 0$. No processo de **reversão à média**, é possível estimar o modelo de regressão dos desvios seguindo um processo de média móvel:

$$D_t = \alpha D_{t-1} + \zeta_t \quad (2.42)$$

sendo:

ζ_t : ruído branco;

α : coeficiente da regressão;

Neste caso, para um processo de reversão à média, o coeficiente α da regressão deve ser negativo e, em módulo, menor que 1, ou seja, $-1 < \alpha < 0$.

Utilizando a equação acima, substitui-se o desvio da média D_t e chega-se à seguinte equação:

$$S_t - \mu = \alpha(S_{t-1} - \mu) + \zeta_t \quad (2.43)$$

Subtraindo-se S_{t-1} dos dois lados da equação:

$$S_t - S_{t-1} = \alpha(S_{t-1} - \mu) - S_{t-1} + \mu + \zeta_t \quad (2.44)$$

E finalmente se chega, conforme Cochrane (2005:458), partindo de um processo $AR(1)$, ao modelo de reversão à média em tempo discreto para o preço da *commodity*:

$$S_t - S_{t-1} = -(1 - \alpha)(S_{t-1} - \mu) + \zeta_t \quad (2.45)$$

Com a versão análoga em tempo contínuo dada por:

$$dS = -\phi(S_t - \mu)dt + \sigma dz_t \quad (2.46)$$

sendo:

ϕ : velocidade de reversão à média.

Nas seções anteriores foram apresentados os processos estocásticos que descrevem a trajetória dos preços de uma *commodity*. Os processos relevantes vistos foram: movimento browniano e reversão à média. Partindo desses processos, foram desenvolvidos modelos que tentam descrever, a partir da combinação de variáveis, o comportamento dos preços das *commodities*. Na seção seguinte é feita breve apresentação dos principais modelos disponíveis na literatura.

2.5 Modelagem de Preços de *Commodities*

2.5.1 Modelo com Base no CAPM

Na literatura existem duas abordagens gerais para os modelos de formação de preços de *commodity*. A primeira baseia-se no retorno de conveniência e na teoria da estocagem. A segunda é baseada no prêmio de risco sistêmico da *commodity* e medida através do beta do modelo de formação de preços de ativos de capital sugerida por Dusak (1973:1393) e tem como base a teoria do CAPM (*Capital Asset Pricing Model*).

O modelo de formação de preços de *commodity* com base no CAPM considera que o preço do contrato futuro de uma *commodity* é formado pelo valor esperado do preço à vista mais um prêmio pelo risco, que pode ser medido em termos de risco sistêmico, ou seja, há um beta referente à *commodity* (β_i). O retorno esperado da *commodity* pode ser calculado conforme a fórmula do CAPM:

$$E(R_i) = r + \beta_i [E(R_m) - r] \quad (2.47)$$

sendo:

$E(R_i)$: retorno esperado da *commodity* i ;

r : taxa de juros livre de risco;

$E(R_m)$: retorno de mercado esperado.

Dusak (1973:1393) define a fórmula do retorno do preço de *commodity* para determinar o preço de contratos futuros também com base no CAPM. Partindo de $\beta_i = \text{Cov}(R_i, R_m) / \sigma^2(R_m)$, tem-se:

$$E(R_i) = r + \left[\frac{E(R_m) - r}{\sigma(R_m)} \right] \frac{\text{Cov}(R_i, R_m)}{\sigma(R_m)} \quad (2.48)$$

sendo:

$\sigma(R_m)$: desvio-padrão dos retornos da carteira teórica de mercado;

$\text{Cov}(R_i, R_m)$: covariância esperada dos retornos da *commodity* e da carteira de mercado.

O retorno esperado de uma *commodity* i para uma determinada data no futuro, representada por T , dependerá do valor esperado da *commodity* no futuro, da seguinte forma:

$$E(R_i) = \frac{E(S_T) - S_t}{S_t} \quad (2.49)$$

sendo:

$E(S_T)$: preço esperado do preço da *commodity* na data T ;

S_t : preço atual no mercado à vista de *commodity*.

Dessa forma, pode-se obter o preço à vista da *commodity* por meio da seguinte equação:

$$S_t = \frac{E(R_i) - S_t \beta_i [E(R_m) - r]}{(1+r)} \quad (2.50)$$

Dusak (1973:1393) afirma que o preço do contrato futuro com vencimento em T de uma determinada *commodity* pode ser estimado apenas por meio do preço à vista e pela taxa livre de risco e supõe que $F_t = S_t(1+r)$. Sendo assim, o preço do contrato deverá ser:

$$F_t = S_t(1+r) = E(R_i) - S_t \beta_i [E(R_m) - r] \quad (2.51)$$

Portanto, é possível estimar a diferença percentual entre o preço esperado da *commodity* e o preço efetivo do contrato futuro com base no prêmio de risco sistêmico da *commodity*, conforme a seguinte equação:

$$\frac{E(R_i) - F_t}{S_t} = \beta_i [E(R_m) - r] \quad (2.52)$$

Dusak (1973:1394) sugere que o coeficiente de risco sistêmico (β_i) pode ser estimado de duas formas distintas: pode-se tomar o preço do ativo no mercado à vista ou o preço do contrato futuro, se o primeiro não estiver disponível para observação.

O modelo CAPM para *commodities* considera o risco sistêmico como variável relevante, mas desconsidera variáveis importantes da teoria da estocagem, como o retorno de conveniência. Copeland e Weston (1988:319) argumentam em favor do modelo com base no CAPM, pois afirmam que “por outro lado, a abordagem tradicional, com base na teoria da estocagem, ignora a possibilidade de que o risco sistêmico possa afetar os preços de equilíbrio dos contratos futuros de *commodity*”. Pode ser levada em consideração a possibilidade de combinação das duas teorias, ou seja, levar em conta não apenas o prêmio de risco sistêmico, mas também o retorno de conveniência e as premissas da teoria de estocagem em um modelo híbrido.

Em seguida serão abordados os principais modelos que utilizam como base o retorno de conveniência e a teoria da estocagem.

2.5.2 Modelos de Um, Dois e Três Fatores

Os modelos de formação de preços de *commodities* evoluíram, de acordo com a literatura aqui examinada, com o objetivo de tentar explicar da melhor forma o comportamento dos preços, com inclusão de variáveis. As variáveis consideradas, ou fatores adicionados ao modelo, foram as seguintes: o preço do ativo, o retorno de conveniência e as taxas de juros.

A modelagem de preços de *commodities*, com base em processos estocásticos dos preços, é abordada inicialmente no trabalho de Brennan e Schwartz (1985:138). Os autores estudam a trajetória do preço da *commodity* para avaliar a viabilidade de investimentos de capital na exploração de minas de cobre. O foco inicial da abordagem era introduzir o preço da *commodity* como variável determinante na decisão de investimento, teoria conhecida na literatura financeira como **opções reais**.

De acordo com a definição de Copeland e Antikarov (2001:6), “uma opção real é o direito, mas não a obrigação, de empreender uma ação (por exemplo, diferir, expandir, contrair ou abandonar) a um custo predeterminado, que se denomina preço de exercício, por um período preestabelecido – a vida da opção”, comumente utilizada em projetos de investimento de capital nos quais o preço de uma *commodity* é decisivo para a viabilidade econômica do projeto.

A opção real permite aos administradores tomarem decisões de investimento de capital com base na avaliação econômico-financeira e na maximização de valor para o proprietário do negócio. É um procedimento comumente utilizado em projetos de exploração de *commodities* não renováveis como o petróleo. De acordo com o preço esperado da *commodity*, o administrador pode tomar decisões de expansão, retração, postergação, antecipação, ou mesmo abandono de um projeto de exploração.

Brennan e Schwartz (1985:139) estudaram a formação de preços de *commodities* com base no **modelo de um fator**, no qual utilizam como única variável aleatória do modelo o preço do ativo à vista, descrevendo um movimento browniano, de acordo com a seguinte equação:

$$\frac{dS}{S_t} = \mu dt + \sigma dz \quad (2.53)$$

em que dz é o choque aleatório de um processo de Wiener padrão, μ é o componente determinístico que representa a tendência no preço da *commodity* e σ é a volatilidade dos retornos.

Brennan e Schwartz (1985:139), em seu trabalho, não incluíram o retorno de conveniência como variável aleatória. Os autores definiram o retorno de conveniência como o prêmio ao

qual o detentor do ativo à vista tem direito, considerando que o produtor pode obter ganhos ao controlar o processo produtivo ou aproveitando-se da falta do produto no mercado à vista. O retorno de conveniência, na visão de Brennan e Schwartz (1985:139) é determinado pelos volumes de estoque do produto no mercado. A taxa de juros livre de risco foi considerada constante no modelo, restando apenas uma variável aleatória, o preço S_t da *commodity*.

Em trabalho posterior, Gibson e Schwartz (1990:960) definem o **modelo de dois fatores** aplicado aos preços do barril de petróleo e assumem que o preço à vista e o retorno de conveniência seguem um processo estocástico conjunto. O preço da *commodity* segue um processo de passeio aleatório em tempo contínuo e o retorno de conveniência, um processo de reversão à média, conforme as seguintes equações:

$$\begin{aligned} \frac{dS}{S_t} &= \mu_s dt + \sigma_s dz_s \\ d\delta &= \phi_\delta (\mu_\delta - \delta_t) dt + \sigma_\delta dz_\delta \end{aligned} \tag{2.54}$$

sendo:

δ_t : retorno de conveniência;

μ_δ : valor de equilíbrio médio estimado para o retorno de conveniência;

ϕ_δ : velocidade de reversão à média do retorno de conveniência.

e dz_s e dz_δ são choques aleatórios correlacionados: $dz_s dz_\delta = \rho dt$.

Em seus resultados com os preços futuros de petróleo WTI na *New York Mercantile Exchange*, Gibson e Schwartz (1990:974) verificaram que o retorno de conveniência possui mais volatilidade do que o preço à vista e que, quanto maior a maturidade, menor a volatilidade do preço do contrato.

Em trabalho posterior, utilizando o filtro de Kalman, Schwartz (1997:925) comparou os modelos: (i) um fator, no qual considera apenas o preço como variável aleatória, seguindo um processo de passeio aleatório em tempo contínuo; (ii) dois fatores, utilizando, além do movimento browniano do preço à vista, um processo de reversão à média para o retorno de conveniência; e (iii) modelo de três fatores, que agrega a taxa de juros r como variável aleatória, seguindo um processo de reversão à média Ornstein-Uhlenbeck.

O **modelo de três fatores** possui a seguinte estrutura:

$$\begin{aligned}\frac{dS}{S_t} &= (r_t - \delta_t)dt + \sigma_S dz_S \\ d\delta &= \phi_\delta (\mu_\delta - \delta_t)dt + \sigma_\delta dz_\delta \\ dr &= \phi_r (\mu_r - r_t)dt + \sigma_r dz_r\end{aligned}\tag{2.55}$$

sendo:

r_t : taxa de juros no tempo t ;

μ_r : valor de equilíbrio médio estimado para as taxas de juros;

ϕ_r : velocidade de reversão à média para as taxas de juros;

e dz_S , dz_δ , dz_r são choques aleatórios correlacionados:

$$\begin{aligned}dz_S dz_\delta &= \rho_1 dt \\ dz_\delta dz_r &= \rho_2 dt \\ dz_S dz_r &= \rho_3 dt\end{aligned}\tag{2.56}$$

Schwartz (1997:972) realizou testes de comparação dos modelos com os mercados futuros de cobre, petróleo e ouro, chegando às seguintes conclusões:

- a) no modelo de um fator, a volatilidade dos preços dos contratos futuros convergiu para zero e o preço, para uma constante, à medida que aumenta a maturidade do contrato futuro;
- b) nos modelos de dois e três fatores, a volatilidade diminuiu, porém convergiu para um valor constante diferente de zero, a estrutura temporal dos preços futuros tendeu para uma taxa de crescimento constante, porém ainda atrás do preço à vista.

2.5.3 Modelo de Curto e Longo Prazo

Em trabalho mais recente, Schwartz e Smith (2000:895) desenvolveram um modelo distinto de dois fatores para preço de *commodity*, utilizando os preços do petróleo para os testes. Diferentemente dos trabalhos anteriores, supõem que o preço à vista sofre um processo de reversão à média e os preços dos contratos futuros de maturidades mais longas tendem a um preço de equilíbrio. Consideram que os choques de curto prazo, definidos como a diferença entre o preço *spot* e o preço de equilíbrio, têm reversão a zero, seguindo um processo Ornstein-Uhlenbeck.

Para Schwartz e Smith (2000:894), os choques aleatórios de curto prazo são “conseqüência de mudanças na demanda de curto prazo, causadas por alterações climáticas ou interrupções de fornecimento intermitentes”. Essas irregularidades de fornecimento são compensadas pela habilidade dos participantes do mercado de ajustar seus estoques.

Por exemplo, em determinada região pode haver falta momentânea de estoque de determinada *commodity* agrícola e conseqüente interrupção no fornecimento. Esse fenômeno poderia causar um aumento de curto prazo no preço da *commodity* na região. Esse choque aleatório faria com que a safra no período seguinte fosse compensada com um aumento na produção, motivada pela alta nos preços da *commodity*. Maior oferta na safra seguinte faria com que o preço revertisse a seu equilíbrio original.

No modelo curto prazo/longo prazo, Schwartz e Smith (2000:894) definem que:

[...] mudanças nos preços dos contratos futuros de longo prazo de maturidade oferecem informações sobre as mudanças no preço de equilíbrio, e as mudanças nas diferenças de preços entre os futuros de curto e longo prazos, oferecem informação sobre os desvios (choques) de curto prazo.

No modelo de Schwartz e Smith (2000:895), o preço à vista da *commodity* é decomposto em dois fatores:

$$\ln(S_t) = \chi_t + \omega_t \quad (2.57)$$

sendo:

χ_t : variações nos preços de curto prazo no tempo t ;

ω_t : nível do preço de equilíbrio no tempo t .

Os desvios de curto prazo seguem um processo de Ornstein-Uhlenbeck de reversão à média e o preço de equilíbrio segue um processo de movimento browniano:

$$\begin{aligned} d\chi_t &= -\phi_\chi \chi_t dt + \sigma_\chi dz_\chi \\ d\omega_t &= \mu_\omega dt + \sigma_\omega dz_\omega \end{aligned} \quad (2.58)$$

em que dz_χ e dz_ω são incrementos aleatórios correlacionados de um movimento browniano padronizado, sendo $dz_\chi \cdot dz_\omega = \rho_{\chi\omega} dt$. O coeficiente de reversão à média ϕ_χ representa a velocidade cujos desvios de curto prazo irão desaparecer, dado que $\mu_\chi = 0$.

Mudanças de curto prazo nos preços, de acordo com Schwartz e Smith (2000:895), são resultantes, por exemplo, de mudanças climáticas, e não persistem com o tempo. Por outro lado, mudanças no preço de equilíbrio são mudanças fundamentais que persistem com o tempo. Em seu trabalho, os autores definem uma versão risco-neutra para formação de preços de opções do tipo europeia para *commodities*. Utilizaram o filtro de Kalman para formular o modelo de curto prazo / longo prazo, trabalhando com as variáveis em tempo discreto, com

base no relacionamento das variáveis de estado e os preços dos contratos futuros da *commodity*.

Na próxima seção é discutida a adição de um componente sazonal ao modelo.

2.5.4 Sazonalidade

Os preços das *commodities* agrícolas possuem sazonalidade decorrente dos períodos de **safr**a e **entressafra**. Sorensen (2002:395) afirma que os preços de *commodities* agrícolas seguem padrões cíclicos: em períodos de entressafra, nos quais a oferta é menor, os preços tendem a ser maiores em relação aos períodos de safra, quando a oferta é maior. O autor sugere que o efeito de sazonalidade pode ser explicado pela teoria da estocagem.

De acordo com essa teoria, existe uma relação negativa entre o volume de estoques e o preço da *commodity*. Isso ocorre porque, quando o benefício marginal de retenção dos estoques é maior do que o custo marginal de carregamento, é mais vantajoso para o produtor reter estoques para obter melhores preços de venda, levando em consideração os custos de armazenagem da mercadoria. Logo, quando os volumes de estoques estão baixos, no período imediatamente anterior ao da safra, o preço da mercadoria é máximo e o retorno de conveniência também.

Existem diversas formas para modelar o componente sazonal em séries de tempo, a abordagem mais tradicional é por meio de *dummies* de sazonalidade aditivas, normalmente utilizando-se coeficientes para cada um dos meses do ano, menos um.

Uma outra abordagem, sugerida por Hannan, Terrell e Tuckwell (1970:26) utiliza **componentes trigonométricos**, que permitem identificar ciclos, amplitude e frequência da sazonalidade. Sorensen (2002:414) adicionou um componente sazonal ao modelo de Schwartz e Smith (2000:895). O componente de sazonalidade implementado foi o seguinte:

$$\lambda_t = \sum_{k=1}^K [\beta_{c,k} \cos(2\pi k\theta) + \beta_{s,k} \sin(2\pi k\theta)] \quad (2.59)$$

sendo:

θ : tempo, medido em anos;

K : frequência dos ciclos sazonais anuais;

$\beta_{c,k}, \beta_{s,k}$: coeficientes estimados cosseno e seno, para cada um dos K ciclos.

A Ilustração 1 mostra a representação gráfica dos componentes cíclicos sugeridos por Sorensen (2002:414), para modelar a sazonalidade em um período de doze meses:

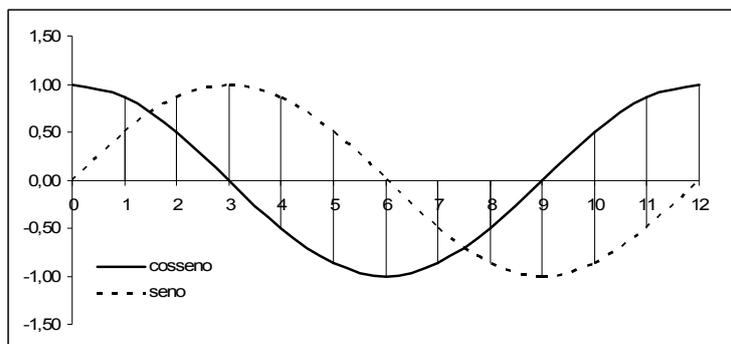


Ilustração 1 - Componentes sazonais seno e cosseno para período de doze meses

No capítulo seguinte são apresentadas as premissas para a especificação do modelo de formação de preços de *commodities* agrícolas proposto neste estudo.

3 O MODELO PROPOSTO

3.1 Evolução dos Modelos Anteriores

Os modelos de formação de preços de *commodities* disponíveis na literatura evoluíram com o tempo, com a adição de variáveis e com a especificação de diferentes processos estocásticos para as variáveis dependentes e explicativas. O intuito dos trabalhos anteriores era obter modelos cada vez melhores, que explicassem a trajetória dos preços das *commodities* e tivessem uma melhor capacidade de previsão. As principais variáveis adicionadas aos modelos foram: o retorno de conveniência, as taxas de juros e componentes de sazonalidade. Com a adição dessas variáveis, houve um aumento do poder de explicação da trajetória dos preços, em detrimento da parcimônia e da simplicidade.

Neste capítulo é apresentado um novo modelo de formação de preços de *commodities*. O objetivo da introdução de um novo modelo é que este tenha um maior poder de explicação da trajetória dos preços de *commodities* do que os modelos anteriores e que seja, ao mesmo tempo, parcimonioso.

3.2 Teoria da Paridade de Duas *Commodities* de Walras

Para desenvolver as premissas necessárias para a especificação do modelo de formação de preços de *commodity*, partiu-se da teoria apresentada por Walras (2003:87) de que uma *commodity*, em um mercado perfeitamente competitivo, pode ser representada, em termos de valor, pela quantidade de **outra commodity**. Duas *commodities* (a) e (b) podem ser representadas em termos de quantidades uma da outra. Isso ocorre porque um indivíduo que possui a *commodity* (a) deseja vendê-la para adquirir uma determinada porção de *commodity* (b); enquanto outro indivíduo que tem a (b) deseja comprar a *commodity* (a). Partindo da equivalência de preços, é possível determinar um valor relativo de uma *commodity* pela outra. Walras (2003:87) define a seguinte equação de troca:

$$mv_a = nv_b \quad (3.1)$$

sendo:

m : quantidade de *commodity* (a);

n : quantidade de *commodity* (b);

v_a : valor unitário da *commodity* (a);

v_b : valor unitário da *commodity* (b).

O preço de uma *commodity* pode ser representado pela razão de troca, ou seja, o preço de (b) pode ser representado em termos da *commodity* (a) por p_b , e o preço de (a) pode ser representado em termos de (b) por p_a , e os quociente das quantidades podem ser representados, respectivamente por $m/n = q$ e $n/m = 1/q$, resultando na relação de preços para as *commodities* conforme as seguintes equações:

$$\begin{aligned} \frac{v_a}{v_b} = p_a = \frac{n}{m} = \frac{1}{q} \\ \frac{v_b}{v_a} = p_b = \frac{m}{n} = q \end{aligned} \quad (3.2)$$

sendo:

q : quociente da relação de troca entre as *commodities* (a) e (b);

p_b : preço da *commodity* (b) representado em termos da *commodity* (a);

p_a : preço da *commodity* (a) representado em termos da *commodity* (b).

Partindo-se dessa teoria, chega-se à relação de preços de duas *commodities*:

$$p_a = \frac{1}{p_b}, \quad p_b = \frac{1}{p_a} \quad (3.3)$$

Walras (2003:87) sugere, em sua teoria, que “o preço de qualquer *commodity*, em termos de outra, é a reciprocidade de preço de uma segunda em termos da primeira”. Partindo-se dessa teoria, é possível, portanto, representar os preços de todas as *commodities*, negociadas em mercados competitivos, em termos de preços relativos de apenas uma *commodity*, ou seja, é possível representar o preço de todas as *commodities* em termos do **petróleo**.

A teoria proposta por Walras (2003:87) descreve uma relação instantânea de preços relativos entre *commodities*. Essa relação pode modificar-se com o tempo e atingir outros preços relativos com o equilíbrio dos mercados, de acordo com as forças de oferta e demanda por cada uma das *commodities*.

3.3 Relação do Preço do Petróleo com Outras *Commodities*

O petróleo, por definição, é um líquido inflamável, encontrado no subsolo em formações rochosas, primordialmente composto por hidrocarbonetos, material orgânico, enxofre, nitrogênio e oxigênio. Em termos de volume, é utilizado principalmente para produção de óleos combustíveis, gasolina, querosene de aviação, óleo para aquecimento, óleo diesel e GLP, Gás Liquefeito de Petróleo, fontes primárias de energia. Graças a sua alta concentração de energia, fácil transporte e relativa abundância, o petróleo tornou-se a principal fonte de energia desde a década de 1950. Além disso, é também matéria-prima de outros subprodutos não combustíveis, como farmacêuticos, solventes, fertilizantes, pesticidas, plásticos e polímeros em geral.

O consumo atual de petróleo está em torno de 84 milhões de barris por dia e cresce à medida que aumenta a demanda de energia pelas novas economias, como a China e a Índia. O consumo crescente do petróleo provoca uma redução no EROEI, *Energy Return Over Energy Invested*, ou seja, com a redução das reservas de fácil exploração, torna-se necessário explorar petróleo de difícil acesso ou com menor concentração de hidrocarbonetos, a um maior custo. O EROEI, ou EROI, é o ganho líquido de energia. De acordo com a Oil Analytics (2008), é o quociente de toda a energia obtida, dividido pela energia gasta na exploração e transporte do petróleo, conforme a seguinte equação:

$$k_e = \frac{E_o}{E_s} \quad (3.4)$$

sendo:

k_e : ganho líquido de energia;

E_o : total de energia obtida;

E_s : energia despendida no processo de exploração;

À medida que o ganho líquido de energia na exploração de petróleo diminui, algumas fontes alternativas de energia tornam-se viáveis economicamente. Portanto, a partir do momento que o EROI dos **combustíveis alternativos** fica maior do que o do petróleo, estes se tornam fontes viáveis de energia substituta aos combustíveis fósseis.

Entre a lista das principais fontes alternativas de energia estão os **biocombustíveis**, que podem ser definidos como combustíveis, sólidos, líquidos ou gasosos produzidos a partir de **material biológico**, fontes de carbono, em sua maioria de origem vegetal. Os biocombustíveis surgem como alternativa renovável, pois, através da fotossíntese, as plantas podem converter energia solar em material orgânico.

A vantagem ambiental dos biocombustíveis é que, além de ser uma fonte de energia renovável, permite a captação de carbono da atmosfera. A desvantagem, apresentada pelos ambientalistas, é que a cultura de vegetais com destino à produção de biocombustível pode ampliar as áreas plantadas e promover o desmatamento de florestas tropicais. O intuito deste estudo não é discutir os impactos ambientais da produção dos biocombustíveis, apesar de ser um tópico de extrema relevância, mas apenas a formação de preços destes.

Uma grande parte das *commodities* de origem vegetal podem ser destinadas a produção de biocombustível, como milho, soja, canola, mamona, beterraba e cana-de-açúcar. Todos esses vegetais possuem a capacidade de ser convertidos em biocombustíveis, porém, de acordo com

a UNICA (2008), a **cana-de-açúcar** é a mais eficiente na produção de energia, pois produz um maior ganho líquido de energia do que os outros vegetais.

Com base na hipótese de que os biocombustíveis podem ser alternativas de substituição do petróleo como fonte primária de energia, foram elaboradas premissas que serão a base para a especificação de um modelo de formação de preços de *commodities*. Essas premissas são apresentadas na próxima seção.

3.4 Premissas do Modelo

A principal premissa, seguindo uma linha um pouco distinta dos modelos anteriormente propostos na literatura, sugere que a formação de preços de uma *commodity* pode ser influenciada pelos **preços de uma outra commodity** e que essa relação pode persistir no tempo.

Ao observar o mercado de *commodities*, verifica-se que algumas possuem certa relação de **interdependência** na formação de preços: o movimento em determinado mercado pode antecipar a trajetória de preços em outro mercado inter-relacionado. Um exemplo é identificado no mercado de *commodities* agrícolas relacionadas com a produção de bioenergia e do petróleo.

A premissa aqui sugerida supõe que a intercambialidade de valor pode persistir no tempo para algumas *commodities*, como é o caso do petróleo e, indiretamente, do **açúcar** proveniente da **cana-de-açúcar**. Com base nesse pressuposto, foram elaboradas premissas fundamentais que serão a base para a especificação do modelo aplicado ao mercado de açúcar, descritas a seguir:

- a) **interdependência do preço do açúcar e do etanol**: os principais subprodutos da cana-de-açúcar são etanol e açúcar cristal. Algumas usinas de moagem de cana podem produzir as duas mercadorias. Dessa maneira, se a perspectiva do preço do etanol, ou do açúcar, estiver mais atraente no próximo período de safra, a usina pode alterar a proporção da produção de forma a maximizar a receita esperada. Com base nisso, assumiu-se a premissa de que uma pressão no preço do álcool, provocada por um

aumento na demanda pelo biocombustível, pode ser transferida para o preço do açúcar, e vice-versa, variações no preço do açúcar também podem provocar alterações persistentes no preço do etanol;

- b) **influência do preço do petróleo:** partiu-se da premissa de que o preço do álcool combustível pode ser influenciado pelo preço do barril de petróleo e, conseqüentemente, essa influência pode ser transferida para o preço do açúcar. Essa premissa tem como base o fato de os biocombustíveis serem produtos parcialmente substitutos dos derivados do petróleo. O retorno do preço do contrato futuro de petróleo WTI da NYMEX será considerada **variável exógena observável** no modelo e componente da média no processo de formação do preço à vista do açúcar;
- c) **preço à vista do açúcar:** o preço à vista do açúcar cristal será considerado variável não observável do modelo. Apesar de ser uma excelente referência de valor de mercado, o preço à vista do açúcar divulgado diariamente pelo CEPEA (2008) pode conter distorções decorrentes do fato de ser uma média de preços de negócios efetuados em localidades diferentes. O preço à vista do açúcar é considerado **variável de estado** do modelo proposto, cuja trajetória é descrita por um **movimento browniano**.
- d) **preço do contrato futuro de açúcar:** os preços do contrato futuro foram considerados **variáveis de sinal** observáveis no modelo, devido à transparência das negociações nos mercados futuro e à possibilidade de arbitragem entre os diferentes mercados futuros de açúcar. Considerou-se também que, em **condição de não arbitragem**, a diferença entre o preço à vista e o futuro é calculada pela diferença entre a taxa livre de risco e o retorno de conveniência;
- e) **retorno de conveniência:** conforme a teoria da estocagem, partiu-se da premissa de que os produtores têm a possibilidade de reter estoques em busca de melhores preços de venda. O retorno de conveniência é o benefício dos detentores de estoque físico, medido em termos de taxa de juros. Considerou-se que a probabilidade de escassez de mercadoria, provocada por fatores climáticos ou interrupções de fornecimento, faz com que, circunstancialmente, o retorno de conveniência seja máximo. Esse choque aleatório deverá ser corrigido pela habilidade dos participantes do mercado de

ajustarem seus estoques, revertendo novamente para uma média de longo prazo. Para o modelo, considerou-se o retorno de conveniência como **variável de estado**, que segue um processo de **reversão à média**;

- f) **influência da volatilidade nos mercados**: outra premissa relevante para a elaboração do modelo considera que um aumento na demanda por *commodities* pode ser provocada pela volatilidade nos mercados financeiros. Quando os mercados financeiros passam por momentos de turbulência provocados por cenários econômicos recessivos ou surtos inflacionários, ocorre um movimento de liquidez, causado pelo fluxo de recursos destinados às alternativas mais seguras de investimentos, em busca da preservação do valor real do capital. As *commodities* mantêm seu valor em qualquer moeda, independentemente das volatilidades nos mercados e da inflação nas economias dos países. A demanda pelas *commodities* pode provocar uma alta de curto prazo nos preços, que não é justificada pela demanda física do produto. No modelo proposto, a volatilidade é tratada como **variável exógena observável** cuja relação com o preço à vista do açúcar é medida por meio de um coeficiente linear contemporâneo;
- g) **presença de sazonalidade**: foi considerada a presença de sazonalidade na formação do preço à vista do açúcar, ocasionada pelos períodos de safra e entressafra da cana. Partiu-se da premissa de que existe uma relação negativa entre o volume de estoques e o preço do açúcar, por isso o componente sazonal é modelado no preço à vista do açúcar por meio de uma **função senoidal**.

3.5 Especificação do Modelo

O modelo proposto de formação de preços de *commodity* aplicado ao mercado de açúcar foi detalhadamente especificado com base nas premissas dispostas na seção anterior. O intuito, neste capítulo, é elaborar um modelo na forma de equações de estado, de forma que seja possível a estimação dos parâmetros por meio do filtro de Kalman. Para obter as equações necessárias para o processo de estimação, a especificação do modelo foi realizada em três etapas, descritas a seguir:

- primeiro, foram definidos os processos estocásticos das variáveis de interesse, conforme premissas preestabelecidas na seção anterior;
- em seguida, as equações em tempo contínuo foram transformadas para tempo discreto;
- por último, colocadas na forma de equações de estado para estimação por meio do filtro de Kalman.

As equações que formam o modelo são abordadas separadamente nas seções seguintes, em que os processos estocásticos são discutidos e algebricamente apresentados. Depois, na última seção deste capítulo, as equações são consolidadas em um sistema único de equações.

3.5.1 Condição de Não Arbitragem do Preço Futuro

O modelo proposto parte da premissa de que os preços no mercado futuro de *commodities* devem estar em **condição de não arbitragem** em relação aos preços no mercado à vista. Sugerir que os preços estão em condição de não arbitragem significa afirmar que nenhum participante do mercado consegue obter lucros anormais sem assumir riscos, simplesmente aproveitando diferenças de preços entre o mercado à vista e o futuro. Significa dizer também que, em condição de não arbitragem, a diferença entre o preço futuro e o preço à vista da *commodity* será apenas a diferença entre a taxa livre de risco e o retorno de conveniência no período.

Se o retorno de conveniência é maior, em termos nominais, do que a taxa livre de risco, o preço futuro ficará abaixo do preço à vista, resultando no fenômeno de *backwardation*, caso contrário, se a taxa do retorno de conveniência for menor do que a livre de risco, haverá a presença de *contango*.

Partindo-se da premissa de que o preço da *commodity* no mercado futuro é igual ao preço à vista acrescido da taxa de juros livre de risco menos o retorno de conveniência, a seguinte equação pode ser adotada no modelo, assumindo-se taxas de juros em forma contínua:

$$F_t = S_t e^{(r-\delta_t)(T-t)} \quad (3.5)$$

sendo:

S_t : preço do ativo à vista no tempo t ;

F_t : preço do contrato futuro no tempo t , com vencimento em T , posterior a t ;

r : taxa de juros do ativo livre de risco;

δ_t : taxa de retorno de conveniência no tempo t .

Após a aplicação do logaritmo dos dois lados de equação, o resultado fica da seguinte maneira:

$$\ln(F_t) = \ln(S_t) + (r - \delta_t)(T - t) \quad (3.6)$$

E finalmente, a equação que determina a relação de preços, fica nesta forma:

$$\ln(F_t) = r(T - t) + \ln(S_t) - \delta_t(T - t) \quad (3.7)$$

A equação 3.7 acima será utilizada posteriormente na consolidação do modelo. O mesmo procedimento será adotado em relação às equações que descrevem o preço à vista e o retorno de conveniência.

3.5.2 Reversão à Média do Retorno de Conveniência

O modelo utiliza uma abordagem semelhante à de Gibson e Schwartz (1990:960) para o tratamento do retorno de conveniência, cuja trajetória é descrita por um processo Ornstein-Uhlenbeck de reversão à média.

É importante reiterar que o retorno de conveniência é um conceito econômico não observável, por isso deve receber tratamento de variável de estado, cujos parâmetros são estimados no modelo. O retorno de conveniência é um **valor medido em termos de taxa de juros**, aplicada sobre o preço do ativo à vista. Tem como base a teoria da estocagem e pode ser entendido como uma vantagem econômica que o detentor da *commodity* física possui, por ter o benefício de poder esperar o melhor momento para vender sua mercadoria, a um preço que maximize sua receita. Em momentos nos quais os estoques globais da *commodity* estão baixos, o retorno de conveniência é máximo.

O processo de reversão à média, escolhido para descrever a variável, considera que desvios de curto prazo provocados por choques aleatórios não persistentes, são corrigidos por um fator de reversão que faz com que a variável convirja para uma média de longo prazo. De forma que, quanto maior seja o choque aleatório, mais rápido será o efeito de correção.

Esse processo é coerente para o retorno de conveniência no mercado de *commodities*. Supondo que ocorra uma falta momentânea de estoques provocada por um fenômeno aleatório não previsível, o detentor da *commodity* física será beneficiado pela distorção de curto-prazo nos preços, fazendo com que o retorno de conveniência seja máximo nesse momento. Nos momentos seguintes, à medida que os estoques sejam regulados novamente, o retorno de conveniência é corrigido para seus níveis médios de longo prazo.

O processo de reversão à média em tempo contínuo para o retorno de conveniência pode ser descrito pela seguinte equação:

$$d\delta_t = \phi(\mu_\delta - \delta_t)dt + \sigma_\delta \sqrt{dt}z \quad ,i.i.d. \quad (3.8)$$

sendo:

δ_t : taxa de retorno de conveniência no termo t ;

ϕ : velocidade de reversão à média do retorno de conveniência;

μ_δ : média de longo prazo do retorno de conveniência;

σ_δ : desvio-padrão do retorno de conveniência.

A equação acima, quando convertida para variações de tempo discretas, para $\Delta t = 1$, pode ser representada da seguinte maneira:

$$\delta_t - \delta_{t-1} = \phi(\mu_\delta - \delta_{t-1}) + \sigma_\delta z \quad \Delta t = 1 \quad (3.9)$$

A Equação 3.9 acima apresenta o processo de reversão à média do retorno de conveniência para uma unidade de tempo. O componente $\sigma_\delta z$ representa o choque aleatório que é corrigido no período subsequente.

Para obter uma equação na forma de espaço de estado, deve-se isolar o δ_t , obtendo-se:

$$\delta_t = \phi(\mu_\delta - \delta_{t-1}) + \delta_{t-1} + \sigma_\delta z \quad (3.10)$$

E finalmente, a equação de reversão à média do retorno de conveniência é apresentada da seguinte forma:

$$\delta_t = \phi\mu_\delta + (1-\phi)\delta_{t-1} + \sigma_\delta z \quad (3.11)$$

A equação acima será utilizada na especificação do modelo na forma de equações de estado. Na etapa seguinte, o processo estocástico do preço à vista é especificado e, em seguida, são incluídos o componente sazonal e a interdependência linear com o preço do petróleo e a volatilidade, o que determina a originalidade do presente trabalho.

3.5.3 Movimento Browniano do Preço à Vista

Em trabalho seminal, Brennan e Schwartz (1985:139) sugerem que a trajetória do preço à vista de uma *commodity* pode ser descrita por meio de um processo de passeio aleatório em tempo contínuo, ou seja, um movimento browniano. Para isso, é necessário assumir que os

retornos do preço à vista da *commodity* possuam distribuição normal, são independentes e identicamente distribuídos no tempo para qualquer instante de tempo t , $R_t^S \sim N(\mu_S, \sigma_S^2)$.

Na presente abordagem, considerou-se também o movimento browniano para descrever a trajetória dos preços à vista da *commodity*, cujo componente determinístico será composto pela relação linear com variáveis exógenas.

Conforme descrito anteriormente, a seguinte equação representa o movimento browniano adotado para descrever os preços da *commodity*:

$$R_t^S = \mu_S dt + \sigma_S \sqrt{dt} z \quad ,i.i.d. \quad (3.12)$$

sendo:

R_t^S : retorno do preço à vista da *commodity* no tempo t ;

μ_S : média dos retornos do preço à vista;

σ_S : desvio-padrão dos retornos do preço à vista.

Convertida para tempo discreto, para a variação de uma unidade de tempo $\Delta t = 1$, a equação de passeio aleatório para o retorno do preço fica da seguinte maneira:

$$R_t^S = \mu_S + \sigma_S z \quad , \Delta t = 1 \quad (3.13)$$

Substituindo-se o retorno contínuo pela diferença dos logaritmos dos preços,

$R_t^S = \ln(S_t) - \ln(S_{t-1})$, obtém-se:

$$\ln(S_t) - \ln(S_{t-1}) = \mu_S + \sigma_S z \quad (3.14)$$

O que permite, assim, isolar o logaritmo do preço à vista, conforme segue:

$$\ln(S_t) = \mu_S + \ln(S_{t-1}) + \sigma_S z \quad (3.15)$$

O componente determinístico μ_S , na Equação 3.15, não é apenas um parâmetro estimado no modelo, mas pode ser escrito através da relação linear com variáveis exógenas observáveis. Nas seções seguintes, o componente determinístico é discutido e as variáveis exógenas apresentadas.

3.5.4 Relação do Petróleo com o Preço da *Commodity*

O processo estocástico que descreve a trajetória dos retornos do preço à vista da *commodity* possui um componente determinístico μ_S , que define a tendência dos retornos, e um componente aleatório $\sigma_S z$. O componente determinístico μ_S pode não ser apenas um parâmetro estimado no modelo. Esse argumento parte da premissa de que existem outras variáveis que podem **antecipar informações sobre a tendência de preços à vista da *commodity***.

No caso das *commodities* agrícolas relacionadas com a produção de biocombustíveis, essa relação com o petróleo pode apresentar uma certa persistência no tempo. Ou seja, havendo um aumento nos preços do petróleo, provocado, por exemplo, por uma queda nos estoques globais, haverá uma alteração na relação de valores sugerida por Walras (2003:87). Essa relação pode ser total ou parcialmente restabelecida com o tempo, o que resultaria em consequente aumento nos preços da *commodity agrícola*, no caso, no preço do açúcar. Essa alteração na tendência dos preços pode estar relacionada a sua capacidade de tornar-se biocombustível, pois uma parte da produção da *commodity* pode ser destinada para suprir o déficit de energia causado pela escassez de petróleo.

Partindo dessa premissa, tem-se que a tendência μ_S dos retornos dos preços da *commodity* pode ser descrita por outras variáveis, dentre elas, os retornos do preço do petróleo com defasagem de uma unidade de tempo. Inicialmente, tem-se que o retorno do preço à vista da *commodity* segue um processo de passeio aleatório. Dado que o componente aleatório

$\sigma_S z = \varepsilon_t$ pode ser considerado um ruído branco, é possível descrever o processo do retorno para uma unidade de tempo de acordo com a seguinte equação:

$$R_t^S = \mu_S + \varepsilon_t \quad (3.16)$$

Partindo-se da premissa de que o componente determinístico pode ser expresso por meio de uma relação linear com variáveis exógenas observáveis, ou seja, sendo $\mu_S = \boldsymbol{\beta}' \mathbf{x}_t$, tem-se que:

$$R_t^S = \boldsymbol{\beta}' \mathbf{x}_t + \varepsilon_t \quad (3.17)$$

sendo:

\mathbf{x}_t : matrix $n \times t$ de variáveis exógenas observáveis;

$\boldsymbol{\beta}$: vetor $n \times 1$ de coeficientes lineares.

A Equação 3.17 acima pode ser escrita também da seguinte forma:

$$\begin{aligned} R_t^S &= \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \dots + \beta_n x_{n,t} + \varepsilon_t \\ R_{t+1}^S &= \beta_0 + \beta_1 x_{1,t+1} + \beta_2 x_{2,t+1} + \dots + \beta_n x_{n,t+1} + \varepsilon_{t+1} \\ &\dots \end{aligned} \quad (3.18)$$

Ao assumir que a média dos retornos da *commodity* pode ser descrita por uma constante mais o retorno defasado do petróleo, obtém-se a seguinte relação linear de retornos:

$$R_t^S = \beta_0 + \beta_1 R_{t-1}^P + \varepsilon_t \quad (3.19)$$

sendo:

R_t^S : retorno do preço à vista da *commodity* no tempo t ;

R_{t-1}^P : retorno do petróleo defasado em uma unidade de tempo, em $t-1$;

β_0, β_1 : coeficientes lineares;

ε_t : ruído branco.

Com base na relação acima, supõe-se que o retorno do petróleo pode antecipar informação sobre a trajetória do retorno de uma determinada *commodity* relacionada com a produção de biocombustível. Na próxima seção será discutida a influência da volatilidade dos mercados na formação de preços das *commodities* e como essa relação pode ser inserida ao modelo.

3.5.5 Relação da Volatilidade do Mercado com o Preço da *Commodity*

As *commodities*, por serem ativos reais e possuírem valor econômico na cadeia produtiva, podem ser utilizadas pelos investidores como alternativas seguras para preservação do valor real do capital. As *commodities*, por definição, são ativos que possuem o mesmo valor em qualquer moeda e podem ser, portanto, referência de valor.

Cenários econômicos turbulentos, provocados por surtos inflacionários, ou crises econômicas, podem provocar uma distorção momentânea no valor relativo das moedas. Nessas situações, os investidores acabam buscando a segurança do mercado de *commodities* para proteger seu capital. Essa corrida pela preservação do valor do capital provoca uma alta momentânea nos preços dessas mercadorias, que não está diretamente relacionada com a demanda física para utilização na cadeia produtiva. Essa demanda é provocada apenas pela migração do capital em busca de ativos reais de preservação de valor dos investimentos financeiros.

Com base nessa teoria, o presente trabalho sugere que, um aumento na volatilidade do mercado, resultante de uma crise econômica ou surto inflacionário, pode provocar uma alteração na tendência dos preços das *commodities*, devido ao fluxo de recursos destinados a investimentos dessa natureza. Essa premissa é reforçada pelos argumentos de Geman (2005:28), que sugere que um aumento na volatilidade provoca um aumento nos preços relativos das *commodities*, diferentemente do que acontece com o mercado de ações, no qual a volatilidade alta ocorre nos momentos de queda nos preços.

A busca pela segurança das *commodities* no mercado financeiro não é sustentável economicamente no longo prazo, pois, logo após o fim da volatilidade provocada pela crise econômica, os investidores retiram o capital e os preços da *commodity* tendem a voltar para patamares normais de equilíbrio entre oferta e demanda.

Portanto, considerou-se que o componente determinístico μ_s do retorno preço da *commodity* deve sofrer alterações persistentes no tempo provocadas pela volatilidade dos mercados. Essa relação pode ser captada pela adição de mais um coeficiente linear relacionado à volatilidade contemporânea. O retorno do preço à vista da *commodity*, portanto, pode ser descrito em função da primeira diferença da volatilidade e do retorno do petróleo, adicionado a um ruído branco, conforme a seguinte equação:

$$R_t^S = \beta_0 + \beta_1 R_{t-1}^P + \beta_2 \Delta \hat{\sigma}_t^M + \varepsilon_t \quad (3.20)$$

sendo:

$\Delta \hat{\sigma}_t^M$: primeira diferença da volatilidade condicional do mercado no tempo t ;

$\beta_0, \beta_1, \beta_2$: coeficientes lineares;

ε_t : ruído branco.

A volatilidade dos mercados, nessa equação, é o desvio-padrão dos retornos contínuos da carteira que representa todos os ativos do mercado. Como não é possível obter uma carteira com essas características, costuma-se utilizar como *proxy*, em estudos empíricos, um índice representativo do mercado de ações. Na seção seguinte, é discutida a adição de um componente sazonal ao modelo, de forma a captar os efeitos da safra nos preços da *commodity* agrícola.

3.5.6 Componente Sazonal

As *commodities* agrícolas sofrem oscilações de preço provocadas pelos períodos de safra e entressafra da lavoura. Isso ocorre porque, imediatamente antes dos períodos de safra, os estoques da *commodity* estão baixos, fazendo com que, de acordo com a teoria da estocagem, os detentores da mercadoria física obtenham melhores preços de venda do que em períodos imediatamente posteriores à safra. O efeito sazonal nos preços das *commodities* agrícolas não está relacionado com a volatilidade dos mercados ou com o petróleo, portanto, um componente deve ser adicionado ao modelo exclusivamente para descrever esse efeito cíclico na formação de preços.

Com base nessa teoria, foi adicionado um componente sazonal determinístico λ_t à equação que descreve o processo estocástico do preço à vista da *commodity*:

$$R_t^S = \beta_0 + \beta_1 R_{t-1}^P + \beta_2 \Delta \hat{\sigma}_t^M + \lambda_t + \varepsilon_t \quad (3.21)$$

λ_t : componente sazonal determinístico no tempo t ;

$\beta_0, \beta_1, \beta_2$: coeficientes lineares;

ε_t : ruído branco.

Existem, na literatura, diversas formas para modelar a sazonalidade, Hannan, Terrell e Tuckwell (1970:26), em seu trabalho, sugerem uma função senoidal para descrever o efeito sazonal em séries de tempo, a qual permite identificar ciclos, amplitude e frequência da sazonalidade durante o ano. Uma outra abordagem alternativa, também disponível na literatura, sugere a adição de variáveis *dummy* mensais, multiplicativas ou aditivas, ao modelo.

Porém, conforme Sorensen (2002:414), a função senoidal é mais indicada porque permite uma maior flexibilidade para lidar com o tempo na análise empírica; não importa se a amostra coletada está disposta em dados mensais ou diários, o modelo será relativamente o mesmo para qualquer intervalo de tempo.

O componente sazonal λ_t , descrito anteriormente, é representado conforme a seguinte função senoidal:

$$\lambda_t = \sum_{k=1}^K [\beta_{c,k} \cos(2\pi k \theta) + \beta_{s,k} \text{sen}(2\pi k \theta)] \quad (3.22)$$

sendo:

θ : tempo, medido em anos;

K : frequência dos ciclos sazonais anuais;

$\beta_{c,k}, \beta_{s,k}$: coeficientes estimados para cada um dos K ciclos.

A equação acima permite identificar a amplitude do efeito sazonal nos preços dos períodos de safra das culturas agrícolas. Em alguns casos, a *commodity* pode ter mais do que um período de safra durante o ano. Nesses casos, quando a série de dados possui mais do que um ciclo, é possível ajustar o modelo aumentando o número de K parâmetros estimados no somatório.

A variável t , que denota a unidade de tempo no modelo de formação de preços de *commodities*, pode ser computada em termos de qualquer unidade de tempo, diferentemente da variável θ do componente sazonal, que sempre será dada em termos anuais. Esse recurso é útil para os estudos empíricos, pois os parâmetros do modelo podem ser estimados conforme a disponibilidade temporal dos dados da amostra.

Com base nos pressupostos aqui estabelecidos, tem-se que o componente determinístico do preço à vista pode ser descrito pelo retorno defasado do petróleo, pela volatilidade dos mercados e pelo efeito cíclico da sazonalidade, conforme a seguinte equação:

$$\mu_s = \beta_0 + \beta_1 R_{t-1}^P + \beta_2 \Delta \hat{\sigma}_t^M + \lambda_t \quad (3.23)$$

Conforme a Equação 3.15, tem-se que logaritmo do preço à vista deve ser dado da seguinte forma:

$$\ln(S_t) = \mu_S + \ln(S_{t-1}) + \sigma_S z \quad (3.24)$$

Sendo assim, a forma final da equação que descreve o processo estocástico do preço à vista de uma *commodity*, é representada pela seguinte equação:

$$\ln(S_t) = \beta_0 + \beta_1 R_{t-1}^P + \beta_2 \Delta \hat{\sigma}_t^M + \lambda_t + \ln(S_{t-1}) + \sigma_S z \quad (3.25)$$

Na próxima seção, a equação acima e as demais discutidas neste capítulo serão dispostas na forma de espaço de estado, permitindo especificar o modelo em um único sistema de equações.

3.5.7 Equações na Forma de Espaço de Estado

O modelo, na forma de espaço de estado, é composto por equações de estado e equações de sinal, conforme a seguinte estrutura:

$$\begin{cases} \mathbf{y}_t = \mathbf{A} + \mathbf{B}\mathbf{x}_t + \boldsymbol{\varepsilon}_t \\ \mathbf{x}_t = \mathbf{C} + \mathbf{D}\mathbf{x}_{t-1} + \mathbf{v}_t \end{cases} \quad (3.26)$$

sendo:

\mathbf{x}_t : vetor $n \times 1$ de variáveis de estado;

\mathbf{y}_t : vetor $m \times 1$ de variáveis de sinal;

\mathbf{A}, \mathbf{C} : vetores estimados no processo, incluindo os parâmetros relativos às variáveis exógenas;

\mathbf{B}, \mathbf{D} : matrizes estimadas no processo, referentes aos coeficientes lineares das variáveis de estado;

$\boldsymbol{\varepsilon}_t, \mathbf{v}_t$: vetores de ruídos brancos serialmente independentes.

Para especificar o modelo de formação de preços de *commodities* na forma de espaço de estado, é necessário adequar os processos estocásticos das variáveis analisadas em tempo discreto. O modelo deverá seguir a seguinte estrutura:

- a) **as equações de sinal** devem ser especificadas em função do vetor \mathbf{x}_t de variáveis de estado em sua forma contemporânea, sem defasagem, cujos parâmetros farão parte da matriz \mathbf{B} . Se forem adicionadas variáveis exógenas às equações de sinal, essas variáveis, e os parâmetros relacionados a elas, estarão no vetor \mathbf{A} . As variáveis exógenas podem ser defasadas, ou não;
- b) **as equações de estado** devem ser especificadas em função do vetor de variáveis de estado com defasagem de uma unidade de tempo \mathbf{x}_{t-1} , os parâmetros relacionados a esse vetor estão na matriz \mathbf{D} . As equações de estado também podem ser escritas em função de variáveis exógenas, defasadas ou não, cujas variáveis e seus parâmetros devem ser alocados no vetor \mathbf{C} .

Partindo da estrutura acima descrita, é possível especificar o modelo de formação de preços de *commodities* na forma de espaço de estado, utilizando um sistema com três equações em tempo discreto, que definem os processos estocásticos das variáveis, conforme segue:

$$\begin{cases} \ln(F_t) = r(T-t) - (T-t)\delta_t + \ln(S_t) \\ \delta_t = \phi\mu_\delta + (1-\phi)\delta_{t-1} + \sigma_\delta z \\ \ln(S_t) = \beta_0 + \beta_1 R_{t-1}^P + \beta_2 \Delta \hat{\sigma}_t^M + \lambda_t + \ln(S_{t-1}) + \sigma_S z \end{cases} \quad (3.27)$$

A primeira equação é de sinal, que define a condição de não arbitragem para o logaritmo do preço futuro da *commodity*. Essa equação possui a taxa livre de risco r como variável exógena. O logaritmo do preço à vista e o retorno de conveniência são as variáveis de estado na forma contemporânea. Não há componente aleatório.

A segunda é uma equação de estado, que define o processo de reversão à média da taxa do retorno de conveniência, variável de estado. Possui um componente aleatório, cuja variância é representada por σ_δ^2 . Essa equação não possui variáveis exógenas, apenas parâmetros estimados no processo.

A terceira, e principal equação do modelo, é uma equação de estado, que representa o movimento browniano do logaritmo do preço à vista, variável de estado cujo processo especificado determina a originalidade do presente estudo. Na equação estão presentes as **variáveis exógenas** que descrevem a tendência do retorno preço à vista, que são: o retorno defasado do petróleo, a primeira diferença da volatilidade do mercado e o componente sazonal. Essa equação possui também o logaritmo do preço à vista com defasagem e um componente aleatório, que possui variância representada por σ_s^2 .

O sistema de equações acima também pode ser representado na forma matricial, sendo seus componentes dados por:

$$\mathbf{y}_t = [\ln(F_t)]; \quad \mathbf{x}_t = \begin{bmatrix} \delta_t \\ \ln(S_t) \end{bmatrix}; \quad \mathbf{x}_{t-1} = \begin{bmatrix} \delta_{t-1} \\ \ln(S_{t-1}) \end{bmatrix};$$

$$\mathbf{A} = [r(T-t)]; \quad \mathbf{B} = [-(T-t) \quad 1]; \quad \mathbf{C} = \begin{bmatrix} \phi\mu_\delta \\ \beta_0 + \beta_1 R_{t-1}^p + \beta_2 \Delta \hat{\sigma}_t^M + \lambda_t \end{bmatrix} \quad (3.28)$$

$$\mathbf{D} = \begin{bmatrix} (1-\phi) & 0 \\ 0 & 1 \end{bmatrix}; \quad \boldsymbol{\varepsilon}_t = [0]; \quad \mathbf{v}_t = \begin{bmatrix} \sigma_{\delta Z} \\ \sigma_{s Z} \end{bmatrix}$$

A forma matricial é utilizada na estimação dos parâmetros com o filtro de Kalman. No capítulo seguinte, é apresentada a metodologia do presente estudo para a aplicação do modelo proposto.

4 METODOLOGIA DE PESQUISA

*O cientista vive num mundo onde a verdade é inatingível,
mas onde sempre é possível encontrar erros
no que foi pensamente estabelecido ou no óbvio.*

(Rudolf Flesch, 1951)

Com relação à metodologia científica, este estudo pode ser classificado como **hipotético-dedutivo**, seguindo a linha de pensamento de Popper (1975:346). Tal método consiste na eliminação de erros do conhecimento científico por meio de tentativa de falseamento de hipóteses. As hipóteses do trabalho científico são elaboradas a partir do estudo de um problema que tem base em teorias e conhecimentos preestabelecidos.

Pelo método hipotético-dedutivo, o pesquisador delimita um problema de pesquisa, uma região do conhecimento científico e questiona suas teorias preestabelecidas, com o intuito de falseá-las, ou confirmá-las, por meio de testes estatísticos de hipóteses. O conhecimento científico, pelo método hipotético-dedutivo, é falível, é provisório, ou seja, não existe *episteme* ou conhecimento definitivo. O conhecimento científico está sujeito a constantes tentativas de refutação e falseamento.

O método de um estudo científico pode ser também classificado como indutivo ou dedutivo. Porém, seria uma atitude extremada, por parte do pesquisador, classificar este estudo como puramente de uma ou outra categoria, porque, na verdade, ele possui simultaneamente ambas as características, é o que Cooper e Schindler (2003:48) classificam como movimento duplo do pensamento científico.

A indução, segundo Marconi e Lakatos (2000:73), consiste em admitir, ou generalizar um fato, sem necessidade de ser demonstrado ou justificado. O pensamento indutivo permite ao pesquisador fazer inferências e generalizações sobre o todo a partir da observação de particularidades. O pensamento dedutivo, por outro lado, consiste em chegar a uma conclusão a partir da verificação de premissas previamente elaboradas, consiste, portanto, em tirar conclusões sobre particularidades a partir da observação do todo. Segundo Marconi e Lakatos (2000:64), “(...) os argumentos indutivos aumentam o conteúdo das premissas, ao passo que os argumentos dedutivos sacrificam a ampliação do conteúdo para atingir a certeza (...)”.

Este trabalho possui características dedutivas porque, partindo do teste estatístico de hipóteses, chega a conclusões sobre o problema que está sendo investigado e, com base em uma pequena amostra observada, são feitas generalizações em relação à população, de uma forma indutiva.

Em relação à abordagem epistemológica, o método adotado neste estudo é empírico-analítico, conforme classificação sugerida por Martins (1994:26). O falseamento das hipóteses desta pesquisa foi feito a partir de testes estatísticos de probabilidade, que inclui coleta, tratamento e análise de dados, além de verificar a relação causal entre as variáveis. A validação da prova científica é buscada através de testes de coeficientes com graus de significância, características do método empírico-analítico.

A Ilustração 2, a seguir, apresenta o esquema do método científico utilizado nesta pesquisa. Analisando os modelos e teorias disponíveis na literatura, foi identificado um problema de pesquisa e, a partir desse problema, foram feitas conjecturas dispostas na forma de hipóteses. Depois dos testes estatísticos dessas hipóteses, na tentativa de falseamento, os resultados foram generalizados e uma nova teoria foi elaborada:

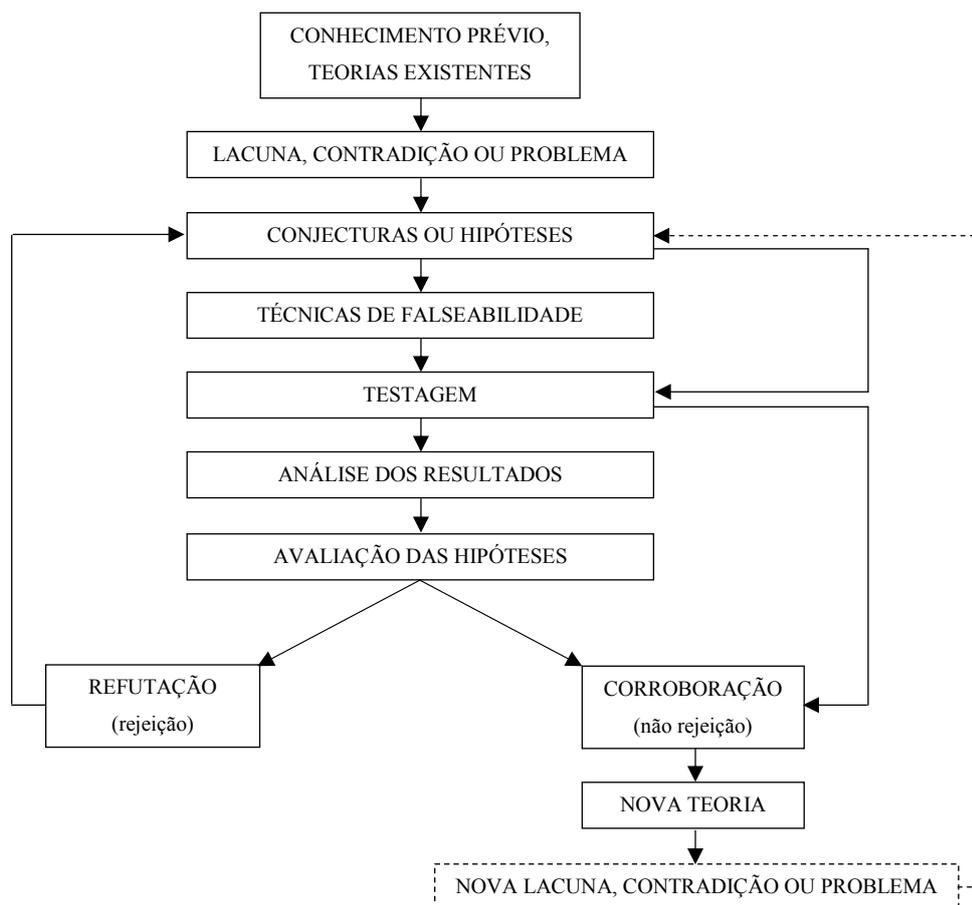


Ilustração 2 - Esquema metodológico da pesquisa científica

FONTE: Adaptado de Marconi e Lakatos (2000:75).

4.1 Hipóteses de Pesquisa

A formulação e teste de hipóteses permite a verificação do conhecimento científico. Martins (1994:33) define a formulação de hipóteses da seguinte maneira:

As hipóteses são uma ferramenta poderosa para o avanço do conhecimento porque, embora formuladas pelo homem, podem ser testadas e mostradas como provavelmente corretas ou incorretas independentemente dos valores e crenças do homem.

A aceitação ou rejeição das hipóteses permite atingir o objetivo proposto pelo trabalho científico, desde que o conhecimento científico possua a característica de verificabilidade.

Para o desenvolvimento do modelo de formação de preços de *commodities* – objetivo central deste estudo científico – foram elaboradas premissas, verificadas por meio de testes estatísticos de hipóteses. A partir dessas premissas, foram feitas generalizações para a especificação do modelo, cujos parâmetros foram estimados, aplicado ao mercado de brasileiro de açúcar. Finalmente, foram realizados testes estatísticos sobre séries de previsão, com o intuito de verificar a hipótese central da pesquisa: saber se o modelo proposto é mais eficiente, em termos de previsão, do que outro modelo disponível na literatura.

Para isso, os testes de hipóteses foram segregados em duas etapas, antes e depois da especificação do modelo proposto:

a) Verificação das premissas do modelo:

Pergunta 1: Existe relação serial linear na formação de preços entre o açúcar e o etanol de cana-de-açúcar?

H_{0P1} – não existe relação linear na formação de preços entre o açúcar e o etanol.

H_{1P1} – hipótese alternativa: existe relação linear persistente na formação de preços entre o açúcar e o etanol, e o etanol Granger causa os preços do açúcar.

H_{2P1} – hipótese alternativa: existe relação linear persistente na formação de preços entre o açúcar e o etanol, e o açúcar Granger causa os preços do etanol.

Pergunta 2: Existe *backwardation* persistente na formação de preços dos contratos futuros de açúcar?

H_{0P2} – não existe *backwardation* nos preços dos contratos futuros de açúcar.

H_{1P2} – hipótese alternativa: existe *backwardation* persistente e fraco nos preços dos contratos futuros de açúcar.

H_{2P2} – hipótese alternativa: existe *backwardation* persistente e forte nos preços dos contratos futuros de açúcar.

Pergunta 3: Existe persistência de relação serial linear na formação de preços entre o açúcar e o petróleo?

H_{0P3} – não existe relação na formação de preços entre o açúcar e o petróleo.

H_{1P3} – hipótese alternativa: existe relação linear contemporânea entre os preços do açúcar e o petróleo, e a trajetória de preços do petróleo não antecipa informações sobre o preço do açúcar.

H_{2P3} – hipótese alternativa: existe relação linear serial entre os preços do açúcar e o petróleo, e a trajetória de preços do petróleo antecipa informações sobre o preço do açúcar.

Pergunta 4: Existe relação linear persistente entre a volatilidade do mercado e a formação de preços do açúcar?

H_{0P4} – não existe relação na formação de preços entre o açúcar e a volatilidade do mercado.

H_{1P4} – hipótese alternativa: existe relação linear persistente entre os preços do açúcar e a volatilidade do mercado.

Pergunta 5: Existe sazonalidade na formação de preços do açúcar?

H_{0P5} – não existe sazonalidade significativa na formação de preços do açúcar.

H_{1P5} – hipótese alternativa: existe sazonalidade significativa na formação de preços do açúcar.

b) Aplicação do modelo:

Pergunta Final: A capacidade de previsão do modelo proposto é melhor, estatisticamente, do que outro modelo disponível na bibliografia?

H_{0F} – não existe diferença na capacidade de previsão entre o modelo proposto e outro modelo disponível.

H_{1F} – hipótese alternativa: o modelo proposto não é melhor, em termos de previsão, do que outro modelo disponível.

H_{2F} – hipótese alternativa: o modelo proposto é melhor, em termos de previsão, do que outro modelo disponível.

4.2 Tratamento Estatístico e Análise dos Dados

O estudo estatístico dos dados foi dividido em três partes inter-relacionadas. Na primeira, foram verificadas as hipóteses de pesquisa e as premissas fundamentais para a aplicação do modelo. As hipóteses foram verificadas por meio de testes estatísticos de coeficientes de regressões lineares com níveis de significância e testes t de média das variáveis. Na segunda parte, os parâmetros do modelo, na forma de espaço de estado, foram estimados por meio do filtro de Kalman. Na terceira parte da pesquisa, a capacidade de previsão do modelo proposto foi comparada com a de outro modelo selecionado na literatura que melhor descreve a trajetória dos preços no mercado de açúcar e os dois modelos foram testados por meio de medidas que quantificam o erro de previsão. As três partes do estudo estatístico serão explicitadas nas próximas seções.

4.2.1 Parte 1 – Verificação das Premissas do Modelo

Nesta parte foram testadas as premissas para a aplicação do modelo utilizando o preço à vista observado do açúcar como variável dependente. O intuito foi tentar responder às cinco primeiras perguntas do processo de investigação. Nesta etapa, foram realizados testes t de média e testes dos coeficientes estimados de regressões lineares múltiplas, pelo método MQO,

Mínimos Quadrados Ordinários. Para responder às perguntas, são analisadas, a seguir, as relações entre as variáveis.

Testes da pergunta 1: Para verificar a relação entre os preços à vista observados do açúcar e do etanol de cana-de-açúcar, foi realizado o teste causalidade de Granger (1969:424)¹ entre as duas variáveis. O teste consiste na verificação da estatística F para a hipótese conjunta de igualdade a zero dos coeficientes de uma regressão linear com defasagem. O intuito do teste é verificar se o açúcar Granger causa o etanol, ou o contrário, se o etanol Granger causa o açúcar. É importante salientar que o teste de causalidade de Granger não verifica a relação causal entre as variáveis, no sentido literal, mas apenas se uma das variáveis pode antecipar a trajetória da outra, supondo ausência de relações espúrias. O teste foi realizado para as duas Equações 4.1, tendo o etanol como variável dependente na primeira equação e o açúcar à vista como dependente na segunda equação. A relação entre as variáveis que se deseja investigar é definida pelos coeficientes das equações (a) e (b):

$$\begin{aligned} R_t^E &= \sum_{j=1}^d \alpha_j R_{t-j}^E + \sum_{j=1}^d \beta_j R_{t-j}^S + \varepsilon_t \quad (\text{a}) \\ R_t^S &= \sum_{j=1}^d \alpha_j R_{t-j}^S + \sum_{j=1}^d \beta_j R_{t-j}^E + \varepsilon_t \quad (\text{b}) \end{aligned} \quad (4.1)$$

sendo:

R_{t-j}^E : retorno do preço à vista do etanol no tempo $t-j$;

R_{t-j}^S : retorno do preço à vista do açúcar no tempo $t-j$;

d : defasagem das variáveis;

$\beta_1, \beta_2, \dots, \beta_d$: coeficientes lineares da regressão relativos à variável explicativa;

¹ Um maior detalhamento sobre o teste de causalidade de Granger está disponível no Apêndice II.

$\alpha_1, \alpha_2, \dots, \alpha_d$: coeficientes lineares da regressão relativos à variável dependente;

ε_t : ruído branco no tempo t .

A estatística F do teste de causalidade de Granger verifica a hipótese conjunta de igualdade a zero dos coeficientes β_j das regressões (a) e (b), conforme segue:

H_{0a} – R_t^S não Granger causa R_t^E se $\beta_1 = \beta_2 = \dots = \beta_d = 0$ na equação (a);

H_{0b} – R_t^E não Granger causa R_t^S se $\beta_1 = \beta_2 = \dots = \beta_d = 0$ na equação (b).

Espera-se que os coeficientes sejam significativamente diferentes de zero em pelo menos uma das equações. Foram verificadas as seguintes hipóteses de pesquisa nos testes, para responder à **Pergunta 1** do processo de investigação:

H_{0P1} – não existe relação causal entre as variáveis, ou seja, não se rejeita H_{0a} e H_{0b} ;

H_{1P1} – o etanol Granger causa o açúcar, logo rejeita-se H_{0a} ;

H_{2P1} – o açúcar Granger causa o etanol, logo, rejeita-se H_{0b} .

Testes da pergunta 2: Para verificar a presença de *backwardation* nos preços dos contratos futuros de açúcar foi utilizada a técnica estatística do teste t unicaudal sobre a média de uma distribuição Normal com variância desconhecida, sugerida por Bussab e Moretin (2004:347). Esse teste verifica se a média μ_B do *backwardation* na amostra é maior do que zero, supondo que a média da amostra possui distribuição normal $\bar{B} \sim N(\mu_B, \sigma_B^2/n)$. Se os resultados dos testes confirmarem a presença de *backwardation* na amostra, indicam também que o retorno de conveniência possui uma média maior do que a da taxa livre de risco, conforme a Teoria da Estocagem. Para a realização dos testes t , foram calculadas, para as N observações da amostra, as médias do *backwardation* fraco B^w e forte B^s , e as respectivas variâncias amostrais, conforme as seguintes equações:

$$\begin{aligned}\bar{B}^w &= \frac{1}{N} \sum_{j=1}^N S_{t,j} - F_{t,j} e^{-r(T-t)}, & S_{B^w}^2 &= \frac{1}{N-1} \sum_{j=1}^N (B_j^w - \bar{B}^w)^2 \\ \bar{B}^s &= \frac{1}{N} \sum_{j=1}^N S_{t,j} - F_{t,j}, & S_{B^s}^2 &= \frac{1}{N-1} \sum_{j=1}^N (B_j^s - \bar{B}^s)^2\end{aligned}\tag{4.2}$$

sendo:

$S_{t,j}$: preço do ativo à vista no tempo t , referente à j -ésima observação;

$F_{t,j}$: preço do contrato futuro no tempo t , referente à j -ésima observação, com vencimento em T , posterior a t ;

r : taxa de juros livre de risco;

\bar{B}^w, \bar{B}^s : médias da amostra de *backwardation* fraco e forte;

$S_{B^w}^2, S_{B^s}^2$: variâncias da amostra de *backwardation* fraco e forte.

Se a amostra possui *backwardation* forte e significativamente positivo, terá também, consequentemente, *backwardation* fraco e positivo, pois uma hipótese sobrepõe-se à outra. Pela estatística t , as médias foram testadas para verificar a hipótese nula e as hipóteses alternativas, conforme segue:

$H_{0P2} - \mu_{B^w} \leq 0$, não existe *backwardation*;

$H_{1P2} - \mu_{B^w} > 0$, existe *backwardation* fraco;

$H_{2P2} - \mu_{B^s} > 0$, existe *backwardation* forte.

Testes da pergunta 3: Para investigar a existência de relação serial entre os preços do açúcar e do petróleo, foi utilizada a técnica estatística de regressão linear múltipla do tipo seção

cruzada, com e sem defasagem temporal, entre os retornos do petróleo WTI da Nymex e os retornos dos preços observados do açúcar à vista, divulgados pelo CEPEA. Para testar a relação entre as variáveis, foram realizados testes t bicaudais de igualdade a zero nos estimadores dos coeficientes da regressão, método sugerido por Hair *et al* (2005:32) para esses tipos de variáveis. Para identificar a existência de defasagem, utilizou-se um correlograma serial cruzado entre variáveis. A relação que se pretende investigar é determinada pelos coeficientes da seguinte equação²:

$$R_t^S = \beta_1 R_{t-d}^P + \dots + \varepsilon_t \quad (4.3)$$

sendo:

R_t^S : retorno do preço à vista do açúcar no tempo t ;

R_{t-d}^P : retorno do petróleo defasado em d unidades de tempo;

β_1 : coeficiente linear da regressão múltipla;

ε_t : ruído branco.

Com os resultados dos testes t , espera-se que o coeficiente β_1 seja significativo, d seja maior do que zero e o sinal de β_1 seja positivo, o que poderia indicar que o petróleo antecipa informação sobre a tendência do preço do açúcar. As hipóteses testadas sobre o coeficiente ficaram da seguinte forma:

$H_{0P3} - \beta_1 = 0$, o preço do petróleo não tem relação com o preço do açúcar;

$H_{1P3} - \beta_1 \neq 0$, para $d = 0$, tem relação, porém não antecipa informação;

² Os “...” na equação indicam que outras variáveis serão adicionadas na especificação da regressão múltipla. Essas variáveis são discutidas mais adiante nesta seção.

$H_{2P3} - \beta_1 \neq 0$, para $d > 0$, tem relação e antecipa informação na tendência do preço do açúcar.

Testes da pergunta 4: para investigar a existência de relação positiva entre a volatilidade do mercado e os preços observados do açúcar à vista, foi utilizada a técnica estatística de regressão linear múltipla do tipo seção cruzada sem defasagem. Para verificar a relação entre as variáveis, foram realizados testes t de igualdade a zero dos estimadores dos coeficientes da regressão. A relação que se pretende investigar é determinada pelos coeficientes da seguinte equação:

$$R_t^S = \beta_1 R_{t-d}^P + \beta_2 \Delta \hat{\sigma}_t^M + \dots + \varepsilon_t \quad (4.4)$$

sendo:

$\Delta \hat{\sigma}_t^M$: primeira diferença da volatilidade condicional do mercado, no tempo t ;

β_1, β_2 : coeficientes lineares da regressão múltipla;

ε_t : ruído branco.

Nos resultados dos testes t , espera-se que os coeficientes sejam significantes e o sinal de β_2 seja positivo, o que indicaria que o preço do açúcar se eleva com o aumento da volatilidade do mercado. A hipótese nula e as hipóteses alternativas ficaram da seguinte forma:

$H_{0P4} - \beta_2 = 0$, não existe relação;

$H_{1P4} - \beta_2 \neq 0$, existe relação da volatilidade com a tendência do preço do açúcar.

A volatilidade não é uma variável diretamente observável, portanto deve ser calculada. Optou-se por estimar a volatilidade condicional por meio de um processo ARMA / GARCH³ de um índice representativo do mercado de ações. Esse método de estimação será apresentado em seção posterior neste capítulo

Testes da pergunta 5: a hipótese de presença de sazonalidade na amostra de preços do açúcar, decorrente da safra da cana-de-açúcar, foi investigada por meio de testes estatísticos de igualdade a zero dos estimadores dos coeficientes sazonais seno e cosseno resultantes de regressão linear múltipla, tendo o retorno do preço do açúcar como variável dependente. Os testes do componente sazonal λ_t permitem a identificação de ciclos nos preços do açúcar. Para verificar a hipótese de sazonalidade foram testados os coeficientes da seguinte equação:

$$R_t^S = \sum_{k=1}^K [\beta_{c,k} \cos(2\pi k\theta) + \beta_{s,k} \sin(2\pi k\theta)] + \dots + \varepsilon_t \quad (4.5)$$

Considerou-se $K = 1$, em decorrência do fato de a cana-de-açúcar apresentar apenas um ciclo de safra predominante no Brasil. Portanto, a equação utilizada para o teste de presença de sazonalidade ficou da seguinte forma:

$$R_t^S = \beta_c \cos(2\pi\theta) + \beta_s \sin(2\pi\theta) + \dots + \varepsilon_t \quad (4.6)$$

sendo:

θ : tempo, medido em anos;

β_c, β_s : coeficientes lineares do cosseno e do seno.

Para confirmar a presença de sazonalidade nos preços do açúcar, pelo menos um dos coeficientes, β_c ou β_s , deve ser estatisticamente diferente de zero e não existe um sinal

³ O processo ARMA / GARCH utilizado nos testes é discutido no item 4.2.3 deste capítulo.

esperado para os coeficientes. Portanto, a hipótese de pesquisa ficaria da seguinte forma, utilizando os coeficientes da regressão:

$H_{0P5} - \beta_c = \beta_s = 0$, não existe sazonalidade;

$H_{1P5} - \beta_c \neq 0$ ou $\beta_s \neq 0$, existe sazonalidade nos preços do açúcar.

Os testes referentes às perguntas 3, 4 e 5 foram realizados por meio da técnica estatística de análise dos coeficientes de uma regressão linear múltipla, tendo como variáveis explicativas: o retorno do petróleo, a volatilidade e o componente sazonal; e como variável dependente: o retorno do preço à vista do açúcar divulgado pelo CEPEA. Hair *et al* (2005:34) defendem a utilização de técnicas estatísticas que envolvem regressões múltiplas:

Regressão múltipla é o método de análise apropriado quando o problema de pesquisa envolve uma única variável dependente métrica considerada relacionada a duas ou mais variáveis independentes métricas. O objetivo da análise de regressão é prever as mudanças na variável dependente como resposta a mudanças nas variáveis independentes. Esse objetivo é alcançado com frequência, por meio da regra estatística dos mínimos quadrados.

Logo, para verificar as hipóteses referentes às perguntas 3, 4 e 5, foram realizados testes t dos estimadores dos coeficientes da regressão linear múltipla das variáveis do modelo. A equação completa da regressão é a seguinte:

$$R_t^S = \beta_1 R_{t-d}^P + \beta_2 \Delta \hat{\sigma}_t^M + \beta_c \cos(2\pi\theta) + \beta_s \sin(2\pi\theta) + \varepsilon_t \quad (4.7)$$

sendo:

$\beta_1, \beta_2, \beta_c, \beta_s$: coeficientes estimados pela regressão MQO, utilizados para os testes de significância e falseamento de hipóteses.

Durante processo de estimação dos coeficientes do modelo de regressão, alguns procedimentos foram adotados para garantir que os estimadores MQO sejam consistentes e não enviesados. Para isso, os seguintes testes foram adotados:

- testes de raiz unitária e cointegração das variáveis, antes de proceder à regressão;
- teste Jarque-Bera de normalidade dos resíduos resultantes da regressão MQO para amostras finitas, das séries de dados com periodicidade mensal.⁴ Para amostra com dados diários, foram consideradas as propriedades assintóticas dos estimadores MQO para grandes amostras. Segundo Greene (2003:65), é possível obter estimadores não enviesados em regressões com grandes amostras e a premissa de normalidade dos resíduos pode ser relaxada;
- teste de correlação serial dos resíduos, com estatística Q *Ljung-Box* dos coeficientes de correlação.

4.2.2 Teste de Raiz Unitária e Cointegração

Antes de realizar as regressões múltiplas e antes da estimação dos parâmetros do modelo pelo filtro de Kalman, foram realizados testes nas variáveis para verificar presença de raiz unitária e testes de cointegração, seguindo os procedimentos de Engle e Granger (1987:252). As variáveis utilizadas para a estimação dos parâmetros do modelo devem ser estacionárias, os resíduos da regressão devem ter média igual a zero, variância finita e não podem apresentar correlação serial ou tendência. O teste de raiz unitária adotado foi o ADF – Dickey e Fuller (1979:428) aumentado, enquanto o teste para verificar se as variáveis do modelo cointegram foi baseado no método de Johansen (1995:80).

⁴ Detalhes sobre o teste Jarque-Bera estão no Apêndice I.

A presença de séries não estacionárias no processo de estimação de parâmetros pode gerar uma regressão espúria, que pode ter um R^2 elevado, e os testes t aparentarem ser significantes, porém os resultados são inconsistentes. Variáveis não estacionárias podem ser transformadas, segundo Enders (2004:164), por meio da diferenciação, para eliminação de tendência. As variáveis utilizadas no modelo foram transformadas em primeira diferença, entretanto, mesmo sendo estacionárias em primeira diferença, as variáveis podem cointegrar. Por isso, além do teste de raiz unitária para nível, primeira e segunda diferenças, foram realizados também testes de cointegração.

O procedimento para identificar cointegração, seguindo a metodologia de Engle e Granger (1987:252), consiste em, primeiro, identificar a ordem de integração das variáveis por meio do teste ADF. No teste de raiz unitária, se a variável analisada y_t possui tendência não estacionária, ou seja, seja integrada de ordem um $I(1)$, a hipótese nula de que o coeficiente $\alpha_1 = 1$ será rejeitada, conforme a seguinte equação:

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t \quad (4.8)$$

sendo:

y_t : variável na qual se deseja identificar presença de raiz unitária.

Ao identificar a presença de pelo menos uma raiz unitária nas variáveis, a segunda etapa do processo consiste em analisar os resíduos da regressão MQO das duas variáveis, para identificar a presença de cointegração:

$$y_t = \phi_0 + \phi_1 x_t + e_t \quad (4.9)$$

sendo:

y_t, x_t : variáveis que podem ser cointegradas.

Se os resíduos e_t da regressão forem estacionários, é um indicativo de que as variáveis y_t e x_t cointegram. Para isso, deve-se realizar um teste de raiz unitária também nos resíduos, conforme a equação:

$$\Delta e_t = \varphi_1 e_t + \xi_t \quad (4.10)$$

Se a hipótese de que $\varphi_1 = 0$ não for rejeitada, é possível concluir que a série de resíduos possui pelo menos uma raiz unitária e que, portanto, as variáveis y_t e x_t não cointegram. A hipótese nula do teste de cointegração é a seguinte:

H_0 – as variáveis y_t e x_t não cointegram entre si.

O objetivo da pesquisa é verificar que as variáveis do modelo: preço do petróleo, preço do açúcar e volatilidade, não cointegram, portanto a hipótese nula não deve ser rejeitada. Em testes cujo objetivo é aceitação da hipótese nula, o pesquisador está sujeito à ocorrência do erro estatístico do tipo II, ou seja, de não rejeição de uma hipótese, que na verdade é inválida.⁵

4.2.3 Volatilidade Condicional GARCH

A volatilidade do mercado é variável de entrada do modelo e precisa ser estimada, pois não é diretamente observável. Existem diversas maneiras de calcular a volatilidade, Galdi e Pereira (2007:74) sugerem três diferentes formas: Volatilidade Estocástica, EWMA (*Exponential Weighted Moving Average*) e GARCH (*Generalized Autoregressive Conditional Heteroskedasticity Model*).

Optou-se pelo GARCH, pois, de acordo com Enders (2004:118), ao especificar um GARCH, a volatilidade é modelada como se fosse um processo ARMA, *Autoregressive Moving Average Model*, possibilitando que o modelo seja mais parcimonioso.

⁵ As limitações do teste de cointegração são discutidas posteriormente, no item 4.5 neste capítulo.

Para poder adotar um processo GARCH, é necessário testar a hipótese de que a volatilidade do mercado siga um processo **heterocedástico**, ou seja, que a variância possa ser diferente para qualquer tempo t . O procedimento utilizado para identificar a presença de processo GARCH na série foi a aplicação **estatística Q Ljung-Box** no correlograma dos resíduos quadráticos resultantes de uma regressão ARMA dos retornos de um índice representativo do mercado de ações, neste estudo, optou-se pelo Índice Dow Jones.⁶

Na Ilustração 3, a seguir, com a representação gráfica dos retornos do Índice Dow Jones, é possível identificar períodos de maior e menor turbulência.

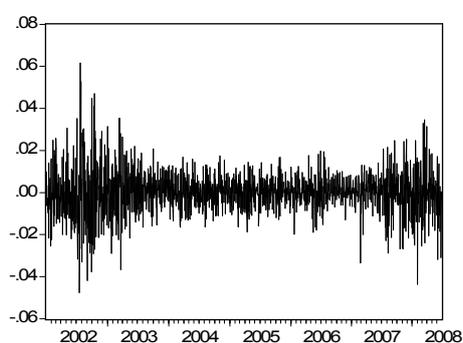


Ilustração 3 - Retornos diários do Índice Dow Jones no período de jan./2002 a jun./2008.

- **Estimando o Modelo ARMA**

Um processo GARCH é utilizado em séries econômicas que apresentem volatilidade elevada, seguida por períodos de tranquilidade. Nesses casos, é inapropriado assumir volatilidade constante. Um modelo do tipo GARCH consegue captar a trajetória da volatilidade no tempo, por meio da regressão dos resíduos quadráticos de um modelo autorregressivo ARMA. Esses dois modelos, ARMA / GARCH, apesar de serem diferentes, têm seus parâmetros estimados simultaneamente, pela minimização da soma dos resíduos quadráticos de uma regressão MQO. A equação seguinte é de modelo ARMA de ordem um, AR(1):

⁶ A escolha do Índice Dow Jones é discutida posteriormente, no item 4.3 neste capítulo.

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + v_t \quad (4.11)$$

sendo:

y_t : retorno da carteira de mercado no tempo t ;

v_t : resíduo da regressão no tempo t , com variância igual a σ_v^2 ;

φ_0, φ_1 : coeficientes lineares do processo AR(1).

- **Teste *Q* Ljung Box de autocorrelação dos resíduos quadráticos**

Os resíduos resultantes da regressão ARMA são testados para identificar a presença de um processo GARCH na série. Segundo Enders (2004:119), devem-se calcular as autocorrelações dos resíduos quadráticos, para defasagens $i = 1, 2, 3, \dots$, da seguinte forma:

$$\rho_i = \frac{\sum_{t=i+1}^N (v_t^2 - \sigma_v^2)(v_{t-i}^2 - \sigma_v^2)}{\sum_{t=1}^N (v_t^2 - \sigma_v^2)^2} \quad (4.12)$$

sendo:

ρ_i : autocorrelação dos resíduos quadráticos para a defasagem i ;

N : número de resíduos da série;

v_t^2 : resíduo quadrático resultante do processo ARMA (p,q);

σ_v^2 : variância dos resíduos v_t .

Os valores da correlação ρ_i significativamente diferentes de zero indicam a presença de um processo GARCH; rejeitar a hipótese nula de que v_t^2 são serialmente não correlacionados é equivalente a rejeitar a hipótese nula de ausência de processo GARCH, sendo assim, a hipótese nula que se pretende testar é:

H_0 – ausência de correlação serial, ou seja, ausência de processo GARCH nos resíduos quadráticos v_t^2 .

O objetivo é que a hipótese nula seja rejeitada. O teste utilizado para verificar a hipótese acima é a estatística *Q Ljung Box*, que possui uma distribuição χ^2 , com n graus de liberdade, conforme a seguinte equação:

$$Q = N(N + 2) \sum_{i=1}^n \rho_i / (N - i) \quad (4.13)$$

- **Especificação do Processo GARCH**

A variável de interesse para o modelo do preço à vista da *commodity* é a volatilidade condicional do mercado um passo à frente $\hat{\sigma}_t^M$, que pode ser obtida por meio da modelagem dos resíduos v_t resultantes da regressão ARMA dos retornos da carteira de mercado. A volatilidade condicional é obtida por meio dos resíduos quadráticos condicionais, conforme a seguinte equação:

$$\hat{\sigma}_{t|t-1}^M = \sqrt{\hat{v}_{t|t-1}^2} \quad (4.14)$$

sendo:

$\hat{\sigma}_t^M$: volatilidade condicional do mercado em t ;

$\hat{v}_{t|t-1}$: resíduo condicional.

O resíduo condicional esperado $\hat{v}_{t|t-1}$ pode ser calculado pela variância condicional h_t , tal que $\hat{v}_{t|t-1} = \sqrt{h_t}$. Abaixo está a equação que descreve a variância condicional h_t , resultante de um processo GARCH (1,1):

$$h_t = \alpha_0 + \alpha_1 v_{t-1}^2 + \beta_1 h_{t-1} \quad (4.15)$$

sendo:

h_t : variância condicional do mercado no tempo t ;

α_0, α_1 : coeficientes lineares do processo GARCH(1,1).

Após obter a variância condicional, calcula-se a volatilidade condicional $\hat{\sigma}_{t|t-1}^M$, que será adicionada ao modelo de formação de *commodities* na forma de espaço de estado. Os resultados dos testes e os coeficientes estimados no modelo GARCH serão apresentados no próximo capítulo.

4.2.4 Parte 2 – Estimação do Modelo com o Filtro de Kalman

Nesta parte do estudo, o modelo de formação de preços de *commodity* na forma de um sistema de equações de estado teve seus parâmetros estimados com o filtro de Kalman. Além do modelo proposto, mais um modelo disponível na literatura também teve seus parâmetros estimados, com o intuito de verificar, na terceira parte do estudo, a capacidade de previsão dos dois modelos.

Nesta etapa, durante o processo de filtragem, o preço à vista do açúcar é tratado como variável não observável. O objetivo é estimar os parâmetros dos modelos, verificar os sinais e as dimensões dos coeficientes relativos às variáveis analisadas na primeira parte do estudo e gerar variáveis de estado filtradas. Além disso, com os coeficientes resultantes do processo de filtragem, obter séries de previsões n -passos à frente dos preços do açúcar, tanto para o modelo proposto, quanto para o modelo de comparação.

O modelo selecionado para comparação da capacidade de previsão é o de **dois fatores** sugerido por Gibson e Schwartz (1990:960), as razões que motivaram a escolha desse modelo são discutidas no item 4.2.6. O processo de filtragem e de previsão dos preços do açúcar por meio do filtro de Kalman será explicitado na próxima seção.

4.2.5 Filtro de Kalman

O filtro de Kalman é definido, segundo Grewal e Andrews (2001:22), como um processo de estimação de **variáveis de estado** em um **sistema dinâmico** linear perturbado por um ruído branco com distribuição normal. É um algoritmo recursivo de atualização sequencial **um passo à frente**. O estimador resultante do processo é ótimo, em termos de erro quadrático de estimação.

Mayback (1979:4), em seu trabalho aplicado em engenharia elétrica, definiu o filtro de Kalman como “um algoritmo ótimo, recursivo de processamento de dados”. O autor considera o filtro como uma ferramenta utilizada para obter a estimação de parâmetros perturbados por ruídos.

O filtro de Kalman é um dos inúmeros recursos que podem ser utilizados na estimação dos parâmetros do modelo de formação de preços de *commodities*. A justificativa por essa opção é a possibilidade de inferência de variáveis não observáveis. Dada uma distribuição de probabilidade da variável de estado e um modelo descrevendo o processo estocástico das variáveis observadas, o filtro de Kalman gera distribuições atualizadas, um passo à frente, para as variáveis de estado.

Aiube, Baidya e Tito (2006:219) sugeriram o filtro de partículas como alternativa ao filtro de Kalman para estimação de parâmetros de modelos de formação de preços de *commodities*. Porém, em análise empírica, verificaram que o filtro de Kalman produz variáveis filtradas com menos erros de previsão do que o filtro de partículas, quando comparadas com dados reais.

O filtro de Kalman torna-se mais relevante quando a variável que se quer modelar, ou não é diretamente observável, ou sua observação direta é distorcida por variáveis exógenas não

controláveis. Este é o caso do preço à vista de *commodities*. Os preços negociados no mercado à vista de uma *commodity* são variáveis que, apesar de serem facilmente observadas nos negócios realizados entre os produtores e a indústria compradora, podem ter problemas de observação ou distorções.

Por essa razão, preferiu-se o filtro de Kalman a outros processos de estimação usualmente utilizados em modelagens similares, como os modelos autorregressivos integrados de média móvel, o ARIMA, *Autoregressive Integrated Moving Average*, ou na versão vetorial, o VAR, *Vector Autoregressive Model*. Apesar de os modelos autorregressivos serem eficientes, pois produzem estimadores ótimos em termos de erro quadrático, utilizam a variável dependente como entrada no processo de estimação. O filtro de Kalman possui a vantagem de não precisar das observações da variável dependente. Além disso, grande quantidade de modelos de séries temporais, incluindo o ARIMA, podem ser escritos como **casos particulares** do filtro de Kalman.

Para melhor entender o filtro de Kalman, deve-se primeiro analisar os **sistemas dinâmicos** nos quais ele pode ser aplicado.

- **Sistema dinâmico:**

É difícil encontrar na natureza algum sistema que seja realmente constante. Praticamente todos os sistemas físicos são dinâmicos de alguma forma. Se alguma estimação desses sistemas tem o intuito de buscar algum nível de precisão, deve levar em consideração suas características dinâmicas. Morettin e Toloí (2006:333) definem os modelos que descrevem esses sistemas como MLD, modelos lineares dinâmicos.

Um sistema dinâmico é um sistema que se modifica com o passar do tempo e pode ser matematicamente representado por meio de um sistema de equações diferenciais lineares que representam a evolução no tempo das **variáveis de estado**.

- **Especificação do modelo de espaço de estado:**

A representação linear do modelo de espaço de estado de um sistema dinâmico pode ser feita por meio de um sistema de equações, como foi feito com o modelo de formação de preços de

commodities no item 3.5.7 do capítulo anterior. Em um sistema de equações de estado, a primeira equação (\mathbf{y}_t), ou conjunto de equações, é definida como equação de **sinal**, ou equação observada, enquanto a segunda (\mathbf{x}_t) é definida como equação de **estado**, conforme segue:

$$\begin{cases} \mathbf{y}_t = \mathbf{A} + \mathbf{B}\mathbf{x}_t + \boldsymbol{\varepsilon}_t \\ \mathbf{x}_t = \mathbf{C} + \mathbf{D}\mathbf{x}_{t-1} + \mathbf{v}_t \end{cases} \quad (4.16)$$

Os ruídos brancos $\boldsymbol{\varepsilon}_t$ e \mathbf{v}_t são dois vetores, cujas variâncias são parâmetros estimados no processo, representadas por $\boldsymbol{\Omega}_t$:

$$\boldsymbol{\Omega}_t = \text{var} \begin{bmatrix} \boldsymbol{\varepsilon}_t \\ \mathbf{v}_t \end{bmatrix} = \begin{bmatrix} \mathbf{H}_t & \mathbf{G}_t \\ \mathbf{G}_t' & \mathbf{Q}_t \end{bmatrix} \quad (4.17)$$

sendo:

\mathbf{H}_t : uma matriz de variância simétrica $n \times n$;

\mathbf{Q}_t : uma matriz de variância simétrica $m \times m$;

\mathbf{G}_t : uma matriz de covariância $n \times m$.

No item seguinte, será abordado o processo de atualização da média e da variância das equações de estado e de sinal.

- **Processo de Filtragem:**

Conforme Grewal e Andrews (2001:29), o sistema dinâmico em análise é **processo contínuo em tempo discreto**, isso significa que a variável de estado varia continuamente em um intervalo de tempo discreto Δt .

No processo de filtragem, parte-se da premissa de que é possível estimar a distribuição condicional do vetor de estado \mathbf{x}_t , dada a informação disponível até o tempo q , sendo $q < t$.

A média e variância condicionais do vetor \mathbf{x}_t podem ser dadas por:

$$\begin{aligned}\hat{\mathbf{x}}_{t|q} &= \mathbf{E}_q(\mathbf{x}_t) \\ \hat{\boldsymbol{\sigma}}_{\mathbf{x}_{t|q}}^2 &= \mathbf{E}_q\left[(\mathbf{x}_t - \hat{\mathbf{x}}_{t|q})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|q})'\right]\end{aligned}\quad (4.18)$$

sendo:

$\hat{\mathbf{x}}_{t|q}$: vetor de esperança condicional de \mathbf{x}_t em t , dada a informação até o tempo q ;

$\hat{\boldsymbol{\sigma}}_{\mathbf{x}_{t|q}}^2$: matriz de variância condicional de \mathbf{x}_t em t , dada a informação até o tempo q .

Se o processo de estimação um passo à frente for adotado, significa assumir que $q = t - 1$, ou seja, t está apenas uma **unidade de tempo** à frente. Nesse caso, é possível estimar um passo à frente a esperança condicional $\hat{\mathbf{x}}_{t|t-1}$ e a variância condicional $\hat{\boldsymbol{\sigma}}_{\mathbf{x}_{t|t-1}}^2$ da variável de estado \mathbf{x}_t , assumida a premissa de que a média e a variância são estimadores ótimos em termos de mínimo erro quadrático.

Com os estimadores dos parâmetros da distribuição gaussiana da variável de estado \mathbf{x}_t , é possível estimar um passo à frente o vetor \mathbf{y}_t , da seguinte forma:

$$\hat{\mathbf{y}}_{t|t-1} = \mathbf{E}_{t-1}(\mathbf{y}_t) = \mathbf{E}(\mathbf{y}_t | \mathbf{x}_{t|t-1}) = \mathbf{A} + \mathbf{B}\hat{\mathbf{x}}_{t|t-1}\quad (4.19)$$

O vetor de erro um passo à frente será dado por:

$$\hat{\boldsymbol{\varepsilon}}_{t|t-1} = \mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}\quad (4.20)$$

E a variância do erro um passo à frente dada por:

$$\hat{\sigma}_{\varepsilon_{t|t-1}}^2 = \text{var}(\hat{\varepsilon}_{t|t-1}) = \mathbf{B}\hat{\sigma}_{x_{t|t-1}}^2\mathbf{B}' + \mathbf{H}_t \quad (4.21)$$

O filtro de Kalman permite realizar a previsão para cada inovação disponível no tempo. Dados os valores iniciais de média e variância para a variável de estado dos valores das matrizes e vetores $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{G}, \mathbf{H}, \mathbf{Q}$ e as observações de y_t , o filtro de Kalman pode ser usado para computar um passo à frente:

- a) a estimativa da variável de estado x_t , cuja esperança e variâncias condicionais são

$$\hat{x}_{t|t-1} \text{ e } \hat{\sigma}_{x_{t|t-1}}^2 ;$$

- b) os valores **filtrados** contemporâneos da variável de estado x_t , cujas médias e variâncias são μ_{x_t} e $\sigma_{x_t}^2$;

- c) a previsão da variável de sinal $\hat{y}_{t|t-1}$, do erro de previsão $\hat{\varepsilon}_{t|t-1}$ e da variância do erro de previsão da variável de sinal $\hat{\sigma}_{\varepsilon_{t|t-1}}^2$.

- **Condições Iniciais do Modelo**

Para o processo de estimação dos parâmetros do modelo utilizando o filtro de Kalman, é necessário dar o primeiro passo utilizando uma **semente** – uma informação utilizada para começar a rotina computacional do algoritmo recursivo do processo de estimação dos parâmetros do modelo.

Para isso é necessário estimar valores iniciais para o vetor de média das variáveis de estado $\hat{x}_{1|0}$ e para a matriz de variância $\hat{\sigma}_{x_{1|0}}^2$. As estimativas iniciais são relevantes, pois permitem uma melhor e mais rápida **convergência** da rotina computacional de estimação.

Para calcular as estimativas de $\hat{x}_{1|0}$ e de $\hat{\sigma}_{x_{1|0}}^2$ para o primeiro passo da rotina no processo de estimação dos parâmetros do modelo, podem-se utilizar dados passados, anteriores ao período da amostra, ou dados da amostra observada.

- **Previsão**

A previsão, utilizando os parâmetros estimados através do processo de filtragem, permite evidenciar a aplicação do modelo de uma forma prática, ou seja, possibilitando a **previsão**, dentro de um intervalo de confiança, **dos preços da commodity** para um determinado período à frente.

O processo de previsão pode ser feito utilizando-se ou não as inovações das variáveis observadas. Seguem-se os principais tipos de previsão, por meio dos parâmetros estimados pelo filtro de Kalman.

- a) **Previsão n passos à frente**: nesse tipo de previsão, tem-se apenas um dado número de observações, para um dado período de tempo, que são utilizadas para realizar a previsão para múltiplos passos à frente, sem utilizar novas informações. Para isso, são consideradas a média e variância condicionais n passos à frente, conforme a equação:

$$\begin{aligned}\hat{\mathbf{x}}_{t+n|t} &= \mathbf{E}_t(\mathbf{x}_{t+n}) \\ \hat{\boldsymbol{\sigma}}_{\mathbf{x}_{t+n|t}}^2 &= \mathbf{E}_t\left[(\mathbf{x}_{t+n} - \hat{\mathbf{x}}_{t+n|t})(\mathbf{x}_{t+n} - \hat{\mathbf{x}}_{t+n|t})'\right]\end{aligned}\quad (4.22)$$

A previsão n passos à frente da variável de sinal é dada por:

$$\hat{\mathbf{y}}_{t+n|t} = \mathbf{E}_t(\mathbf{y}_{t+n}) = \mathbf{E}(\mathbf{y}_{t+n} | \mathbf{x}_{t+n|t}) = \mathbf{A} + \mathbf{B}\hat{\mathbf{x}}_{t+n|t} \quad (4.23)$$

e a correspondente matriz de mínimos erros quadrados dada por:

$$\hat{\boldsymbol{\sigma}}_{\boldsymbol{\varepsilon}_{t+n|t}}^2 = \mathbf{var}(\hat{\boldsymbol{\varepsilon}}_{t+n|t}) = \mathbf{B}\hat{\boldsymbol{\sigma}}_{\mathbf{x}_{t+n|t}}^2\mathbf{B}' + \mathbf{H}_t \quad (4.24)$$

O processo de previsão em $t+n$ consiste em realizar a previsão utilizando a informação disponível apenas até t . O procedimento de previsão consiste em iniciar o filtro em $t+1$, com os valores de média e variância estimados das variáveis de estado até o tempo t e executar o modelo n passos à frente, sem utilizar **nenhuma nova informação** das variáveis de sinal.

- b) **Previsão suavizada**: nesse processo de previsão, toda a amostra de variáveis de sinal é utilizada para estimação dos valores de $\hat{x}_{t+n|t+n}$, até a data $t+n$. Na prática, essa modalidade de previsão é útil apenas quando se deseja interpolar dados entre o intervalo de tempo da amostra, ou testar estatisticamente a qualidade de predição entre diferentes modelos. Para estimar o preço da *commodity* para uma determinada data $t+n$ no futuro, esse procedimento não será útil, se os dados das variáveis de sinal não estiverem disponíveis para a janela de previsão.

O processo de previsão utilizado para a comparação dos modelos na terceira parte dos testes estatísticos será o de ***n* passos à frente**, ou seja, sem utilizar as inovações das variáveis observadas.

4.2.6 Parte 3 – Testes de Capacidade de Previsão dos Modelos

O modelo proposto, na forma de espaço de estado, foi aplicado ao mercado de açúcar no Brasil, cujos parâmetros foram estimados através do filtro de Kalman. Para isso foram utilizados dados dos mercados futuros de açúcar e omitidos os preços à vista divulgados pelo CEPEA. O objetivo da análise empírica nesta etapa é comparar os resultados obtidos no modelo proposto com o modelo de dois fatores de Gibson e Schwartz (1990:960) e determinar qual dos dois modelos é superior em termos de previsão *n* passos à frente.

Schwartz (1997:957), em seu estudo empírico, comparou os modelos de um, dois e três fatores com dados de três *commodities* diferentes. Verificou, em seus testes, que os resultados dos modelos de dois e três fatores são superiores aos de um fator e concluiu que o modelo de formação de preços de *commodities* não pode ignorar o processo de reversão à média no retorno de conveniência. O autor também observou que o modelo de três fatores não produziu ganho de informação comparado com o de dois fatores, sendo este último mais parcimonioso. Segundo Schwartz (1997:925), os modelos de dois e três fatores “(...) muitas vezes são indistinguíveis e sempre estão muito próximos um do outro”, ou seja, a adição do terceiro fator, de reversão à média das taxas de juros, não melhorou a capacidade de previsão do modelo.

Com base nos resultados obtidos por Schwartz (1997), optou-se pelo modelo de dois fatores para comparação com o novo modelo proposto, pois o modelo de dois fatores é mais eficiente, em termos de previsão, do que os outros disponíveis na literatura. O modelo de curto e longo prazo poderia ser uma alternativa para comparação, porém tem uma limitação: depende de um mercado futuro que tenha liquidez em contratos de curta e de longa maturidade, ou seja, que se tenha uma **curva** amplamente negociada em vários vencimentos de contratos futuros da *commodity*. Essa não é uma realidade para o mercado futuro de açúcar negociado na BM&FBOVESPA, em São Paulo, por essa razão, o modelo de curto e longo prazo não poderia ser utilizado para a comparação neste estudo empírico.

O **modelo de dois fatores**, utilizado para comparação com o modelo proposto, foi especificado na forma de espaço de estado, cujos parâmetros também foram estimados pelo filtro de Kalman, conforme o seguinte sistema de equações:

$$\begin{cases} \ln(F_t) = r(T-t) - (T-t)\delta_t + \ln(S_t) \\ \delta_t = \phi\mu_\delta + (1-\phi)\delta_{t-1} + \sigma_\delta z \\ \ln(S_t) = \beta_0 + \sigma_s z \end{cases} \quad (4.25)$$

sendo:

S_t : preço à vista do açúcar na data t , com variância σ_s^2 ;

F_t : preço futuro do açúcar na data t com vencimento em T ;

δ_t : retorno de conveniência, com média μ_δ e variância σ_δ^2 ;

ϕ : fator de reversão à média do retorno de conveniência.

Para comparar os resultados obtidos nos dois modelos, foram feitas previsões n passos à frente, para todas as N observações da amostra, e os resultados obtidos foram comparados, usando-se as medidas de erro de previsão entre as variáveis observada S_o e prevista S_f . O processo de comparação dos dois modelos foi feito seguindo os seguintes passos:

- **primeiro passo:** os parâmetros dos dois modelos foram estimados por meio do filtro de Kalman, utilizando dados do mercado futuro de açúcar na BM&FBOVESPA;
- **segundo passo:** com os parâmetros dos modelos, foram geradas séries de previsão n passos à frente para os preços do açúcar à vista;
- **terceiro passo:** os valores previstos de açúcar à vista S_f dos dois modelos foram comparados com o preço real do açúcar à vista S_o divulgado pelo CEPEA e foram calculados os erros de previsão $S_o - S_f$. Utilizaram-se medidas de erro de previsão para a comparação dos dois modelos e o teste Granger-Newbold (1976) para verificar, estatisticamente, qual dos dois modelos é melhor em termos de previsão.

As medidas de erro de previsão adotadas foram também utilizadas no trabalho de Aiube, Baidya e Tito (2006:222) para testes de previsão em modelos de formação de preços de *commodities*. O intuito dessas medidas é quantificar o erro de previsão dos dois modelos, com base na diferença entre o preço observado do açúcar S_o e o preço previsto S_f . As medidas adotadas são as seguintes:

$$\begin{aligned}
 RMSE &= \sqrt{\frac{\sum_{j=1}^N (S_{ot} - S_{ft})^2}{N}} \\
 MAE &= \sum_{j=1}^N \frac{|S_{ot} - S_{ft}|}{N} \\
 MAPE &= \frac{1}{N} \sum_{j=1}^N \left| \frac{S_{ot} - S_{ft}}{S_{ot}} \right|
 \end{aligned} \tag{4.26}$$

sendo:

N : número de previsões da série.

A primeira medida – $RMSE$ – é a raiz do erro quadrático médio, a segunda – MAE – é o erro absoluto médio e a terceira – $MAPE$ – é o erro absoluto percentual médio.

Além da análise das medidas de erros, verificou-se estatisticamente qual dos dois modelos é o melhor em termos de previsão. Para isso, utilizou-se o método Granger-Newbold, que permite realizar testes estatísticos de hipóteses de comparação de previsão entre dois modelos. Granger e Newbold (1976:195) definem duas variáveis para os testes de previsão:

$$w_t = e_{1t} + e_{2t} ; \quad g_t = e_{1t} - e_{2t} \quad (4.27)$$

sendo:

e_{1t} : erro de previsão do modelo 1 no tempo t ;

e_{2t} : erro de previsão do modelo 2 no tempo t ;

Os erros de previsão e_{1t} e e_{2t} são calculados da seguinte forma:

$$e_{1t} = S_{1ft} - S_{ot} \quad \text{e} \quad e_{2t} = S_{2ft} - S_{ot} \quad (4.28)$$

sendo:

S_{1ft} : valor previsto do preço do açúcar no tempo t pelo modelo 1;

S_{2ft} : valor previsto do preço do açúcar no tempo t pelo modelo 2;

S_{ot} : valor observado do açúcar à vista no tempo t .

A hipótese nula do teste de previsão é que os modelos são iguais, em termos de erros de previsão, para isso, as variáveis w_t e g_t devem ser não correlacionadas, sendo a correlação das variáveis dada por:

$$\text{corr}(w, g) = E[e_{1t}^2 - e_{2t}^2] \quad (4.29)$$

Se o modelo 1 tiver maior erro de previsão, $\text{corr}(w, g)$ será positiva; se o modelo 2 errar mais, a correlação será negativa. De acordo com Granger e Newbold (1976), a seguinte equação, com o coeficiente de correlação amostral $\rho_{w,g}$, tem uma distribuição t com $N-1$ graus de liberdade:

$$\rho_{w,g} / \sqrt{(1 - \rho_{w,g}^2) / (N - 1)} \quad (4.30)$$

Foram realizados testes t bicaudal para os dois modelos, com as séries de previsões n passos à frente para os modelos de dois fatores e para o modelo proposto. Os resultados são apresentados no capítulo seguinte.

4.3 Definição Teórica e Operacional das Variáveis

Nesta seção são apresentadas as variáveis utilizadas no processo investigatório das hipóteses de pesquisa discutidas anteriormente. O intuito nesta seção é consolidar as características de todas as variáveis utilizadas nos testes estatísticos.

- **Preço à Vista do Açúcar**

O preço a vista do açúcar no Brasil é a variável de interesse que se deseja modelar. Variável de estado no modelo de formação de preços. Para coleta dos dados da amostra, considerou-se o padrão do CEPEA (2008), pois é amplamente utilizado em estudos e pesquisas equivalentes e é ativo-objeto dos contratos futuros na BM&FBOVESPA. A unidade de negociação do açúcar à vista é a saca de 50 kg de açúcar cristal, conforme especificação do CEPEA (2008):

Açúcar Cristal – com mínimo de polarização de 99,7 graus, máximo de 0,10% de umidade, cor ICUMSA mais frequente 130-180, máximo de 0,07% de cinzas, ensacado em sacas novas de polipropileno, destinado ao mercado interno.

Conforme a metodologia do CEPEA, os preços são coletados diariamente das negociações sem frete entre unidades produtoras – usinas, compradores e grandes intermediários de vendas

no mercado de açúcar, referentes a negócios efetuados no Estado de São Paulo, nas regiões de Piracicaba, Ribeirão Preto, Jaú e Assis.

O CEPEA calcula uma média ponderada dos preços coletados por região, por volume médio regional do produto, direcionado ao mercado doméstico, considerando as três últimas safras. Apesar de ser um excelente indicador dos preços do açúcar à vista, existem limitações, pois pode haver distorções causadas pela média de preços negociados em diferentes localidades. Conforme discutido anteriormente, a escassez de açúcar em uma ou mais localidades pode causar distorção no preço médio.

O preço à vista do açúcar, para os cálculos, foi convertido em dólares por saca de 50 kg, utilizando a cotação média diária do dólar comercial divulgada pelo Banco Central. O intuito dessa conversão foi padronizar todos os ativos do modelo em uma mesma moeda.

A série de preços à vista do açúcar foi utilizada em apenas duas das três etapas dos estudos estatísticos:

Na primeira parte dos testes estatísticos, a série de preço à vista do açúcar foi utilizada como variável dependente de regressão linear múltipla, para os testes de hipótese das premissas do modelo.

Na segunda parte, o preço à vista foi omitido, pois o intuito, nessa etapa, foi modelá-lo como variável não observável. O preço à vista do açúcar foi estimado no processo de filtragem do modelo.

Na terceira e última parte, a série de preços à vista do açúcar foi novamente utilizada para realizar os testes estatísticos de comparação da capacidade de previsão n passos à frente dos modelos. Nessa parte, o preço à vista foi considerado variável observada.

- **Preço à Vista do Etanol**

A série histórica de preços à vista do álcool anidro carburante – etanol – foi utilizada na primeira parte dos testes estatísticos, para verificar a hipótese de pesquisa sobre a relação do etanol com os preços do açúcar à vista. Considerou-se também o padrão utilizado pelo

CEPEA (2008), que está de acordo com as normas da ANP, Agência Nacional de Petróleo, Gás Natural e Biocombustíveis. A unidade de negociação da série de preços do álcool divulgado pelo CEPEA é de Reais por litro, divulgada em periodicidade semanal.

As especificações do etanol são apresentadas no Quadro 3:⁷

Quadro 3 - Especificações do álcool anidro

Característica	Especificação
Aspecto	Aspecto límpido e isento de impurezas.
Cor	Incolor, ou com cor levemente amarelada.
Acidez	Acidez total de 30mg/l.
Massa Específica	Máximo de 791,5 quilogramas por metro cúbico a 20°C.
Teor de etanol	Mínimo de 99,6%.

FONTE: BM&FBOVESPA (2008).

Conforme a metodologia do CEPEA, o etanol é calculado sem frete, retirado na unidade de produção. A região de referência para o cálculo de preços é o Estado de São Paulo, como preços ponderados pelo volume comercializado nos negócios realizados entre usinas e distribuidores de combustíveis. Para os testes estatísticos, os preços semanais do etanol à vista foram convertidos para dólares por metro cúbico.

- **Preço do Contrato Futuro de Açúcar**

O preço do contrato futuro de açúcar é variável de sinal do modelo de formação de preços, ou seja, variável observável no modelo. As cotações do contrato futuro são diariamente divulgadas pela BM&FBOVESPA (2008). O ativo-objeto do contrato futuro de açúcar negociado na BM&FBOVESPA possui as mesmas especificações do açúcar à vista divulgado

⁷ Especificações conforme contrato futuro de etanol na BM&FBOVESPA (2008).

pelo CEPEA. O tamanho do contrato é de 270 sacas de 50 kg, com cotação em centavos de dólar por saca. Os vencimentos ocorrem nos meses de fevereiro, abril, julho, setembro e novembro.

Foram observados, na amostra, os preços de ajuste diário do primeiro vencimento do contrato futuro de açúcar. À medida que o contrato se aproximava do vencimento, foi feita a rolagem para o vencimento seguinte, em um prazo mínimo de dez dias antes do vencimento. Esse procedimento foi adotado para evitar distorções de curto prazo provocadas pelos últimos dias de ajuste antes do vencimento do contrato.

Com a divulgação diária dos preços dos contratos futuros, também são divulgadas as datas de vencimento dos contratos. Com essas datas, foi possível calcular os prazos $T-t$ em dias corridos até o vencimento do contrato futuro.

- **Preço do Contrato Futuro de Petróleo**

O preço do contrato futuro de petróleo WTI negociado na NYMEX foi considerado como variável exógena observável no modelo e componente da média no processo de formação do preço à vista do açúcar. O preço do petróleo foi utilizado, nas duas primeiras partes dos estudos estatísticos, como variável explicativa nas regressões múltiplas e como variável exógena no processo de filtragem.

O contrato futuro de petróleo utilizado como referência neste estudo foi o WTI por ser o contrato futuro de maior liquidez em termos de volume de negociação, comparado a outros contratos como o *Brent* da IPE, em Londres. O tamanho do contrato WTI é de mil barris de petróleo, com cotação em dólares por barril.

A série histórica de cotações do petróleo foi coletada a partir dos preços de ajuste do primeiro vencimento do contrato futuro. O contrato de primeiro vencimento, sempre o mais negociado na NYMEX, é referência mundial de cotação dos preços do barril de petróleo.

- **Retorno de Conveniência**

Como definido nos capítulos anteriores, o retorno de conveniência é um conceito econômico não observável, um valor medido em termos de taxa de juros aplicada sobre o preço do ativo à vista. Pode ser entendido como uma vantagem econômica que o detentor da *commodity* física possui, por ter o benefício de poder esperar o melhor momento para vender sua mercadoria a um preço que maximize sua receita. Em momentos nos quais os estoques globais da *commodity* estão baixos, o retorno de conveniência é máximo.

A série de taxas diárias do retorno de conveniência não pode ser diretamente observada, portanto será estimada durante o processo de filtragem do modelo, inclusive seu valor médio no período da amostra. A unidade de medida do retorno de conveniência no modelo proposto é taxa contínua, com prazo de referência anual, na base de 360 dias corridos.

- **Volatilidade do Mercado**

A volatilidade do mercado, medida em termos de desvio-padrão dos retornos contínuos diários do Índice Dow Jones, é considerada variável exógena de entrada no modelo proposto e será componente da média no processo de formação do preço à vista do açúcar.

O índice selecionado para os testes foi o Dow Jones, pois é calculado com uma carteira representativa no mercado de ações negociadas na NYSE, *New York Stock Exchange*. A abrangência desse índice no mercado de ações nos Estados Unidos é relevante para captar a volatilidade. O *Dow Jones Industrial Average* é baseado em uma carteira teórica das 30 ações mais negociadas nos Estados Unidos e é um importante indicador de comportamento do mercado naquele país.

A volatilidade foi estimada utilizando-se os retornos do Índice Dow Jones, conforme explicado no item 4.2.3, por meio da variância condicional de um processo ARMA / GARCH. A volatilidade do mercado foi utilizada nas duas primeiras etapas do processo de investigação.

- **Componente Sazonal**

Conforme apresentado no item 3.5.6 do capítulo anterior, o componente sazonal é variável exógena de entrada no modelo, a qual tenta captar os padrões cíclicos nos preços do açúcar provocados pela safra da cana-de-açúcar. Utilizou-se apenas um ciclo de sazonalidade anual, que possui dois coeficientes estimados durante o processo de filtragem: um coeficiente relativo a **seno** e outro relativo a **coseno**.

O componente sazonal está presente na primeira e segunda etapas dos testes estatísticos, nos testes das premissas e durante o processo de filtragem do preço do açúcar à vista.

- **Taxa de Juros Livre de Risco**

A taxa de juros do ativo livre de risco, juntamente com o retorno de conveniência, é utilizada na equação de arbitragem de preços entre mercado futuro e mercado à vista do preço da *commodity*. A taxa de juros livre de risco, no conceito teórico, é uma variável que possui correlação nula com a carteira teórica de mercado, deve ter variância igual a zero e não pode haver restrições, em termos de volume, para captação ou aplicação de recursos pela taxa livre de risco. O risco de um ativo pode ser medido em termos de volatilidade dos retornos, logo, a taxa de retorno do ativo livre de risco deve possuir volatilidade nula.

Um ativo livre de risco é aquele em que o investidor sabe exatamente o valor que receberá ao final do prazo de investimento, é a taxa básica de uma economia, ou taxa de retorno obtida pela simples restrição ao consumo. Para não haver incerteza quanto ao valor a ser recebido, deve-se pressupor, portanto, um desvio-padrão dos retornos do ativo igual a zero e, conseqüentemente, um beta igual a zero, conforme definição teórica sugerida por Copeland, Koller e Murrin (2002:220).

É difícil obter um ativo, em qualquer economia, cuja taxa de retorno possua todas as características acima mencionadas. Uma *proxy* para a taxa de juros do ativo livre de risco pode ser a taxa de juros de curto prazo das operações de transferência de reservas no mercado interbancário. Neste trabalho foram utilizadas as taxas médias das operações dos *Fed Funds*. Essa escolha é coerente com as outras variáveis do modelo, pois todas estão referenciadas em dólar. A taxa dos *Fed Funds* é divulgada diariamente pelo *Federal Reserve* (2008), autoridade

monetária nos Estados Unidos responsável pelas transações que envolvem as reservas bancárias naquele país. A taxa livre de risco é variável exógena no modelo proposto, utilizada na segunda parte dos estudos estatísticos, durante o processo de filtragem.

- **Cotação do Dólar**

A cotação do dólar foi utilizada apenas para converter os preços do açúcar e do etanol, divulgados pelo CEPEA em reais, para preços em dólares, portanto não é variável de entrada do modelo. A série histórica de taxa de câmbio foi obtida pela cotação média do dólar, divulgada diariamente pelo BACEN –, Banco Central do Brasil. A cotação é divulgada em reais por dólar dos Estados Unidos da América, das operações negociadas no mercado de câmbio comercial, para entrega pronta, cotação de venda, regulamentada pelo CMN, Conselho Monetário Nacional e divulgada por intermédio do SISBACEN, Sistema do Banco Central, na transação PTAX800, opção “5”, cotação de fechamento, para liquidação em dois dias.

O Quadro 4, a seguir, apresenta um resumo das variáveis utilizadas durante o estudo empírico e uma breve descrição de cada uma delas:

Quadro 4 - Variáveis analisadas no estudo estatístico

Sigla	Variável	Finalidade	Fonte e Descrição
SV1	Retorno de conveniência	Variável de Estado	Filtrada no modelo na forma de taxa de juros contínua anual, base 360 dias corridos.
SV2 (acucarbr)	Preço à vista do açúcar cristal no Brasil.	Variável de Estado	CEPEA – preço médio das negociações de sacas de 50 kg de açúcar convertidos em dólar.
Fut	Preço do contrato futuro de açúcar cristal.	Variável de Sinal	BM&FBOVESPA – preços de ajuste diário dos contratos futuros de açúcar cristal, cotados em dólar por saca de 50 kg.
Etanol	Preço à vista do álcool anidro – etanol no Brasil.	Variável utilizada para testes de hipótese	CEPEA – preço médio das negociações de álcool anidro convertido em dólares por metro cúbico.
Petr	Preço futuro de petróleo WTI de primeiro vencimento.	Variável Exógena	NYMEX – contrato futuro de petróleo de primeiro vencimento do tipo WTI, cotado em dólares por barril.
R	Taxa de juros livre de risco na forma contínua.	Variável Exógena	<i>Federal Reserve</i> – taxa dos <i>Fed Funds</i> , divulgadas diariamente em base anual.
Vol (retdjones)	Volatilidade do Índice Dow Jones	Variável Exógena	Volatilidade condicional do retorno diário do índice Dow Jones, calculada por meio de um processo ARMA / GARCH.
seno e cosseno	Componente Sazonal	Variável Exógena	Componentes trigonométricos seno e cosseno que identificam ciclo anual de sazonalidade do preço do açúcar.
Dol	Cotação diária do dólar comercial	Utilizado para converter o preço do açúcar e do álcool do CEPEA para dólar.	Banco Central do Brasil – dólar médio PTAX800 de venda, divulgado diariamente.
Dc	Prazo em dias corridos até o vencimento do contrato futuro de açúcar.	Utilizado para o ajuste temporal da taxa livre de risco e do retorno de conveniência.	BM&FBOVESPA – informação disponibilizada com os ajustes dos contratos futuros de açúcar.

4.4 População, Amostragem e Coleta de Dados

Para caracterização do universo populacional da pesquisa, Martins (1994:34) sugere que a região física ou geográfica seja definida simultaneamente com a unidade que representa a população. O universo populacional que se deseja estudar é o dos preços do mercado de açúcar. O preço de referência do açúcar utilizado no estudo é divulgado pelo CEPEA e pela BM&FOVESPA, que utilizam dados coletados, para composição do preço médio, de negociações ocorridas em regiões dentro do Estado de São Paulo. Logo, não é equivocado afirmar que o universo populacional está localizado geograficamente no Estado de São Paulo.

Contudo também é conveniente levar em consideração que o açúcar é uma *commodity* negociada internacionalmente, logo a formação de preços, apesar de regionalmente localizada, é influenciada por forças de oferta e demanda no mercado internacional. Os preços do açúcar são amplamente negociados em bolsas de mercados futuros em Nova York e Londres, permitindo arbitragens de preços nos diferentes mercados. Portanto, devido a essa característica, os resultados obtidos na pesquisa podem ser generalizados para um contexto geograficamente um pouco mais amplo do que apenas no âmbito regional ou nacional.

Segundo Cooper e Schindler (2003:151), a qualidade de uma amostra é determinada por quanto essa amostra retrata bem as características da população que representa. A validade de uma amostra depende de duas características: acuidade e precisão.

Acuidade é o grau em que os vieses ficam fora da amostra. Em uma amostra com alto grau de acuidade, os resultados obtidos na amostra são geralmente muito próximos aos resultados apresentados na população. A precisão denota os erros no processo de observação dos dados. Uma amostra pode possuir erros decorrentes do processo de coleta de dados, inerentes a flutuações aleatórias, erros de observação, erros de medida ou elementos omitidos na observação⁸, ou seja, pode ter baixa precisão na coleta.

Para aumentar a acuidade da amostra observada, procurou-se obter uma quantidade ampla de dados que fosse representativa no universo populacional. Foram coletadas séries históricas

⁸ As limitações observacionais são discutidas posteriormente neste capítulo.

para uma janela temporal de 6,5 anos de dados, o que permitiu que o estudo incorporasse, na formação de preços do açúcar, diferentes contextos político-econômicos.

A amostra coletada compreende séries históricas de dados observados no período de 2 de janeiro de 2002 até 30 de junho de 2008, totalizando 1631 observações diárias, 343 semanais e 78 mensais. A série semanal foi coletada apenas para a amostra de etanol, porque o CEPEA divulga os preços de etanol apenas com essa periodicidade. Os dados foram coletados somente a partir de 2 de janeiro de 2002, porque havia, nas amostras mais antigas, muitos dias sem negociação no mercado futuro de açúcar na BM&FBOVESPA. Considerou-se a possibilidade de que falta de liquidez pudesse comprometer a formação dos preços no mercado futuro. Por isso, a base de dados de séries anteriores a essa data foi desconsiderada.

A origem dos dados coletados é predominantemente de fontes secundárias, portanto a precisão e qualidade da amostra dependem dessas origens nas quais os dados foram obtidos. Durante o processo de coleta de dados, procurou-se utilizar fontes fidedignas e amplamente utilizadas em trabalhos semelhantes. No Quadro 5, a seguir, são apresentadas as origens dos dados coletados.

Quadro 5 - Origem e periodicidade das séries de dados coletadas

Sigla	Dados Coletados	Origem	Periodicidade da série
Acucarbr	Preço à vista do açúcar	Site do CEPEA (2008)	Diária
Etanol	Preço à vista do álcool anidro – etanol	Site do CEPEA (2008)	Semanal
Fut	Preço do contrato futuro de açúcar cristal. (cód. ISU)	Sistema de recuperação de informações da BM&FBOVESPA (2008)	Diária
Petr	Preço futuro de petróleo WTI. (cód. <i>CLI Comdty</i>)	Sistema Bloomberg	Diária
R	Taxa dos <i>Fed Funds</i> (<i>Federal Funds Effective Rate</i>)	Site do <i>Federal Reserve</i> (2008)	Diária
Djones	Índice Dow Jones	Sistema Económica	Diária
Dol	Cotação do dólar PTAX800	Sistema Económica	Diária

4.5 Limitações da Pesquisa

O processo de especificação do modelo de formação de preços no mercado de açúcar e dos testes estatísticos pode apresentar problemas econométricos ou de observação na amostra.

Os problemas na especificação do modelo e amostragem podem ser causados por:

- cointegração entre as variáveis do modelo: ocorre quando as variáveis não estacionárias possuem pelo menos uma relação linear que seja estacionária;
- causalidade reversa e endogeneidade: ocorre quando as variáveis são assumidas como exógenas, quando na verdade são endógenas;
- variáveis omitidas e redundantes: quando uma ou mais variáveis relevantes não foram adicionadas ao modelo, ou quando uma variável desnecessária é adicionada ao modelo;
- seletividade da amostra: ocorre quando as amostras não são selecionadas aleatoriamente;
- erro de mensuração das variáveis: ocorre quando as variáveis são mensuradas com muito erro.

A seguir os problemas e limitações da pesquisa serão apresentados.

4.5.1 Não Identificação de Variáveis Cointegradas

Um problema econométrico comum ocorre quando variáveis não estacionárias são incluídas no processo de estimação dos parâmetros do modelo. Para evitar essa situação, as variáveis são testadas para verificar a presença de raiz unitária por meio do Teste ADF, Dickey-Fuller Aumentado. A hipótese nula do teste ADF de raiz unitária é de que as variáveis possuem pelo menos uma raiz unitária. Quando o pesquisador rejeita a hipótese nula, corre o risco de

cometer o erro do tipo I (α), ou seja, de rejeitar uma hipótese que realmente é verdadeira. Para tornar as variáveis estacionárias, Enders (2004:164) sugere a diferenciação para a eliminação de tendência, porém esse procedimento é válido apenas se as variáveis não cointegrarem.

Se as variáveis forem integradas de ordem um, $I(1)$, é necessária a verificação da presença de cointegração, ou seja, apesar de as variáveis não serem estacionárias, a relação linear entre elas pode ser estacionária. O teste de cointegração é feito por meio do estudo dos resíduos resultantes da regressão entre as variáveis; se os resíduos forem estacionários, as variáveis cointegram. O problema enfrentado durante os testes é que a hipótese nula do teste de cointegração é de que as variáveis não cointegram. Nesse caso, se os resultados dos testes indicarem a não rejeição da hipótese nula, o pesquisador corre o risco de cometer o erro do tipo II (β), ou seja, de não rejeitar uma hipótese nula, que na verdade é falsa. Quando os resultados dos testes levam à não rejeição de uma hipótese, os resultados não são totalmente conclusivos. A consequência de uma decisão errada no processo investigatório, nesse caso, é que o pesquisador pode ignorar a presença de uma relação linear estacionária das variáveis não transformadas e especificar o modelo de uma forma equivocada.

4.5.2 Causalidade Reversa e Endogeneidade

No modelo, o preço do açúcar é a variável dependente, enquanto o preço do petróleo, a volatilidade e a sazonalidade são variáveis exógenas. É possível que as variáveis consideradas exógenas, na verdade, sejam endógenas, ou seja, variáveis explicativas podem ser, na verdade, variáveis dependentes.

Nas premissas fundamentais do modelo, considerou-se que a volatilidade do mercado causa um aumento nos preços da *commodity*. Um aumento na volatilidade dos mercados, resultante de uma crise econômica ou surto inflacionário, pode provocar uma alteração na tendência dos preços das *commodities*, devido ao fluxo de recursos destinados a investimentos dessa natureza para proteção do valor do capital. Porém, não se pode ignorar a possibilidade de que um aumento no preço das *commodities*, inclusive do petróleo, possa causar aumento na volatilidade do mercado, e não o contrário, o que geraria uma relação espúria no modelo.

Relações espúrias ocorrem entre variáveis que possuem alta correlação e cujos coeficientes são estatisticamente significantes, porém a relação entre as variáveis é inexistente, ou é irrelevante, ou pode ser causada por uma outra variável exógena omitida no modelo. Para minimizar a possibilidade de especificação de um modelo cujo coeficiente seja enviesado e inconsistente devido a uma relação espúria, foram feitos correlogramas cruzados entre as variáveis, cujos coeficientes de correlação foram estatisticamente testados, e também foram realizados testes com os resíduos resultantes das regressões.

4.5.3 Variáveis Omitidas e Redundantes

Apesar dos testes de significância estatística de igualdade a zero dos estimadores dos coeficientes lineares, existe a possibilidade de que variáveis do modelo sejam redundantes, ou seja, quando duas variáveis do modelo explicam as mesmas variações na variável dependente. Quando esse fenômeno sobrevém, as variáveis do modelo podem apresentar multicolinearidade – elevada correlação entre as variáveis independentes. Ao retirar uma variável redundante do modelo, este fica mais parcimonioso, com pouca perda incremental de informação. Variáveis omitidas ou redundantes fazem com que os estimadores MQO fiquem enviesados e inconsistentes, inviabilizando os testes de estatísticos dos coeficientes.

Podem existir também variáveis não identificadas, que poderiam ter sido incluídas, pois são mais relevantes na formação de preços do que as variáveis consideradas no modelo. Algumas dessas variáveis poderiam ser consideradas omitidas e outras redundantes, como nos seguintes casos:

- o retorno de conveniência, incluído no modelo, identifica o benefício não pecuniário que o detentor da mercadoria física possui na possibilidade de falta de estoque global. No lugar do retorno de conveniência para identificar esse benefício, poderia ser incluído o volume global de estoque físico da mercadoria. Nesse caso, o retorno de conveniência poderia ser redundante e o volume de estoques a variável omitida;

- a volatilidade do mercado interfere na tendência dos preços do açúcar, pois os investidores, em situações de crise, buscam o mercado de *commodities* para manutenção do valor do capital. Esse fluxo de recursos para o mercado de *commodities* pode ser causado pela expectativa de inflação nas moedas, e não pela volatilidade do mercado. Nesse caso, os índices de inflação poderiam ser variáveis omitidas e a volatilidade poderia ser redundante.

Muitas outras variáveis, além das acima citadas, poderiam ser adicionadas, em substituição às variáveis do modelo. Contudo, as variáveis atuais foram selecionadas com base em trabalhos anteriormente publicados na literatura e também na premissa fundamental de que a formação de preços de *commodities* – relacionadas com a produção de bioenergia – possui relação significativa com o preço do petróleo.

4.5.4 Seletividade da Amostra

Durante o processo de coleta dos dados para composição da amostra, procurou-se adotar procedimentos que minimizassem o viés de seleção – que ocorre quando uma amostra não é aleatoriamente coletada dentro do universo populacional. Uma amostra não aleatória, ou não representativa da população, resulta em estimadores enviesados, comprometendo o resultado da pesquisa.

Foram identificadas três lacunas no processo de coleta de dados, que poderiam causar viés de seleção: (i) a representatividade das janelas temporais das séries históricas; (ii) a periodicidade das observações; e (iii) as fontes de dados secundários.

O pesquisador pode cometer erros ao escolher a (i) janela temporal das séries de dados, pode selecionar janelas temporais muito pequenas, que não incorporariam os diversos cenários político-econômicos que interferem na formação de preços da *commodity*; ou janelas convenientes, nas quais o modelo apresentaria maior aderência aos resultados desejados. Para evitar esse tipo de viés de seleção, optou-se por uma janela temporal extensa, de 6,5 anos, com 1631 observações diárias e 78 mensais. Acredita-se que, quanto maior a janela temporal, maior também é a representatividade da amostra em relação à população. Além disso,

segundo Greene (2003:65), algumas premissas de estimação podem ser relaxadas para as regressões com grandes amostras e algumas propriedades assintóticas dos estimadores podem ser consideradas.

Além das janelas temporais, também foi tomada precaução quanto à (ii) periodicidade das observações. A unidade de tempo das observações é relevante quando se trata de modelos estocásticos com variáveis defasadas em tempo discreto. Nesses tipos de modelos, os resultados podem ser diferentes, dependendo da periodicidade dos dados. Portanto, para minimizar esse viés, foram coletados dados em duas periodicidades: diária e mensal; e todos os testes estatísticos foram executados para as duas unidades de tempo.

Outro problema observacional é identificado na escolha das (iii) fontes dos dados secundários. Em amostras de dados secundários, o pesquisador não consegue controlar o processo gerador de dados em sua origem primária, ou seja, não consegue garantir que a amostra foi coletada aleatoriamente. Para diminuir esse risco, o procedimento adotado nesta pesquisa foi a utilização de fontes de dados fidedignas, amplamente utilizadas em estudos acadêmicos. Uma outra limitação da pesquisa, que deve ser levada em consideração no momento da coleta de dados secundários, é a escolha entre as diferentes fontes de dados disponíveis. Algumas variáveis podem ter várias fontes de dados diferentes, como as cotações do contrato futuro de açúcar, que podem ser obtidas na BM&FBOVESPA ou nas bolsas de Nova York – NYMEX – e Londres – LIFFE. Optou-se pela BM&FBOVESPA porque o universo amostral que se deseja estudar é o de formação de preços no mercado brasileiro de açúcar, o que não garante que a formação de preços na bolsa brasileira seja tão eficiente quanto nas outras bolsas. Esse viés pode ser mitigado pelo fato de o açúcar ser uma *commodity* internacionalmente negociada e pela possibilidade de realização de operações de arbitragem entre as bolsas.

4.5.5 Erro de Mensuração das Variáveis

Outro problema identificado durante o processo de coleta de dados são erros de mensuração dos dados, que podem ocorrer quando as variáveis exógenas ou endógenas são mensuradas incorretamente. Em dados de origem secundária, o problema de erro de mensuração dificilmente é identificado, devido à não supervisão direta no processo de coleta primária dos

dados. Um erro de mensuração nas variáveis endógenas não enviesava os coeficientes estimados no modelo, pois o termo de erro captura as variações adicionais. Porém, os resultados estimados apresentarão significância estatística menor. Já erros de mensuração nas variáveis exógenas podem fazer com que os coeficientes estimados sejam viesados e inconsistentes.

Os preços do açúcar à vista, utilizados para os testes estatísticos das premissas do modelo, são divulgados pelo CEPEA em base diária, com uma metodologia de coleta de dados publicamente conhecida. Porém, como discutido no item 4.2.5 deste capítulo, os preços à vista do açúcar são médias ponderadas dos preços de negócios realizados entre produtores e compradores, em localidades distintas e, portanto, podem conter distorções devido à falta de estoque da mercadoria física em uma ou mais localidades.

Essas distorções, ou erros de medida, causadas pela média dos preços, podem ser minimizadas quando o modelo é especificado na forma de espaço de estado e estimado pelo filtro de Kalman. Esse procedimento permite que variáveis com possíveis problemas observacionais possam ser tratadas como variáveis de estado não observadas e outras variáveis, mais transparentemente coletadas, como os preços dos contratos futuros, possam ser utilizadas como variáveis de sinal observadas.

5 RESULTADOS DA PESQUISA

Neste capítulo são apresentados os resultados dos testes estatísticos realizados para a aplicação do modelo proposto ao mercado brasileiro de açúcar. O intuito principal, neste capítulo, é tentar responder às perguntas e testar as hipóteses da pesquisa apresentadas no capítulo anterior. Os resultados são apresentados em ordem sequencial, dividida em três partes interdependentes.

- a) **Parte I – resultados dos testes das premissas:** nesta etapa foram verificadas as premissas apresentadas, fundamentais para a especificação do modelo proposto por meio de testes de hipóteses. Testes estatísticos de coeficientes de regressões lineares com níveis de significância e testes t de média das variáveis foram usados para verificação das hipóteses.
- b) **Parte II – resultados do filtro de Kalman:** nesta etapa foram estimados os parâmetros do modelo proposto e do modelo de dois fatores por meio do filtro de Kalman, foram geradas séries de previsão n passos à frente para os dois modelos.
- c) **Parte III – resultados dos testes de capacidade de previsão:** nesta etapa os modelos foram testados e os resultados interpretados com o intuito de responder à pergunta final da pesquisa: se o modelo proposto é mais eficiente, em termos de previsão, do que outro modelo da literatura adequado ao mercado de açúcar.

Antes dos resultados acima mencionados, foram apresentadas as estatísticas descritivas das variáveis e os resultados dos testes nas variáveis utilizadas nos modelos com defasagem temporal, como os testes de raiz unitária e cointegração. Também foram apresentados antecipadamente os resultados da estimação do modelo ARMA / GARCH para a volatilidade.

5.1 Estatísticas Descritivas

A Tabela 2, a seguir, apresenta as estatísticas descritivas das séries de dados diários e semanais das variáveis brutas, ou seja, sem transformação.

Tabela 2 – Estatística descritiva das variáveis brutas – dados diários

Estatística	FUT	ACUCARBR	PETR	R	DJONES	ETANOL
Média	10,439	11,629	53,573	0,028	10743,357	333,369
Mediana	9,843	11,240	52,020	0,022	10542,550	329,489
Desvio-padrão	3,523	4,156	24,690	0,016	1531,391	113,856
Curtose	2,733	2,958	3,952	1,495	2,429	1,994
Assimetria	0,691	0,750	1,005	0,340	0,198	0,155
Mínimo	5,334	5,212	17,970	0,009	7286,270	126,916
Máximo	20,185	23,007	140,210	0,053	14164,530	575,750
Jarque-Bera	134,442	152,997	336,267	185,290	32,841	15,836
p-valor (JB)	0,000	0,000	0,000	0,000	0,000	0,000
Número de Observações	1631	1631	1631	1631	1631	343
Periodicidade	Diária	Diária	Diária	Diária	Diária	Semanal

Na nona linha da Tabela 2, são apresentados os p-valores do testes Jarque-Bera⁹ de normalidade das variáveis; com base na interpretação dos resultados, rejeita-se a hipótese nula de normalidade para todas as variáveis.

⁹ O teste conjunto de assimetria e curtose Jarque-Bera está explicitado no Apêndice I.

A Tabela 3 apresenta as estatísticas descritivas das séries de dados mensais. Pela análise dos resultados, não se rejeita a hipótese de normalidade, ao nível de 5% de significância, para o preço futuro do açúcar e Índice Dow Jones.

Tabela 3 – Estatística descritiva das variáveis brutas – dados mensais

Estatística	FUT	ACUCARBR	PETR	R	DJONES
Média	10,461	11,659	54,561	0,029	10753,140
Mediana	10,070	11,263	51,865	0,025	10489,005
Desvio-padrão	3,514	4,203	25,541	0,016	1552,525
Curtose	2,599	2,880	3,913	1,536	2,432
Assimetria	0,629	0,691	0,987	0,344	0,177
Mínimo	5,534	5,375	19,480	0,009	7591,930
Máximo	19,123	22,424	140,000	0,053	13930,010
Jarque-Bera	5,672	6,249	15,365	8,505	1,458
p-valor (JB)	0,059	0,044	0,000	0,014	0,482
Número de Observações	78	78	78	78	78
Periodicidade	Mensal	Mensal	Mensal	Mensal	Mensal

Na Ilustração 4, a seguir, estão representados os gráficos das séries diárias da amostra das variáveis: preço futuro do açúcar, preço à vista do açúcar, preço do barril de petróleo WTI e o Índice Dow Jones, no período de 2 de janeiro de 2002 a 30 de junho de 2008.

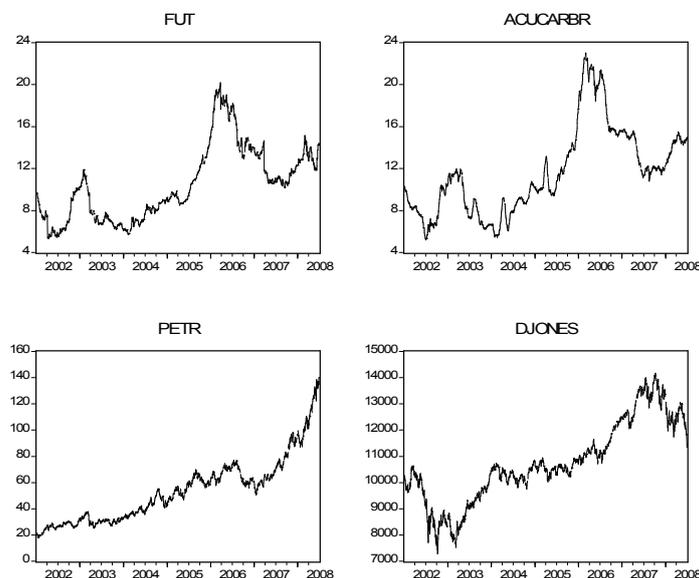


Ilustração 4 - Gráficos das variáveis – dados diários

Na Ilustração 5, está a representação gráfica das amostras com dados de preços do etanol e açúcar à vista, em periodicidade semanal. Os dados estão dispostos nessa periodicidade porque os preços do etanol à vista são divulgados em base semanal, apesar de serem coletados diariamente, segundo o CEPEA(2008).

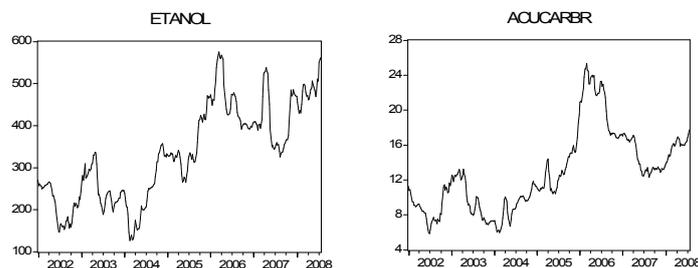


Ilustração 5 - Gráficos do etanol e açúcar a vista

5.2 Estimação da Volatilidade do Mercado

Para o cálculo da volatilidade do mercado, foram utilizados os retornos contínuos do Dow Jones, índice representativo do mercado de ações. A volatilidade foi estimada pela variância condicional resultante de um processo ARMA / GARCH. Para especificar o processo heterocedástico, foram analisados os resíduos quadráticos resultantes de uma regressão AR(1) dos retornos do Índice Dow Jones.

5.2.1 Autocorrelação dos Resíduos Quadráticos

Na Tabela 4, a seguir, são apresentados os coeficientes de autocorrelação dos resíduos quadráticos resultantes de um processo AR(1) com os retornos diários do Índice Dow Jones. A primeira coluna da tabela indica a defasagem da correlação dos resíduos quadráticos, a segunda e terceira colunas são os índices de autocorrelação e autocorrelação parcial. As duas últimas colunas da tabela contêm as estatísticas *Q Ljung Box* e seus respectivos p-valores:

Tabela 4 – Autocorrelação dos resíduos quadráticos – dados diários

Defasagem	Autocorrelação	Autocorrelação Parcial	Estatística <i>Q Ljung Box</i>	P-valor
1	0,1320	0,1320	28,6130	0,0000
2	0,2270	0,2130	112,9100	0,0000
3	0,3870	0,3570	357,2800	0,0000
4	0,1770	0,0970	408,7600	0,0000
5	0,1950	0,0510	471,0600	0,0000

Os resultados do correlograma acima levam à rejeição da hipótese nula de autocorrelação igual a zero, pois os p-valores da coluna à direita são inferiores a 0,05, o que indica a presença de um processo heterocedástico GARCH nas séries diárias de dados. Na Tabela 5 são

apresentados os resultados do mesmo teste, realizados, porém, com os dados mensais dos resíduos quadráticos do processo AR(1) do Índice Dow Jones.

Tabela 5 – Autocorrelação dos resíduos quadráticos – dados mensais

Defasagem	Autocorrelação	Autocorrelação Parcial	Estatística Q <i>Ljung Box</i>	P-valor
1	0,2801	0,2801	6,2023	0,0128
2	0,1764	0,1063	8,6959	0,0129
3	0,2168	0,1567	12,5110	0,0058
4	0,0317	-0,0841	12,5936	0,0134
5	0,0491	0,0203	12,7945	0,0254

Da mesma forma que nos dados diários, os resultados dos testes estatísticos com dados mensais, apresentados na última coluna da Tabela 5, indicam a rejeição da hipótese nula a 5%, ou seja, indicam presença de processo GARCH nos resíduos quadráticos também para os dados mensais. Na primeira linha da tabela, é apresentado o coeficiente de autocorrelação com defasagem de um período igual a 0,2801.

5.2.2 Resultados do Modelo GARCH

Partindo dos resultados dos testes com os resíduos quadráticos, foi possível especificar um modelo de volatilidade condicional ARMA / GARCH. Foram estimados modelos AR(1), GARCH(1,1) com os retornos do Índice Dow Jones, cujas estatísticas z dos coeficientes são apresentadas na penúltima coluna da Tabela 6. Os resultados indicam a rejeição da hipótese nula de igualdade a zero do coeficiente GARCH nos modelos. Os resultados são apresentados nas Tabelas 6 e 7, a seguir, para os dados diários e mensais, respectivamente.

Tabela 6 – Coeficientes do processo GARCH – dados diários

	Coefficiente	Erro-Padrão	Estatística z	P-valores
AR(1)	-0,0651	0,0274	-2,3740	0,0176
Equação da Variância				
C	0,0000	0,0000	4,5332	0,0000
Resíduo quadrático (-1)	0,0554	0,0083	6,6931	0,0000
GARCH (-1)	0,9358	0,0093	100,5469	0,0000

As tabelas 6 e 7 foram divididas em duas partes, a parte superior das tabelas apresenta os coeficientes autorregressivos e suas respectivas estatísticas; na parte inferior, estão as estatísticas dos coeficientes da equação de variância condicional GARCH. A segunda coluna da tabela apresenta o valor dos coeficientes estimados no modelo GARCH e a última coluna, à direita, apresenta os p-valores dos coeficientes; p-valores inferiores a 0,05 indicam a rejeição do coeficiente igual a zero. Os resultados levam à rejeição da hipótese nula do coeficiente GARCH nas duas tabelas.

A Tabela 7 apresenta os resultados com dados mensais.

Tabela 7 – Coeficientes do processo GARCH – dados mensais

	Coeficiente	Erro-Padrão	Estatística z	P-valores
AR(1)	0,0100	0,1627	0,0618	0,9507
Equação da Variância				
C	0,0000	0,0000	0,8660	0,3865
Resíduo quadrático (-1)	0,1601	0,1673	0,9568	0,3387
GARCH(-1)	0,7889	0,1654	4,7686	0,0000

O resultado do teste do coeficiente GARCH acima indica a presença de um processo heterocedástico, porém, o coeficiente AR(1) não é significativamente diferente de zero nos resultados com dados mensais, apenas o coeficiente GARCH é estatisticamente diferente de zero. Com o intuito de tornar o modelo mais parcimonioso, o coeficiente AR foi excluído e considerou-se apenas o coeficiente GARCH.

Os resultados do modelo sem o componente AR para dados mensais são apresentados na Tabela 8, o coeficiente GARCH permanece significativa.

Tabela 8 – Coeficientes do processo GARCH sem processo AR – dados mensais

Equação da Variância	Coeficiente	Erro-Padrão	Estatística z	P-valores
C	0,0000	0,0001	0,7120	0,4765
Resíduo quadrático (-1)	0,2013	0,1725	1,1672	0,2431
GARCH(-1)	0,7646	0,1875	4,0772	0,0000

A volatilidade condicional foi calculada pela raiz quadrada da variância condicional gerada pelo modelo GARCH. A Ilustração 6 representa graficamente os retornos do Índice Dow Jones – RETDJONES – nos gráficos superiores e a volatilidade condicional calculada pelo processo GARCH – VOL – nos gráficos inferiores, para dados diários e mensais. No gráfico, é possível identificar os momentos de maior oscilação nos retornos do índice e, conseqüentemente, um maior índice de volatilidade.

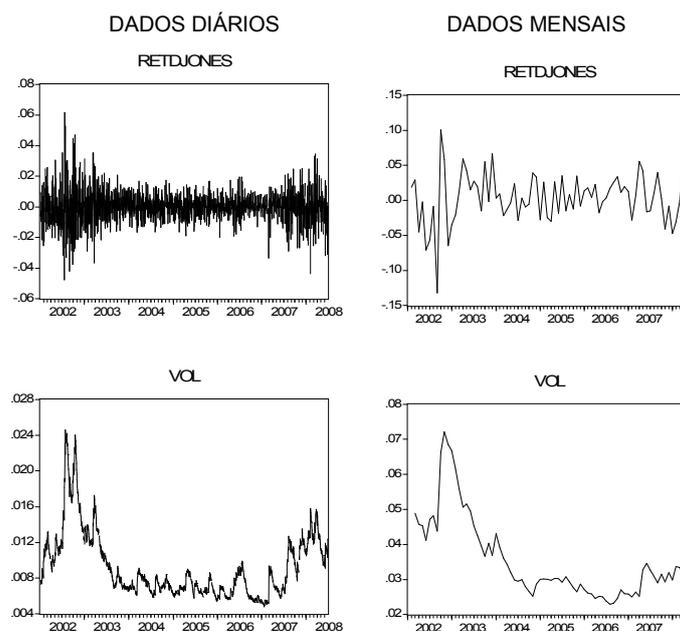


Ilustração 6 - Volatilidade GARCH do Índice Dow Jones

Na seção seguinte são apresentados os resultados dos testes de raiz unitária e cointegração nas variáveis, os testes foram necessários para verificar a estacionariedade das variáveis de entrada do modelo.

5.3 Resultados dos Testes de Raiz Unitária e Cointegração

Antes de realizar os testes de hipóteses das premissas, foram realizados testes de raiz unitária Dickey e Fuller (1979:428) Aumentado – ADF – nas variáveis do modelo. O objetivo desses testes é verificar se as variáveis são estacionárias, condição necessária para que os testes de significância dos estimadores dos coeficientes das regressões tenham validade. Se os resultados dos testes indicarem que as variáveis não são estacionárias, elas devem ser

transformadas; caso contrário, os coeficientes resultantes das regressões serão enviesados e inconsistentes. As variáveis foram testadas em nível, primeira e segunda diferenças e os resultados dos testes ADF com dados diários são apresentados na Tabela 9, a seguir. Entre parênteses estão os p-valores do teste t de raiz unitária.

Tabela 9 – Estatísticas dos testes ADF de raiz unitária – dados diários

Variável	Nível	Primeira diferença	Segunda diferença
Preço do Açúcar à vista	-1,3547 (0,6057)	-11,8934 (0,0000)	-24,2093 (0,0000)
Preço do Petróleo	1,7214 (0,9997)	-41,9535 (0,0000)	-20,7435 (0,0000)
Volatilidade	-2,3161 (0,1454)	-21,1357 (0,0000)	-19,2263 (0,0000)

Conforme o resultado dos testes apresentados na tabela acima, todas as variáveis são integradas de ordem um, $I(1)$, e são estacionárias em primeira e segunda diferença, ou seja, rejeitou-se a hipótese nula de não estacionariedade, com 5% de significância, para as variáveis transformadas em primeira diferença.

Os resultados com dados mensais são apresentados na Tabela 10 e foram similares aos verificados na tabela anterior, ou seja, as variáveis também são integradas de ordem um $I(1)$:

Tabela 10 – Estatísticas dos testes ADF de raiz unitária – dados mensais

Variável	Nível	Primeira diferença	Segunda diferença
Preço do Açúcar à vista	-1,3293 (0,6121)	-8,2037 (0,0000)	-9,3591 (0,0000)
Preço do Petróleo	2,2916 (0,9999)	-7,7158 (0,0000)	-9,0444 (0,0000)
Volatilidade	-1,7136 (0,4204)	-8,5736 (0,0000)	-5,6948 (0,0000)

Variáveis $I(1)$ podem cointegrar, por essa razão; além do teste ADF, tomou-se o cuidado de também testar a presença de cointegração. Os resultados dos testes, para amostras de dados diários, são apresentados na Tabela 11 e para os dados mensais na Tabela 12. A hipótese nula do teste de cointegração supõe que as variáveis não cointegram. O objetivo da pesquisa é a não rejeição da hipótese nula. A primeira coluna das tabelas indica o número de cointegrações entre as variáveis. Caso a hipótese nula não seja rejeitada na primeira linha da tabela, ou seja, nenhuma das variáveis cointegrarem, indica que as duas outras hipóteses também não serão rejeitadas, por predominância. A terceira coluna apresenta a estatística do teste, cujos valores críticos estão nas duas últimas colunas da tabela:

Tabela 11 – Estatísticas do teste de cointegração – dados diários

Número de Cointegrações	<i>Eigenvalue</i>	Estatística Trace	Valor crítico a 5%	Valor crítico a 1%
Nenhuma	0,0077	20,4215	29,6800	35,6500
No máximo uma	0,0041	7,8667	15,4100	20,0400
No máximo duas	0,0007	1,1791	3,7600	6,6500

De acordo com os resultados do teste, não se rejeitou a hipótese nula de não cointegração, ou seja, exceto na ocorrência de erro estatístico do tipo II, não se pode afirmar que as variáveis sejam cointegradas. A tabela seguinte apresenta os resultados do mesmo teste, porém com dados mensais. A não rejeição da hipótese nula na primeira linha da tabela corrobora os resultados obtidos com os dados diários.

Tabela 12 – Estatísticas do teste de cointegração – dados mensais

Número de Cointegrações	<i>Eigenvalue</i>	Estatística Trace	Valor crítico a 5%	Valor crítico a 1%
Nenhuma	0,1667	20,7316	29,6800	35,6500
No máximo uma	0,0730	7,2366	15,4100	20,0400
No máximo duas	0,0217	1,6269	3,7600	6,6500

Os resultados indicam que as variáveis, apesar de serem integradas de ordem um, não cointegram. Portanto, com base nesses resultados, é permitido transformar as variáveis para que se tornem estacionárias em primeira diferença antes de adicioná-las ao modelo. As variáveis (i) preço do açúcar à vista e (ii) do petróleo, foram transformadas pela primeira diferença dos logaritmos, enquanto que (iii) a volatilidade do mercado foi transformada apenas pela primeira diferença.

- **Resultados com o Etanol**

O teste de raiz unitária também foi realizado para a série de preços do etanol e açúcar à vista, cujos dados foram coletados em base semanal. Verificou-se a estacionariedade dessas duas variáveis, condição necessária para a realização dos testes estatísticos referentes à **Pergunta 1** da pesquisa. Na Tabela 13, a seguir, são apresentados os resultados do teste ADF para as séries de etanol e açúcar à vista, os p-valores estão entre parênteses.

Tabela 13 – Estatística do teste ADF para etanol e açúcar à vista

Variável	Nível	Primeira diferença	Segunda diferença
Preço do Etanol	-1,4353 (0,5652)	-11,2538 (0,0000)	-18,2648 (0,0000)
Preço do Açúcar à Vista	-1,3609 (0,6017)	-11,3397 (0,0000)	-13,1483 (0,0000)

Fonte: elaboração própria

De acordo com os resultados do teste acima, o etanol é uma variável integrada de ordem um, $I(1)$, que se torna estacionária em primeira diferença.

O açúcar, confirmando os resultados anteriores, também é uma $I(1)$. Verificou-se também a presença de cointegração entre essas duas variáveis; os resultados do teste estão disponíveis na Tabela 14.

Tabela 14 – Estatísticas do teste de cointegração entre etanol e açúcar à vista

Número de Cointegrações	<i>Eigenvalue</i>	Estatística Trace	Valor crítico a 5%	Valor crítico a 1%
Nenhuma	0,0274	10,8432	15,4100	20,0400
No máximo uma	0,0043	1,4387	3,7600	6,6500

De acordo com os resultados do teste de cointegração do açúcar e do etanol na tabela acima, não se rejeita a hipótese nula de não cointegração, ou seja, as variáveis também podem ser transformadas em primeira diferença para os testes estatísticos das premissas, cujos resultados são apresentados na próxima seção.

5.4 Parte 1 – Resultados dos Testes das Premissas

Nesta seção serão analisados os resultados do testes das hipóteses da pesquisa. O objetivo desses testes, nesta seção, é verificar as premissas fundamentais para o modelo de formação de preços de *commodities* e tentar responder às cinco primeiras perguntas do processo de investigação, com base em inferência estatística. Os testes, cujos resultados são apresentados nesta seção, tiveram a metodologia discutida no item 4.2.1 do capítulo anterior.

5.4.1 Testes Relativos à Pergunta 1

O teste de hipótese relativos à primeira pergunta da pesquisa procurou identificar a relação entre os preços do etanol e do açúcar à vista, divulgados pelo CEPEA. O intuito deste teste foi verificar a premissa de que variações no preço do etanol, provocadas pelas forças de oferta e demanda por esse biocombustível, podem ser transferidas para o preço do açúcar à vista, ou o contrário, se oscilações no preço do açúcar de cana-de-açúcar podem provocar variações nos

preços do etanol. O que motivou essa relação foi o fato de que as usinas de cana-de-açúcar podem alterar a composição da produção dessas mercadorias, com o intuito de maximizar a receita esperada.

A relação que existe entre os preços foi verificada pelo teste de causalidade de Granger, que verificou, por meio de um teste de hipótese nula conjunta de coeficientes defasados, se o açúcar antecipa os preços do etanol, ou vice-versa. Os testes são relativos às Equações 4.1., definidas no capítulo anterior. Os resultados da estatística F do teste de causalidade de Granger são apresentados na Tabela 15 para defasagem $d = 2$. Os resultados do teste indicam que o preço à vista do açúcar Granger causa o preço à vista do etanol e não o contrário. Essa inferência foi possível porque o p-valor da última linha da tabela abaixo é menor do que 0,05.

Tabela 15 – Teste de causalidade de Granger para etanol e açúcar – dados semanais

Hipótese Nula	Estatística F	p-valor
Etanol não Granger causa o Açúcar (a)	0,0166	0,9836
Açúcar não Granger causa o Etanol (b)	7,2916	0,0008

Resultado do teste de hipótese – Pergunta 1: de acordo os resultados do teste de causalidade de Granger, rejeitou-se a hipótese nula de que não existe relação na formação de preços entre o açúcar e o etanol de cana-de-açúcar. Também foi rejeitada a hipótese alternativa um, na qual o açúcar é Granger causado pelo etanol. Portanto, é possível inferir que a trajetória do preço do açúcar à vista determina o preço do etanol, ou seja, a hipótese alternativa número dois é a correta. Partindo do pressuposto de que os preços dessas duas mercadorias estejam intimamente relacionados, seria correto também inferir que o modelo de formação de preço do açúcar aqui proposto também pode ser aplicado ao mercado de etanol.

Na próxima seção serão apresentados os resultados dos testes relativos à presença de *backwardation* nos preços do açúcar.

5.4.2 Testes Relativos à Pergunta 2

A pergunta dois da pesquisa investiga a presença de *backwardation* na amostra de preços do mercado futuro de açúcar na BM&FBOVESPA. *Backwardation* é um valor em moeda que mede a diferença entre o preço à vista e futuro da *commodity*. O *backwardation* forte ocorre quando os preços à vista da mercadoria são maiores do que os preços no mercado futuro, na maior parte da amostra, ou seja, $S_t > F_t$. O *backwardation* fraco ocorre quando o preço à vista é maior do que o preço no mercado futuro descontado a valor presente pela taxa livre de risco, ou seja, $S_t > F_t e^{-r(T-t)}$. A presença de *backwardation* forte e significativamente positivo na amostra indica também que a taxa de retorno de conveniência no mercado de açúcar é maior do que a taxa livre de risco, conforme a Teoria da Estocagem. Ou seja, o benefício, medido em termos de taxas de juros, que o detentor da mercadoria física possui é maior do que a taxa livre de risco, ou seja, $(r - \delta_t)$ terá um valor esperado negativo se $S_t > F_t$ ¹⁰.

Para verificar a hipótese de presença de *backwardation* fraco e forte foram realizados testes *t* unicaudais sobre a média, com variância desconhecida calculados com os dados do mercado futuro e à vista de açúcar.

¹⁰ Mais detalhes sobre a premissa de *backwardation* e retorno de conveniência foram apresentados nos itens 2.3.2 e 2.3.4.

Os resultados do teste e as estatísticas descritivas do *backwardation* são apresentados na Tabela 16 para dados diários, e na Tabela 17 para dados mensais. A primeira linha da Tabela 16 apresenta o percentual das observações da amostra que apresentaram *backwardation*, ou seja, 87,49% da amostra em dados diários apresentaram *backwardation* forte. Na última linha da tabela estão os p-valores das estatísticas *t* unicaudais para média. A interpretação dos resultados dos p-valores leva à rejeição da hipótese nula das médias do *backwardation*, tanto o fraco quanto o forte, serem menores do que zero, ou seja, os resultados indicam presença de *backwardation* forte na amostra:

Tabela 16 – Teste *t* unicaudal de média do *backwardation* – dados diários

Estatística	<i>Backwardation Forte</i>	<i>Backwardation Fraco</i>
Presença na amostra	87,49%	88,60%
Média	1,1909	1,2407
Mediana	0,9437	0,9872
Desvio-padrão da amostra	1,1202	1,1421
Variância da amostra	1,2549	1,3045
Curtose	2,9835	2,9648
Assimetria	0,7671	0,7647
Mínimo	-1,0850	-1,0793
Máximo	4,8868	4,9442
Jarque-Bera	159,9666	159,0477
p-valor (JB)	0,0000	0,0000
Numero de Observações	1631	1631
Hipótese nula	$\mu_{B^s} < 0$	$\mu_{B^w} < 0$
Estatística <i>t</i> para média	42,9319	43,8695
p-valor unicaudal (<i>t</i>)	0,0000	0,0000

Para realização do teste de média de *backwardation*, foi necessário assumir que a média da amostra possui distribuição normal, com $\bar{B} \sim N(\mu_B, \sigma^2/n)$. O teste Jarque-Bera de

normalidade, cujo p-valor está apresentado na décima linha da Tabela 16 acima, indica a rejeição de normalidade. Essa rejeição ocorre porque, em amostras muito grandes, com $N \rightarrow \infty$, como é o caso da amostra com dados diários, o poder do teste Jarque-Bera é aumentado, e se a curtose for um pouco diferente de 3, ou a assimetria diferente de 0, o teste indicará rejeição da hipótese nula de normalidade. Contudo, para grandes amostras, pode-se utilizar o Teorema do Limite Central – TLC e considerar que a média possui distribuição normal, conforme sugerido por Bussab e Morettin (2004:272).

Com dados mensais, não se pode fazer a mesma assunção, pois a amostra não é grande o suficiente para se adotar o TLC. O tamanho necessário da amostra depende do erro amostral da média. Utilizando o desvio-padrão da amostra apresentado na Tabela 17, o erro amostral da média estimado para o *backwardation* forte é de $1,1642/\sqrt{78} = 0,1318$. Logo, o tamanho da amostra deveria ser, no mínimo¹¹, no nível de 5% de significância, de $N = 1,3553(1,6449)^2 / (0,1318)^2 = 211$ observações, para se assumir o TLC. Como a amostra de dados mensais possui apenas 78 observações, o resultado do teste Jarque-Bera deve ser levado em consideração na análise.

¹¹ A determinação do tamanho da amostra, segundo Bussab e Morettin (2004:280) é $N = \sigma^2 z^2 / \varepsilon^2$, em que ε é o erro amostral da média e z o valor da variável Normal Padronizada para um dado nível de significância e σ^2 a variância estimada.

Os resultados dos testes e estatísticas descritivas para as amostras mensais são apresentados na Tabela 17, a interpretação dos resultados corrobora os resultados com dados diários, ou seja, o *backwardation* forte é estatisticamente maior do que zero, e o p-valor do teste Jarque-Bera indica a não rejeição da hipótese de normalidade no nível de 5% de significância. Na última linha da Tabela 17, o p-valor do teste *t* unicaudal indica a rejeição da hipótese nula de $\mu_B < 0$, para *backwardation* fraco e forte.

Tabela 17 – Teste *t* unicaudal de média do *backwardation* – dados mensais

Estatística	<i>Backwardation</i> Forte	<i>Backwardation</i> Fraco
Presença na amostra	85,90%	85,90%
Média	1,1977	1,2566
Mediana	0,9557	1,0494
Desvio-padrão da amostra	1,1642	1,1923
Variância da amostra	1,3553	1,4216
Curtose	2,5388	2,5334
Assimetria	0,5844	0,5838
Mínimo	-0,7787	-0,7301
Máximo	3,7292	3,7975
Jarque-Bera	5,1315	5,1381
p-valor (JB)	0,0768	0,0766
Número de observações	78	78
Hipótese nula	$\mu_{B^s} < 0$	$\mu_{B^w} < 0$
Estatística <i>t</i> para média	9,0862	9,3077
p-valor unicaudal (<i>t</i>)	0,0000	0,0000

Resultado do teste de hipótese – Pergunta 2: de acordo com os resultados dos testes *t* de média, com variância desconhecida, rejeitou-se a hipótese nula de ausência de *backwardation* nas amostras de preços do açúcar. Com base nesses resultados, é possível inferir que os preços no mercado futuro de açúcar na BM&FBOVESPA apresentam *backwardation* forte e

significativamente positivo. A hipótese de presença de *backwardation* forte é predominante sobre o *backwardation* fraco, pois, havendo presença do primeiro, o segundo necessariamente também estará presente. Os resultados também indicam que o retorno de conveniência é maior do que a taxa livre de risco.

5.4.3 Testes Relativos às Perguntas 3, 4 e 5

Os resultados dos testes estatísticos de hipóteses relativos às perguntas 3, 4 e 5 da pesquisa são apresentados nesta seção. Para o processo de investigação, nesta etapa, foi utilizada a técnica estatística de regressão linear múltipla, cujos coeficientes são estimados e analisados simultaneamente por meio de testes *t* bicaudais de igualdade a zero. O que se deseja investigar, com a análise dos resultados dos testes, é a relação dos preços do petróleo, da volatilidade do mercado e do componente sazonal na formação dos preços do açúcar à vista. A relação dessas variáveis são premissas fundamentais para a especificação do modelo de formação de preços de *commodities* na forma de espaço de estado.

A aplicação da técnica de regressão múltipla MQO foi feita em quatro etapas:

- na primeira etapa, foi verificada a **estacionariedade** das variáveis, por meio dos testes de raiz unitária, cujos resultados foram apresentados no item 5.3 anterior. Verificou-se que as variáveis são estacionárias em primeira diferença e não cointegram;
- na segunda etapa dos testes, foi analisado o **correlograma** serial cruzado entre o retorno do petróleo e o retorno do açúcar à vista para identificar a defasagem *d* da variável na regressão;
- na terceira etapa, são apresentados os resultados da **regressão** e as estatísticas dos estimadores;

- na quarta e última etapa deste processo, são apresentadas as **análises dos resíduos**: dois testes foram realizados – testes de normalidade e correlação serial dos resíduos – e os resultados foram discutidos.

Para os testes de hipótese, adicionou-se o retorno contínuo do petróleo ao modelo multivariado com e sem defasagem, pois se pretendia verificar se o petróleo pode antecipar informação na tendência dos preços do açúcar. Essa é a premissa fundamental do modelo, pois sugere que a trajetória dos preços do açúcar pode ser antecipada pelos preços do petróleo. Além disso, foi analisada a premissa de que os preços do açúcar se elevam com o aumento da volatilidade do mercado, com base na teoria de que os investidores procuram esses mercados em tempos de turbulência econômica, em busca da proteção do valor real do capital. Testou-se também a hipótese de presença de sazonalidade nos preços do açúcar. Acredita-se que a safra da cana-de-açúcar no Brasil possa interferir na formação dos preços. A premissa de presença de ciclos no açúcar foi testada por meio dos estimadores dos coeficientes trigonométricos seno e cosseno do componente sazonal.

Os resultados do correlograma serial cruzado dos retornos do petróleo com o açúcar são apresentados na Tabela 18, para dados diários, e na Tabela 19, para dados mensais. A primeira coluna da tabela apresenta a defasagem medida em unidade de tempo; na segunda coluna estão as correlações calculadas com o petróleo defasado; na terceira coluna, com o açúcar defasado. Para análise, foram observados os valores da segunda coluna, ou seja, procurou-se identificar correlações positivas, em ordem crescente de defasagem, tendo o petróleo como variável defasada. De acordo com os resultados apresentados na Tabela 18, não se verificou correlação positiva com defasagem crescente nos dados diários. Defasagem zero indica correlação contemporânea entre as variáveis.

Tabela 18 – Correlograma serial cruzado entre petróleo e açúcar – dados diários

Defasagem	Correlação serial com petróleo defasado	Correlação serial com açúcar defasado
0	0,0394	0,0394
1	-0,0240	-0,0141
2	0,0220	-0,0175
3	0,0043	-0,0457

Com base no resultados do correlograma cruzado com dados diários, não se identificou correlação defasada entre o petróleo e o açúcar, logo, no modelo de regressão com dados diários, adotou-se defasagem igual a zero, $d = 0$.

Os resultados do correlograma com dados mensais, disponíveis na Tabela 19, foram distintos dos resultados com dados diários. Identificou-se correlação positiva com defasagem de uma unidade de tempo na segunda coluna da tabela, o índice de correlação calculado foi de 0,2083.

Tabela 19 – Correlograma serial cruzado entre petróleo e açúcar – dados mensais

Defasagem	Correlação serial com petróleo defasado	Correlação serial com açúcar defasado
0	0,0019	0,0019
1	0,2083	-0,0222
2	0,0638	-0,0014
3	-0,2337	-0,0101

Os resultados apresentados na tabela acima levaram à inclusão do retorno do petróleo com defasagem de um período ao modelo multivariado, quando estimado com dados mensais.

As Tabelas 20 e 21 apresentam os resultados da regressão multivariada para dados diários e mensais, respectivamente. A segunda coluna das tabelas apresenta os valores estimados dos coeficientes, na coluna à direita estão os valores da estatística t dos coeficientes, com os respectivos p-valores entre parênteses.

Tabela 20 – Regressão linear múltipla MQO – dados diários

Variável	Coefficiente	Erro-padrão (*)	Estatística t
Retorno do petróleo (t)	0,027	0,015	1,800 (0,0721)
Primeira diferença da volatilidade do mercado	0,551	0,842	0,654 (0,513)
Sazonalidade cosseno	0,001	0,001	2,100 (0,0359)
Sazonalidade seno	-0,001	0,001	-1,443 (0,1491)

NOTA: (*) estimador robusto de Newey e West (1987:704).

Com base nos resultados apresentados na Tabela 20, com dados diários, não se rejeita a hipótese nula de igualdade a zero para os coeficientes no nível de 5% de significância, exceto para o coeficiente cosseno do componente sazonal. Com 10% de significância estatística, é possível rejeitar a hipótese nula do coeficiente do retorno do petróleo, na forma contemporânea. O coeficiente do petróleo é positivo, como se pretendia obter, o que significa uma relação positiva com o retorno do preço do açúcar à vista. A volatilidade do mercado também possui coeficiente positivo, porém não significativo. O coeficiente de pelo menos um dos componentes trigonométricos é significativo a 5%, indicando presença de sazonalidade na amostra de dados diários.

Na Tabela 21, a seguir, estão os coeficientes estimados e as estatísticas da regressão com dados mensais.

Tabela 21 – Regressão linear múltipla MQO – dados mensais

Variável	Coefficiente	Erro-padrão	Estatística <i>t</i>
Retorno do petróleo ($t-1$)	0,3973	0,1609	2,470 (0,0159)
Primeira diferença da volatilidade do mercado	5,5075	2,9992	1,836 (0,0704)
Sazonalidade cosseno	0,0414	0,0193	2,138 (0,0359)
Sazonalidade seno	-0,0083	0,0194	-0,428 (0,6701)

Na regressão MQO com dados mensais, o retorno do petróleo, defasado em uma unidade de tempo, possui um coeficiente de 0,3973 positivo e significativo a 5%, o que indica que o petróleo antecipa informações sobre o mercado de açúcar quando são utilizados dados em periodicidade mensal. O coeficiente relativo à volatilidade do mercado é positivo e significativo apenas a 10% e pelo menos um dos componentes trigonométricos é significativo a 5%, confirmando presença de sazonalidade.

A Tabela 22 apresenta as estatísticas descritivas e os resultados dos testes de normalidade Jarque-Bera dos resíduos das duas regressões, cujos p-valores são apresentados na última linha da tabela. A hipótese nula do teste de normalidade é que os resíduos possuem distribuição normal.

De acordo com os resultados na tabela abaixo, rejeitou-se a hipótese nula de normalidade, a 5% de significância, apenas para a série de resíduos com dados diários. Ou seja, não se pôde rejeitar que os resíduos mensais sejam normais.

Tabela 22 – Estatísticas dos resíduos da regressão – teste de normalidade

Estatística	Resultados com dados diários	Resultados com dados mensais
Média	0,000	0,000
Mediana	0,000	-0,013
Máximo	0,097	0,385
Mínimo	-0,053	-0,293
Desvio-padrão	0,013	0,116
Assimetria	0,293	0,189
Curtose	6,877	4,245
Numero de Observações	1628	76
Hipótese nula	Distribuição Normal	Distribuição Normal
Jarque-Bera	1,043	5,361
p-valor (JB)	0,000	0,069

A hipótese de normalidade, segundo Greene (2003:65), pode ser relaxada pelas propriedades assintóticas de estimadores de parâmetros de grandes amostras, à medida que $N \rightarrow \infty$. Como o número de observações dos resíduos diários é muito grande, os estimadores podem ser considerados assintoticamente consistentes e não viesados.

Além dos testes de normalidade, foram realizados testes para identificar correlação serial nos resíduos. Os resíduos resultantes da regressão MQO devem possuir variância constante e não podem ser serialmente correlacionados. A presença de correlação serial nos resíduos resulta em erros-padrão incorretos que não podem ser utilizados para inferência estatística.

Na tabela 23 são apresentados os resultados da estatística *Q Ljung-Box* para correlação dos resíduos da regressão com dados diários. Os resultados indicam rejeição de correlação igual a zero e, conseqüentemente, presença de correlação serial:

Tabela 23 – Autocorrelação dos resíduos da regressão – dados diários

Defasagem	Autocorrelação	Estatística <i>Q Ljung Box</i>	p-valor
1	0,344	193,240	0,000
2	0,209	264,300	0,000
3	0,216	340,680	0,000

A presença de correlação serial nos resíduos diários indica que os erros-padrão estimados na regressão MQO podem não estar corretos. Greene (2003:267) sugere que, quando os resíduos resultantes da regressão apresentam heterocedasticidade ou correlação serial, é conveniente utilizar a matriz de covariância robusta de Newey e West (1987:704) para corrigir os erros-padrão e permitir a inferência estatística nos testes de hipóteses. Os erros-padrão apresentados na Tabela 20 tiveram correção da matriz de covariância de Newey-West¹², portanto são robustos, o que permitiu a inferência estatística com dados diários.

¹² Mais detalhes sobre a matriz de covariância robusta de Newey-West estão no Apêndice III.

A Tabela 24 apresenta os resultados dos testes de correlação serial dos resíduos para os dados mensais, os resultados na tabela indicam ausência de correlação serial para os resíduos mensais.

Tabela 24 – Autocorrelação dos resíduos da regressão – dados mensais

Defasagem	Autocorrelação	Estatística <i>Q</i> <i>Ljung Box</i>	p-valor
1	-0,093	0,678	0,410
2	-0,225	4,727	0,094
3	0,157	6,732	0,081

Resultado do teste de hipótese – Pergunta 3: para investigar a relação dos preços do petróleo WTI da NYMEX com o açúcar à vista divulgado pelo CEPEA, foram realizados testes estatísticos de hipóteses dos coeficientes de regressão linear múltipla. Os resultados dos testes indicaram a rejeição da hipótese nula e confirmaram a segunda hipótese, na qual se afirma a existência de relação serial linear entre os preços do açúcar e do petróleo, sendo que a trajetória de preços do petróleo antecipa informações sobre o preço do açúcar. O coeficiente do petróleo na regressão, com defasagem de uma unidade de tempo, é positivo e estatisticamente diferente de zero, para a estimação com dados mensais. Na amostra com dados diários, verificou-se relação contemporânea e positiva entre as variáveis, significativa apenas no nível de 10%.

Resultado do teste de hipótese – Pergunta 4: os resultados dos testes de coeficientes indicaram rejeição da hipótese nula com 10% de significância para a amostra de dados mensais, ou seja, pode-se inferir, de forma não conclusiva, que existe relação linear entre os preços do açúcar à vista e a volatilidade do mercado. O coeficiente é positivo, com um p-valor de 0,07 para dados mensais e de 0,51 para dados diários, ou seja, a hipótese nula não foi rejeitada nos testes com dados diários.

Resultado do teste de hipótese – Pergunta 5: rejeitou-se a hipótese nula na amostra com dados diários e mensais. Com 5% de significância, o coeficiente cosseno foi estatisticamente diferente de zero nas duas séries de dados. Para identificar sazonalidade, pelo menos um dos

coeficientes do componente sazonal, seno ou cosseno, deve ser estatisticamente diferente de zero. Logo, rejeitou-se a hipótese nula da pesquisa e foi confirmada a existência de sazonalidade nos preços do açúcar à vista.

5.5 Parte 2 – Resultados do Filtro de Kalman

Nesta seção são apresentados os resultados dos parâmetros do modelo, estimados por meio do filtro de Kalman. Com o modelo estimado, foram geradas séries de dados filtradas e previsões n passos à frente. Nesta parte da pesquisa, o preço do açúcar à vista foi tratado como variável não observável, portanto as cotações de preços do açúcar divulgadas pelo CEPEA foram omitidas no processo de filtragem. Foram utilizadas as cotações do açúcar no mercado futuro divulgadas pela BM&FBOVESPA.

O processo foi executado para obter os parâmetros estimados do modelo proposto e do modelo de dois fatores de Gibson e Schwartz (1990:960)¹³, o objetivo foi gerar séries de previsão para os dois modelos, para que os resultados fossem comparados na terceira parte dos testes estatísticos.

O modelo proposto neste trabalho, conforme especificado no item 3.5.7 no terceiro capítulo, é composto pelo seguinte do sistema de equações na forma de espaço de estado:

$$\begin{cases} \ln(F_t) = r(T-t) - (T-t)\delta_t + \ln(S_t) \\ \delta_t = \phi\mu_\delta + (1-\phi)\delta_{t-1} + \sigma_\delta z \\ \ln(S_t) = \beta_0 + \beta_1 R_{t-1}^P + \beta_2 \Delta \hat{\sigma}_t^M + \lambda_t + \ln(S_{t-1}) + \sigma_S z \end{cases} \quad (5.1)$$

O filtro de Kalman, conforme definido anteriormente, é um processo de estimação de variáveis de estado não observáveis em um sistema dinâmico linear, perturbado por um ruído branco com distribuição normal. O estimador resultante do processo é ótimo, em termos de erro quadrático de estimação.

¹³ O modelo de dois fatores foi especificado no item 4.2.6.

A Tabela 25 a seguir apresenta os resultados dos parâmetros estimados dos dois modelos para dados diários e mensais por meio do filtro de Kalman. A tabela é dividida em duas partes: nas colunas da esquerda estão os parâmetros estimados com os dados diários e à direita estão os estimadores com dados mensais. Paralelamente, na tabela, é possível comparar os estimadores para o modelo proposto e o de dois fatores, isso é possível porque o modelo de dois fatores também possui parâmetros de movimento browniano para o preço à vista e reversão à média para o retorno de conveniência.

Tabela 25 – Resultados do filtro de Kalman

Parâmetro	Dados diários		Dados Mensais	
	Modelo Proposto	Dois Fatores	Modelo Proposto	Dois Fatores
ϕ	0,008398	0,008339	1,112879	0,124878
μ_δ	0,116495	0,111897	0,143086	0,083884
$\exp(\sigma_\delta)$	-6,556159	-6,588269	-39,29111	-39,29111
β_0	0,000311	0,000254	0,011527	0,006749
β_1	0,022868	-	0,049383	-
β_2	1,069168	-	1,187362	-
$\beta_{c,1}$	0,001367	-	0,002386	-
$\beta_{s,1}$	-0,001836	-	-0,035554	-
$\exp(\sigma_s)$	-8,686838	-8,661802	-4,934495	-4,793600
<i>Log likelihood</i>	4555,53	4548,11	64,31	60,32

Com base nos resultados apresentados na Tabela 25 acima, existe uma relação positiva entre o retorno defasado do petróleo e o logaritmo do preço à vista do açúcar, tanto para dados diários, quanto para dados mensais, pois o coeficiente do petróleo β_1 é positivo. O estimador do coeficiente β_2 da volatilidade também é positivo. Esse resultado pode ser devido ao fato de que, quando há aumento da volatilidade, os preços do açúcar também aumentam.

Em termos de verossimilhança (*likelihood*), o modelo proposto possui maior verossimilhança do que o de dois fatores, tanto para dados diários, quanto mensais. Analisando os parâmetros estimados, verificou-se que a taxa média anual de longo prazo do retorno de conveniência μ_δ é de 11,64% a.a. e 14,31% a.a., para dados diários e mensais, respectivamente, no modelo proposto. A taxa média do retorno de conveniência é superior à taxa de juros média anual dos *Fed Funds* de 2,81% (taxa contínua). Esse resultado é coerente com a presença significativa de *backwardation* forte na amostra. Observando os resultados com dados mensais, verificou-se que o coeficiente de velocidade de reversão à média ϕ é substancialmente maior no modelo proposto do que no modelo de dois fatores.

5.5.1 Séries Filtradas e Resultados da Previsão n Passos à Frente

Utilizando os parâmetros estimados no processo de filtragem, é possível gerar séries de dados das variáveis filtradas e séries de previsão n passos à frente. Na Ilustração 7, está a representação gráfica da série de dados diários de preços de açúcar à vista observados no CEPEA e os preços do açúcar à vista filtrados do modelo proposto. Na Ilustração 8, estão os gráficos com dados mensais.

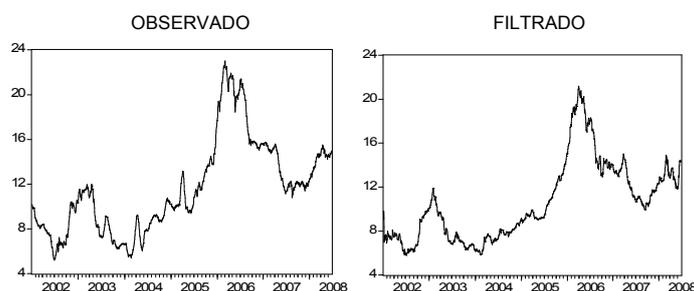


Ilustração 7 - Gráfico dos preços à vista do açúcar observados e filtrados – dados diários

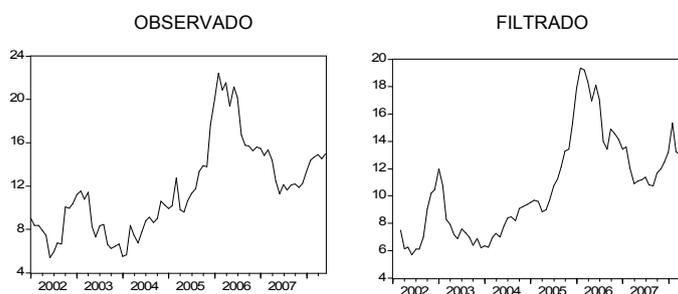


Ilustração 8 - Gráfico dos preços à vista do açúcar observados e filtrados – dados mensais

Com os parâmetros estimados, foram geradas séries de previsão para os dois modelos, para cada modelo foram geradas quatro séries de previsão: um mês e três meses à frente, com dados mensais e diários. Para as séries com dados mensais, as previsões foram geradas para uma e três unidades de tempo à frente. Nas séries de previsão com dados diários, considerou-se o mês com 21 dias úteis em média, logo a previsão foi feita para 21 passos à frente, enquanto, na previsão para três meses, consideraram-se 63 unidades de tempo à frente. Esse procedimento foi adotado porque a série de observações foi coletada nos dias em que houve pregão na BM&FBOVESPA, que ocorrem apenas em dias úteis.

Na Ilustração 9 estão as representações gráficas das previsões um mês e três meses à frente com dados diários do modelo proposto e do modelo de dois fatores; a Ilustração 10 apresenta os gráficos das previsões com dados mensais.

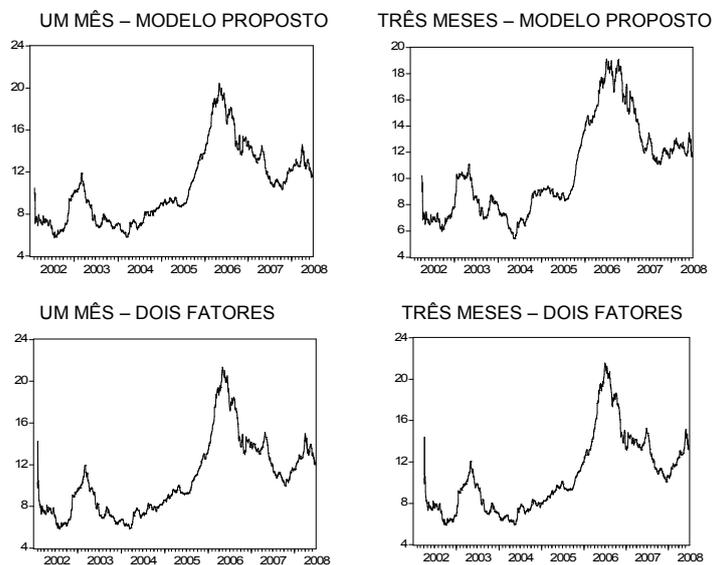


Ilustração 9 - Previsão n passos à frente do preço à vista do açúcar – dados diários

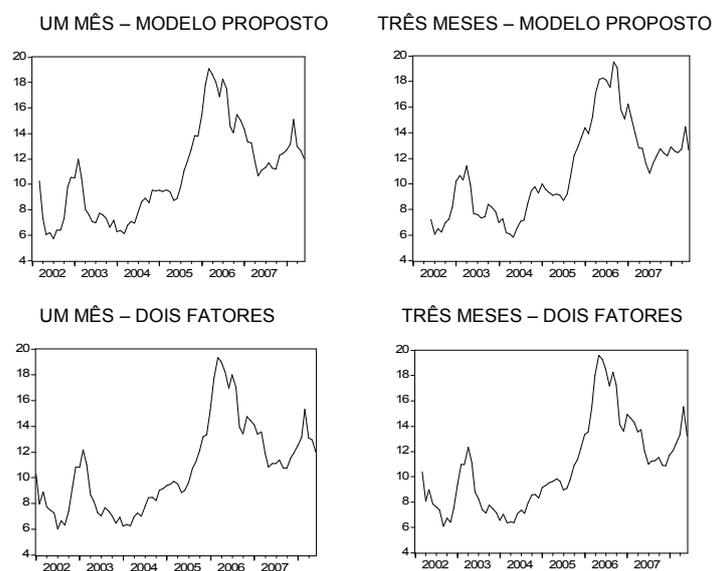


Ilustração 10 - Previsão n passos à frente do preço à vista do açúcar – dados mensais

5.6 Parte 3 – Resultados do Testes de Capacidade de Previsão

Nesta parte da pesquisa, os dois modelos, cujos parâmetros foram estimados pelo filtro de Kalman, foram comparados em termos de capacidade de previsão. Utilizando os parâmetros estimados dos modelos, foram geradas séries de previsão n-passos à frente para dados mensais e diários de preço à vista do açúcar. Para responder à pergunta final da pesquisa, as séries de preços previstos do açúcar dos dois modelos foram comparadas com os valores observados do açúcar à vista divulgados pelo CEPEA. As séries previstas e observadas foram emparelhadas, o preço observado do açúcar S_{ot+n} foi comparado com o preço previsto para a mesma data, $S_{f_{t+n}}$, ou seja, $\hat{S}_{ot+n|t} = S_{f_{t+n}} = E_t(S_{ot+n})$, n passos à frente.

Os testes de previsão dos modelos foram realizados em duas etapas, descritas a seguir:

- na primeira etapa foram calculadas medidas de erro de previsão para quantificar o erro entre a amostra observada e a série prevista. As medidas de erro foram RMSE, MAE e MAPE. O procedimento foi também adotado por Aiube, Baidya e Tito (2006:222) para mensurar o erro de previsão em modelos de formação de preços de *commodities*;
- na segunda etapa foi realizado o teste estatístico de hipótese de Granger e Newbold (1976:195), conforme sugerido por Enders (2004:83). Este teste verificou a hipótese final da pesquisa e identificou, estatisticamente, qual dos dois modelos é melhor em termos de previsão.

5.6.1 Medidas de Erro de Previsão

Nesta seção são apresentados os resultados das medidas de erro de previsão dos dois modelos para um e três meses à frente. A primeira medida – RMSE – é a raiz do erro quadrático médio, a segunda medida – MAE – é o erro absoluto médio e a terceira medida – MAPE – é o erro absoluto percentual médio. Na Tabela 26, são apresentados os resultados das medidas

para os dados diários e, na Tabela 27, os resultados com dados mensais. Quanto maior o valor absoluto das medidas, maior é o erro de previsão do modelo.

Tabela 26 – Medidas de erro de previsão – dados diários

Modelo	Previsão	RMSE	MAE	MAPE
Modelo proposto	Um mês	1,5877	1,2202	0,1169
	Três meses	2,2243	1,6887	0,1656
Modelo de dois fatores	Um mês	1,6198	1,2553	0,1191
	Três meses	2,4285	1,8093	0,1792

Tabela 27 – Medidas de erro de previsão – dados mensais

Modelo	Previsão	RMSE	MAE	MAPE
Modelo proposto	Um mês	1,6127	1,2045	0,1125
	Três meses	2,1638	1,5659	0,1477
Modelo de dois fatores	Um mês	1,6476	1,2431	0,1131
	Três meses	2,2769	1,6996	0,1655

Com base nos resultados apresentados nas Tabelas 26 e 27 acima, verificou-se que, em termos de previsão, o modelo proposto é superior. Nos três indicadores, o modelo proposto resultou em menores erros do que no modelo de dois fatores, tanto para dados diários, quanto para dados mensais. Verificou-se, também, que o modelo estimado a partir de dados mensais produz menos erros do que a partir de dados diários. Quanto mais distante o horizonte de previsão, mais eficiente é o modelo proposto, comparado ao modelo de dois fatores. O ganho de informação, com a redução do erro absoluto percentual médio (MAPE), para a previsão um mês à frente, para dados diários e mensais foi de 1,86% e 0,54%, respectivamente. Enquanto o ganho de informação para a previsão de três meses foi de 7,59% e 10,79%, para dados diários e mensais, ou seja, houve uma redução de erros de previsão de mais de 10% no modelo proposto para o horizonte de três meses, quando comparado com o modelo de dois fatores em dados mensais.

5.6.2 Resultado do Teste de Previsão Granger e Newbold

Os modelos foram testados estatisticamente para identificar qual dos dois modelos é mais eficiente em termos de previsão. O teste de Granger e Newbold (1976) permite fazer essa comparação e, para um dado nível de significância, permite determinar qual dos dois modelos é estaticamente melhor. Para executar o teste, foram necessárias séries de dados observados e previstos dos dois modelos, o erro de previsão foi calculado pela diferença entre o valor observado e o valor previsto¹⁴. A hipótese nula do teste é que os modelos são iguais, em termos de erros de previsão. Se o modelo 1 (proposto) tiver maior erro de previsão, a correlação $\rho_{w,g}$ será positiva; se o modelo 2 (dois fatores) errar mais, a correlação será negativa. De acordo com Granger e Newbold (1976), o coeficiente de correlação amostral $\rho_{w,g}$ tem uma distribuição t com $N-1$ graus de liberdade.

¹⁴ Os cálculos da estatística do teste de Granger e Newbold (1976) foram apresentados no item 4.2.6.

A Tabela 28 apresenta os valores de $\rho_{w,g}$ e, na última coluna à direita da tabela, as estatísticas t do teste bicaudal para os dois modelos, os testes foram realizados com dados diários e mensais e para previsões de um e três meses à frente, os p-valores estão entre parênteses.

Tabela 28 – Medidas de erro de previsão – dados mensais

Dados	Previsão	Graus de liberdade	$\rho_{w,g}$	Estatística t
Diários	Um mês	1606	-0,1024	-4,1261 (0,0000)
	Três meses	1564	-0,9504	-120,8194 (0,0000)
Mensais	Um mês	74	-0,1138	-0,9853 (0,3277)
	Três meses	71	-0,5022	-4,8941 (0,0000)

De acordo com os resultados dos testes da Tabela 28, todos os valores de $\rho_{w,z}$ são negativos, ou seja, em todas as amostras analisadas o modelo proposto resultou em menos erros de previsão do que o modelo de dois fatores. Além disso, rejeitou-se, com 99,9% de confiança, a hipótese nula de que os modelos sejam iguais em termos de previsão, com dados diários, para as previsões de um e três meses à frente. Com dados mensais, rejeitou-se a hipótese nula apenas para previsão de três meses.

Resultado do teste de hipótese – Pergunta Final: a pergunta final da pesquisa questionava se o modelo proposto é melhor, em termos de previsão, do que um outro modelo disponível na literatura adequado ao mercado de açúcar, no caso, o modelo selecionado para comparação foi o de dois fatores. Os resultados dos testes estatísticos indicaram a rejeição da hipótese nula da pergunta final da pesquisa, no nível de 1% de significância, para três das quatro amostras de erros de previsão analisadas. Com base nos resultados, é possível inferir que a segunda hipótese alternativa de pesquisa está correta, ou seja, que o modelo proposto é melhor, em termos de previsão, do que o modelo de dois fatores. Os resultados dos testes de hipóteses são corroborados pelas medidas de erros de previsão RMSE, MAE e MAPE. Em todas as

amostras, com dados mensais e diários, o modelo proposto resultou em menores erros de previsão, para um e três meses à frente.

5.7 Resumo dos Resultados da Pesquisa

Esta pesquisa envolveu um estudo empírico no mercado brasileiro de açúcar com base em testes estatísticos de hipóteses. O objetivo na análise dos resultados dos testes foi verificar a aplicabilidade do modelo proposto de formação de preços de *commodities* e das premissas previamente elaboradas.

Os testes estatísticos foram divididos em três partes inter-relacionadas e interdependentes. Na primeira parte foram testadas as premissas do modelo e respondidas as cinco primeiras perguntas da pesquisa; na segunda parte, os parâmetros do modelo foram estimados por meio do filtro de Kalman; e, na terceira parte do estudo estatístico, foi testada a capacidade de previsão do modelo aplicado ao mercado de açúcar, comparado ao modelo de dois fatores.

O estudo empírico foi iniciado com a estimativa da série de volatilidade de mercado calculada a partir da variância condicional gerada de um modelo ARMA / GARCH com os retornos do Índice Dow Jones. Foram realizados testes de raiz unitária e cointegração nas variáveis de entrada do modelo, nas séries de dados diárias, semanais e mensais. Os resultados dos testes indicaram que as variáveis são estacionárias em primeira diferença e não cointegram.

Na primeira parte dos testes, as premissas fundamentais do modelo foram verificadas por meio de testes estatísticos de hipótese. A primeira pergunta da pesquisa foi referente à relação de interdependência na formação de preços do etanol e do açúcar de cana-de-açúcar. O teste realizado nessa etapa foi o de causalidade de Granger e os resultados indicaram que o açúcar Granger causa os preços do etanol.

A segunda pergunta da pesquisa respondida nos testes foi referente à presença de *backwardation* na amostra de preços futuros e à vista no mercado de açúcar. Os resultados dos testes unicaudais de média, com variância desconhecida, indicaram a presença de *backwardation* forte, positivo e significativo tanto para as amostras de dados diários, quanto mensais. A presença de *backwardation* forte na amostra também indicou que o retorno de

conveniência é maior do que a taxa livre de risco, conforme a Teoria da Estocagem, resultado corroborado posteriormente com os parâmetros estimados no processo de filtragem.

As perguntas 3, 4 e 5 foram respondidas simultaneamente por testes dos coeficientes de uma regressão linear múltipla com as variáveis: petróleo, volatilidade do mercado e componente sazonal, tendo o preço à vista do açúcar, divulgado pelo CEPEA, como variável dependente. O modelo multivariado foi especificado em 4 etapas: (i) teste de raiz unitária e transformação das variáveis em primeira diferença; (ii) correlograma serial cruzado entre os retornos do petróleo e do açúcar para identificar defasagem; (iii) estimação MQO dos coeficientes do modelo; e (iv) teste dos resíduos do modelo.

Com relação à pergunta 3, os resultados confirmaram a relação de preços entre o petróleo e o açúcar. Com defasagem de um período para os dados mensais, o coeficiente foi positivo e significativo. Com dados diários, o coeficiente contemporâneo foi significativo no nível de 10%. Os resultados relativos à pergunta 4, que investigaram a relação da volatilidade do mercado com os preços do açúcar, não foram totalmente conclusivos, pois a hipótese nula foi rejeitada apenas no nível de 10% de significância para os dados mensais. Os testes com os coeficientes do componente sazonal responderam à pergunta 5 da pesquisa e confirmaram a presença de sazonalidade na amostra em todos os resultados, tanto para dados diários, quanto mensais, no nível de 5% de significância.

Na segunda parte do estudo estatístico, o modelo proposto e o modelo de dois fatores tiveram seus parâmetros estimados por meio do filtro de Kalman. Os parâmetros resultantes do processo de filtragem confirmaram a relação positiva entre o retorno defasado do petróleo e o preço à vista do açúcar, como também para o coeficiente estimado da volatilidade do mercado. A taxa média do retorno de conveniência ficou acima da taxa de juros média dos *Fed Funds*. Esse resultado é coerente com a presença de *backwardation* forte na amostra. Com os parâmetros estimados do modelo, foram geradas séries de previsão um e três meses à frente para os preços do açúcar.

Na terceira parte do estudo, os dois modelos foram comparados em termos de erro previsão. Nas três medidas de erro de previsão – RMSE, MAE e MAPE –, o modelo proposto resultou em menores erros de previsão do que no modelo de dois fatores, tanto para dados diários, quanto para dados mensais e, quanto maior o horizonte de previsão, melhor foi o resultado

obtido no modelo proposto. A previsão do modelo com dados mensais foi melhor do que com dados diários. O modelo proposto apresentou um ganho percentual médio (MAPE) de 10,79% na redução dos erros de previsão de três meses com dados mensais.

Na terceira parte foi realizado um teste de hipótese para identificar qual dos dois modelos é superior na previsão dos preços do açúcar. Foi realizado o teste de Granger e Newbold (1976), os resultados indicaram que o modelo proposto é mais eficiente estatisticamente do que o modelo de dois fatores para previsão dos preços do açúcar. Os resultados foram significantes no nível de 1% para as previsões de um e três meses à frente com dados diários, e três meses à frente para dados mensais. O coeficiente de correlação dos erros indicou que o modelo proposto é superior, em termos de previsão, ao modelo de dois fatores.

Os resultados dos testes permitem inferir que o modelo sugerido neste trabalho – que parte da teoria de que os preços de *commodities* utilizadas na produção de biocombustível têm uma relação de dependência com os preços do petróleo e que leva em consideração o efeito da volatilidade do mercado e do componente sazonal na formação de preços – é melhor, em termos de previsão, do que outro modelo que não incorpora essas variáveis.

6 CONCLUSÕES

O problema estudado nesta pesquisa foi o de formação de preços de *commodities* dentro do conceito de energia renovável, ou seja, de *commodities* agrícolas que, de alguma forma, estão relacionadas com a produção de biocombustíveis. No momento em que houve um aumento da demanda de energia e de insumos por países como China e Índia, ocorreu uma escassez circunstancial do petróleo, resultando na escalada momentânea de preços. Por esse motivo, a viabilidade de produção de bioenergia a partir de *commodities* agrícolas passou a ser tópico relevante de discussão econômica internacional. Em alguns casos, países como Brasil e Estados Unidos anteciparam-se ao processo de substituição dos combustíveis de origem fóssil e à produção de combustíveis alternativos, como o etanol derivado do milho ou da cana-de-açúcar.

O uso dos biocombustíveis produzidos a partir de gêneros alimentícios tornou-se assunto polêmico. Seus defensores afirmam que os produtos de origem vegetal podem sequestrar carbono da atmosfera e reduzir os efeitos do aquecimento global. Por outro lado, os críticos alegam que o uso de produtos agrícolas como fonte de energia pode causar escassez de alimentos e acelerar o desmatamento de florestas tropicais. Contudo, não é possível negar o fato de que as *commodities* agrícolas relacionadas com a produção de biocombustíveis estão em um novo contexto em termos de formação de preços. Por esse motivo, o modelo que pretende determinar preços dessas mercadorias deve levar em consideração os fatores nesse novo contexto.

O objetivo deste trabalho foi de elaborar um modelo formação de preços de *commodities*, tendo como base modelos anteriores da bibliografia e teorias pertinentes ao mercado de *commodities*, que incorporasse o fato de que as *commodities* agrícolas passaram a fazer parte da matriz energética. O modelo proposto de formação de preços teve como base premissas sobre a interdependência de preços com o petróleo, a Teoria de Estocagem, a sazonalidade e a volatilidade do mercado.

O modelo foi aplicado ao mercado brasileiro de açúcar, utilizando os preços dos negócios nos contratos futuros na BM&FBOVESPA e os preços do mercado à vista de açúcar e etanol divulgados pelo CEPEA. Para a especificação do modelo, foram elaboradas premissas

fundamentais, disputas na forma de cinco perguntas de pesquisa. Para responder as perguntas, foram formuladas hipóteses, sobre as quais foi possível realizar testes estatísticos de significância. Os testes foram realizados com séries de dados em periodicidade diária, mensal e, no caso do etanol, em periodicidade semanal. Antes de sua execução, foi verificada a presença de raiz unitária e cointegração nas variáveis, os resultados indicaram que as variáveis são estacionárias em primeira diferença e não cointegram.

Após a verificação das premissas, o modelo foi especificado na forma de equações de estado. O sistema possui três equações fundamentais. A primeira equação do modelo é de sinal e tem como base a condição de não arbitragem entre os preços no mercado futuro e à vista, em que o preço do contrato futuro é calculado pela diferença entre a taxa livre de risco e o do retorno de conveniência, aplicada sobre o preço à vista da *commodity*. A segunda equação do modelo define o processo de reversão à média do retorno de conveniência, considerada variável de estado no modelo. A terceira e mais importante equação do modelo é uma equação de estado que define o movimento browniano do retorno do preço à vista da *commodity*. Nessa equação, foram adicionados mais três fatores exógenos ao componente determinístico, dentre eles: o retorno do petróleo, a primeira diferença da volatilidade condicional do mercado e o componente sazonal. Os parâmetros das três equações foram estimados por meio do filtro de Kalman e os resultados analisados.

Com o intuito de verificar a capacidade de previsão do modelo estimado, foram geradas séries de previsão n passos à frente, e os resultados do erro de previsão comparados com os de outro modelo da literatura adequado para o mercado de açúcar. O modelo selecionado para a comparação foi o de dois fatores, proposto por Gibson e Schwartz (1990:960).

A primeira premissa verificada nos testes de hipótese foi a de interdependência na formação de preços do etanol e do açúcar à vista. Os resultados indicaram que os preços do etanol são Granger causados pelos preços do açúcar. Esse resultado permitiu inferir que o modelo pode ser aplicado tanto para o mercado de açúcar, quanto para o de etanol.

A segunda pergunta da pesquisa foi relativa à presença de *backwardation* nas séries de preços de açúcar à vista e futuro da BM&FBOVESPA. Foram realizados testes de média no nível de 5% de significância. Os resultados indicaram presença de *backwardation* forte, positivo e significativo na amostra, ou seja, na maior parte da amostra, os preços à vista do açúcar eram

maiores do que aqueles no mercado futuro. Se os preços do açúcar possuem *backwardation* forte, é possível concluir que, conforme a Teoria da Estocagem, a média do retorno de conveniência é maior do que a taxa livre de risco.

Foram realizados testes nos coeficientes de uma regressão linear múltipla MQO, para responder às perguntas 3, 4 e 5 da pesquisa. De acordo com os resultados, o petróleo antecipa informações sobre a tendência dos preços do açúcar no mercado brasileiro. O retorno do petróleo, com defasagem de uma unidade de tempo, possui relação positiva e significativa estatisticamente com o retorno do açúcar nas amostras com dados mensais. Esses resultados corroboraram a premissa de que as *commodities* agrícolas relacionadas com a produção de biocombustível sofrem alteração na tendência de preços com as variações na cotação do barril de petróleo. É possível concluir, a partir da análise dos resultados, que o modelo deve incorporar o conceito de interdependência na formação de preços com o petróleo.

A quarta premissa do modelo sugere que um aumento na volatilidade dos mercados também provocou uma elevação nos preços do açúcar, esse aumento aconteceu porque os participantes do mercado financeiro buscam o mercado de *commodities*, em períodos de turbulência econômica, ocasionados por crises ou surtos inflacionários, para proteger o valor real do capital. De acordo com os resultados dos testes, essa relação foi significativa apenas com dados mensais, no nível de 10% de significância. Os resultados não foram totalmente conclusivos, mas permitiram inferir, com risco majorado de ocorrência de erro estatístico do tipo I, que os investidores procuram o mercado de *commodities* em busca de proteção do capital. Porém, é importante salientar que a relação da volatilidade com o preço das *commodities* não é sustentável no longo prazo, pois não está relacionada com a demanda física pela mercadoria.

Os resultados dos testes de coeficientes também indicaram a presença de sazonalidade nos preços do açúcar no Brasil, devido aos ciclos de safra da cana-de-açúcar. O coeficiente de sazonalidade foi significativo no nível de 5% para as séries de dados diários e mensais. Os resultados reforçam a teoria de que, em períodos de safra, os preços do açúcar tendem a ser menores do que em períodos de entressafra. Esses resultados estão em conformidade com os conceitos da Teoria da Estocagem, pois, em momentos de estoques baixos, o preço da mercadoria tende a ser máximo.

O modelo especificado na forma de equações de estado teve seus parâmetros estimados pelo filtro de Kalman. Os parâmetros do petróleo e da volatilidade resultantes do processo de filtragem ficaram positivos. A média da taxa do retorno de conveniência, parâmetro do processo de reversão à média dessa variável, foi de 11,64% a.a. e de 14,31% a.a., para dados diários e mensais, respectivamente, ou seja, foi superior à taxa de juros média anual dos *Fed Funds*, de 2,81% no período analisado, o que também confirmou as conclusões sobre o *backwardation*.

Para comparar quantitativamente a capacidade de previsão dos dois modelos, foram geradas séries de previsão n passos à frente do preço de açúcar no mercado à vista, a partir dos parâmetros estimados no filtro de Kalman. Para cada observação da amostra, foram feitas previsões um mês e três meses à frente e os resultados foram confrontados com o real valor observado do açúcar divulgado pelo CEPEA.

Foram calculadas três medidas de erros de previsão – RMSE, MAE e MAPE – e os resultados mostraram que o modelo proposto possui menos erros de previsão do que o modelo de dois fatores para as três medidas, para amostra em dados diários e mensais. Para responder à pergunta final da pesquisa e verificar se o modelo proposto, em termos de previsão, foi melhor do que o modelo de dois fatores, foi realizado o teste estatístico de Granger e Newbold (1976) no nível de 1% de significância.

Os resultados do teste indicaram que o modelo proposto é superior estatisticamente ao modelo de dois fatores, para as previsões de um e três meses nos dados diários, e para três meses nos dados mensais. Além disso, os resultados indicaram que, quanto maior o horizonte de previsão, maior é o ganho de informação relativo do modelo proposto. O ganho de informação com a redução percentual de erros (MAPE) foi superior a 10%, na previsão de três meses à frente com dados mensais, quando comparado ao modelo de dois fatores.

Com base nesses resultados, portanto, é possível concluir que o modelo que incorpora a relação de preços com o petróleo, a volatilidade do mercado e a sazonalidade, é melhor, em termos de previsão, do que um outro modelo sugerido na literatura que não leva essas premissas em consideração, quando aplicado ao mercado brasileiro de açúcar.

O presente estudo científico resultou na elaboração de um modelo superior em termos de previsão, que pode ter aplicação pragmática na Administração de Empresas do setor sucroalcooleiro, pois permite realizar estimativas futuras mais precisas de preços do açúcar, que podem ser relevantes em processos de orçamento e planejamento no setor. Além disso, o modelo também pode ser utilizado para formação de preços no mercado de derivativos de *commodities*.

Como continuidade ao presente estudo, o modelo proposto pode ser implementado para outras *commodities* agrícolas que também são influenciadas pelas variáveis e premissas estudadas.

REFERÊNCIAS BIBLIOGRÁFICAS

AIUBE, Fernando A. L.; BAIDYA, Tara K. N.; TITO, Edson E. A. Processos estocásticos dos preços das commodities: uma abordagem através do filtro de partículas. **Revista Brasileira de Economia**. v. 60, n. 3, p. 215-233, Jul. 2006.

BENIRSCHKA, Martin; BINKLEY, James K. Optimal storage and marketing over space and time. **American Journal of Agricultural Economics**. v. 77, n. 3, p. 512-524, Aug. 1995.

BLACK, Fischer; SCHOLES, Myron S. The pricing of options and corporate liabilities. **Journal of Political Economy**. v. 81, n. 3, p. 637-654, May. 1973.

BM&FBOVESPA. *Bolsa de Valores, Mercadorias e Futuros*. Disponível em: <<http://www.bmf.com.br/>>. Acesso em: 10 jul. 2008.

BRENNAN, Michael J. The supply of storage. **The American Economic Review**. v. 48, n. 1, p. 50-72, Mar. 1958.

BRENNAN, Michael J.; SCHWARTZ, Eduardo S. Evaluating natural resource investments. **The Journal of Business**. v. 58, n. 2, p. 135-157, Apr. 1985.

BUSSAB, Wilton O.; MORETIN, Pedro A. **Estatística básica**. 5. ed. São Paulo: Saraiva, 2004.

CBOT. *Organizational profile*. Disponível em: <<http://www.cbot.com/cbot/pub/page/0,3181,760,00.html>>. Acesso em: 30 mar. 2008.

CEPEA. *Indicadores diários*. Disponível em: <<http://cepea.esalq.usp.br/indicador/>>. Acesso em: 10 jul. 2008.

COCHRANE, John H. **Asset pricing**. New Jersey: Princeton, 2005.

COOPER, Donald R.; SCHINDLER, Pamela S. **Métodos de pesquisa em administração**. Tradução: Luciana de Oliveira da Rocha. 7. ed. Porto Alegre: Bookman, 2003.

COPELAND, Thomas E.; ANTIKAROV, Vladimir. **Opções reais**: um novo paradigma para reinventar a avaliação de investimentos. Tradução: Maria José Cyhlar. Rio de Janeiro: Campus, 2001.

COPELAND, Thomas E.; KOLLER, Tim; MURRIN, Jack. **Avaliação de empresas - valuation**: calculando e gerenciando o valor das empresas. Tradução: Allan Vidigal Hastings. 3. ed. São Paulo: Makron Books, 2002.

COPELAND, Thomas E.; WESTON, J. Fred. **Financial theory and corporate policy**. 3rd. ed. New York: Addison Wesley, 1988.

DICKEY, David A.; FULLER, Wayne A. Distribution of the estimators for autoregressive time series with a unit root. **Journal of the American Statistical Association**. v. 74, n. 366, p. 427-431, June. 1979.

DUSAK, Katherine. Futures trading and investor returns: an investigation of commodity market risk premiums. **The Journal of Political Economy**. v. 81, n. 6, p. 1387-1406, Nov. 1973.

ENDERS, Walter. **Applied econometric time series**. 2nd. ed. New Jersey: Wiley, 2004.

ENGLE, Robert F.; GRANGER, Clive. W. J. Co-integration and error correction: representation, estimation, and testing. **Econometrica**. v. 55, n. 2, p. 251-276, Mar. 1987.

EPE. *Empresa de Pesquisa Energética*. Disponível em: <http://www.epe.gov.br/PressReleases/20080508_1.pdf>. Acesso em: 8 mai. 2008.

FAMA, Eugene F.; FRENCH, Kenneth R. Commodity futures prices: some evidence on forecast power, premiums, and the theory of storage. **The Journal of Business**. v. 60, n. 1, p. 55-73, Jan. 1987.

FEDERAL RESERVE. *Fed Funds*. Disponível em: <<http://www.federalreserve.gov/econresdata/default.htm>>. Acesso em: 10 jul. 2008.

FRECHETTE, Darren L.; FACKLER, Paul L. What causes commodity price backwardation? **American Journal of Agricultural Economics**. v. 81, n. 4, p. 761-771, Nov. 1999.

GALDI, Fernando C.; PEREIRA, Leonel M. Value at risk using volatility forecasting models: EWMA, GARCH and stochastic volatility. **Brazilian Business Review**. v. 4, n. 1, p. 74-94, Jan. 2007.

GEMAN, Hélyette. **Commodities and commodity derivatives**: modeling and pricing for agriculturals, metals and energy. Chichester: Wiley, 2005.

GIBSON, Rajna; SCHWARTZ, Eduardo S. Stochastic convenience yield and the pricing of oil contingent claims. **The Journal of Finance**. v. 45, n. 3, p. 959-976, July. 1990.

GRANGER, Clive. W. J. Investigating casual relations by econometric models and cross-spectral methods. **Econometrica**. v. 37, n. 3, p. 424-438, Aug. 1969.

GRANGER, Clive. W. J.; NEWBOLD, Paul. Forecasting transformed series. **Journal of the Royal Statistical Society**. v. 38, n. 2, p. 189-203, 1976.

- GREENE, William H. **Econometric analysis**. 5th. ed. New Jersey: Prentice Hall, 2003.
- GREWAL, Mohinder S.; ANDREWS, Angus P. **Kalman filtering: theory and practice using MATLAB**. 2nd. ed. New Jersey: Wiley, 2001.
- HAIR, Joseph F. et al. **Análise multivariada de dados**. Tradução: Adonai Schlup Sant'Anna, Anselmo Chaves Neto. 5. ed. Porto Alegre: Bookman, 2005.
- HAMILTON, James D. **Time series analysis**. New Jersey: Princeton, 1994.
- HANNAN, E. J.; TERRELL, R. D.; TUCKWELL, N. E. The seasonal adjustment of economic time series. **International Economic Review**. v. 11, n. 1, p. 24-52, Feb. 1970.
- HULL, John C. **Options, futures, and other derivatives**. 6th. ed. New Jersey: Prentice Hall, 2006.
- ICE. *Intercontinental exchange crude oil futures*. Disponível em: <<https://www.theice.com/crude.jhtml>>. Acesso em: 2 abr. 2008.
- ITÔ, Kiyosi. On stochastic differential equations. **Memoirs of the American Mathematical Society**. v. 4, n. 1, p. 1-51, 1951.
- JOHANSEN, Soren. **Likelihood-based inference in cointegrated vector autoregressive models**. Oxford: Oxford University Press, 1995.
- JOHNSON, Leland L. The theory of hedging and speculation in commodity futures. **The Review of Economic Studies**. v. 27, n. 3, p. 139-151, June. 1960.
- KALDOR, Nicholas. Speculation and economic stability. **The Review of Economic Studies**. v. 7, n. 1, p. 1-27, Oct. 1939.
- LIFFE. *White sugar futures*. Disponível em: <<http://www.euronext.com/trader/trader-21313-EN.html>>. Acesso em: 31 mar. 2008.
- LITZENBERGER, Robert H.; RABINOWITZ, Nir. Backwardation in oil futures markets: theory and empirical evidence. **The Journal of Finance**. v. 50, n. 5, p. 1517-1545, Dec. 1995.
- MACKINNON, James G. Numerical distribution functions for unit root and cointegration tests. **Journal of Applied Econometrics**. v. 11, n. 6, p. 601-618, Apr. 1996.
- MARCONI, Marina A.; LAKATOS, Eva M. **Metodologia científica**. 3. ed. São Paulo: Atlas, 2000.

MARQUES, Pedro V.; MELLO, Pedro C. **Mercados futuros de commodities agropecuárias: exemplos e aplicações para os mercados brasileiros**. São Paulo: Bolsa de Mercadorias & Futuros, 1999.

MARTINS, Gilberto A. **Manual para elaboração de monografias e dissertações**. 2. ed. São Paulo: Atlas, 1994.

MAYBACK, Peter S. **Stochastic models, estimation and control**. Academic Press: New York, 1979.

MERTON, Robert C. Theory of rational option pricing. **The Bell Journal of Economics and Management Science**. v. 4, n. 1, p. 141-183, *Spring*. 1973.

MORETTIN, Pedro A.; TOLOI, Célia M. C. **Análise de séries temporais**. 2. ed. São Paulo: Edgar Blücher, 2006.

NEWBY, Whitney K.; WEST, Kenneth D. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. **Econometrica**. v. 55, n. 3, p. 703-708, May. 1987.

NYMEX. *Number 11 sugar futures*. Disponível em: <http://www.nymex.com/YO_desc.aspx>. Acesso em: 30 mar. 2008.

OIL ANALYTICS. *Net Energy*. Disponível em: <<http://web.archive.org/web/20041019220802/http://www.oilanalytics.org/neten/neten.html#measure>>. Acesso em: 21 abr. 2008.

PINDYCK, Robert S.; RUBINFELD, Daniel L. **Microeconomics**. 3rd. ed. New Jersey: Prentice Hall, 1994.

POPPER, Karl S. **A lógica da pesquisa científica**. 2. ed. São Paulo: Cultrix, 1975.

SCHOUCHANA, Félix; MICELI, Wilson M. **Introdução aos mercados futuros e de opções agropecuários no Brasil**. 3. ed. São Paulo: Bolsa de Mercadorias & Futuros, 2004.

SCHWARTZ, Eduardo S. The stochastic behavior of commodity prices: implications for valuation and hedging. **The Journal of Finance**. v. 52, n. 3, p. 923-973, July 1997.

SCHWARTZ, Eduardo S.; SMITH, James E. Short-term variations and long-term dynamics in commodity prices. **Management Science**. v. 46, n. 7, p. 893-911, July 2000.

SECURATO, José R. **Decisões financeiras em condições de risco**. São Paulo: Atlas, 1996.

SORENSEN, Carsten. Modeling seasonality in agricultural commodity futures. **The Journal of Futures Markets**. v. 22, n. 5, p. 393-426, Mar. 2002.

TELSER, Lester G. Futures trading and the storage of cotton and wheat. **The Journal of Political Economy**. v. 66, n. 3, p. 233-255, June 1958.

UNICA. *União da Indústria de Cana-de-açúcar*. Disponível em: <<http://www.portalunica.com.br/portalunica/>>. Acesso em: 21 abr. 2008.

WALRAS, Léon. **Elements of pure economics**: or the theory of social wealth. Tradução: Willian Jaffe. London: Routledge, 2003.

WORKING, Holbrook. Theory of the inverse carrying charge in futures markets. **Journal of Farm Economics**. v. 30, n. 1, p. 1-28, Feb. 1948.

_____. The theory of price of storage. **The American Economic Review**. v. 39, n. 6, p. 1254-1262, Dec. 1949.

GLOSSÁRIO DE VARIÁVEIS

Abaixo estão representadas as variáveis utilizadas nas equações, elas estão em ordem de aparição no texto:

S_t : preço do ativo à vista no tempo t ;

F_t : preço do contrato futuro no tempo t , com vencimento em T , posterior a t ;

r : taxa de juros livre de risco;

W_t : custo marginal de armazenagem no tempo t ;

C_t : valor do ganho marginal de conveniência no tempo t ;

δ_t : retorno de conveniência no tempo t ;

B_t^w : *backwardation* fraco;

B_t^s : *backwardation* forte;

σ_{δ_t} : volatilidade do retorno de conveniência no tempo t ;

σ_{S_t} : volatilidade do retorno do preço à vista no tempo t ;

σ_{F_t} : volatilidade do retorno do preço futuro no tempo t ;

R_t : taxa de retorno contínua do preço à vista, sendo $R_t \sim N(\mu, \sigma^2)$;

z : variável normal padronizada;

y_t : em diversas equações representa a variável dependente;

x_t : em diversas equações representa variável independente ou explicativa;

ε_t : ruído branco (em algumas equações, resíduo das regressões);

ζ_t : ruído branco;

e_t : resíduo das regressões;

ξ_t : resíduo das regressões;

β_j : coeficientes das regressões;

α_j : coeficientes das regressões;

φ_1 : coeficiente da regressão;

D_t : desvio em relação à média do processo $AR(1)$;

ϕ : velocidade de reversão à média;

\mathbf{x}_t : matriz (n x m) de variáveis explicativas nos modelos de regressão;

$\boldsymbol{\beta}$: vetor (m x 1) de coeficientes da regressão;

$E(R_m)$: retorno do mercado esperado na equação do CAPM para *commodities*;

β_i : coeficiente de risco sistêmico da *commodity i* no modelo CAPM;

μ_s : componente determinístico da equação do preço à vista;

μ_δ : valor de equilíbrio médio estimado para o retorno de conveniência;

ϕ_δ : velocidade de reversão à média do retorno de conveniência;

μ_r : valor de equilíbrio médio estimado para as taxas de juros;

ϕ_r : velocidade de reversão à média para as taxas de juros;

χ_t e ω_t : fatores do modelo de curto e longo prazo;

ϕ_χ : velocidade de reversão à média do fator de curto prazo;

σ_χ : volatilidade do fator de curto prazo;

μ_ω : componente determinístico do fator de longo prazo;

σ_ω : volatilidade do fator de longo prazo;

λ_t : componente sazonal no tempo t ;

θ : tempo medido em anos no componente sazonal;

K : frequência dos ciclos sazonais anuais;

$\beta_{c,k}, \beta_{s,k}$: coeficientes lineares cosseno e seno para cada um dos K ciclos sazonais;

m : quantidade de *commodity* (a) no modelo de Walras (2003);

n : quantidade de *commodity* (b) no modelo de Walras (2003);

v_a : valor unitário da *commodity* (a) no modelo de Walras (2003);

v_b : valor unitário da *commodity* (b) no modelo de Walras (2003);

q : quociente da relação de troca entre as *commodities* (a) e (b);

p_b : preço da *commodity* (b) representado em termos da *commodity* (a);

p_a : preço da *commodity* (a) representado em termos da *commodity* (b);

k_e : ganho líquido de energia no modelo EROI;

E_o : total de energia obtida;

E_s : energia despendida no processo de exploração;

R_t^S : retorno do preço à vista da *commodity* no tempo t ;

R_{t-1}^P : retorno do petróleo com defasagem de uma unidade de tempo, em $t-1$;

$\hat{\sigma}_t^M$: volatilidade condicional dos mercados no tempo t ;

$\Delta \hat{\sigma}_t^M$: primeira diferença da volatilidade condicional do mercado no tempo t ;

\mathbf{x}_t : vetor $n \times 1$ de variáveis de estado nas equações de estado;

\mathbf{y}_t : vetor $m \times 1$ de variáveis de sinal;

\mathbf{A}, \mathbf{C} : vetores estimados no processo, incluindo os parâmetros relativos às variáveis exógenas;

\mathbf{B}, \mathbf{D} : matrizes estimadas no processo, referentes aos coeficientes lineares das variáveis de estado;

$\boldsymbol{\varepsilon}_t, \mathbf{v}_t$: vetores de ruídos brancos serialmente independentes;

R_t^E : retorno do preço à vista do etanol no tempo t ;

d : defasagem das variáveis;

\bar{B}^w, \bar{B}^s : médias da amostra de *backwardation* fraco e forte;

$S_{B^w}^2, S_{B^s}^2$: variâncias da amostra de *backwardation* fraco e forte;

v_t : resíduo de um processo ARMA;

$\hat{v}_{t|t-1}$: resíduo condicional;

ρ_i : autocorrelação dos resíduos quadráticos v_t^2 para defasagem i ;

σ_v^2 : variância dos resíduos de um processo ARMA;

h_t : variância condicional dos resíduos v_t ;

Ω_t : matriz de covariância dos vetores de ruído branco no filtro de Kalman;

$\hat{x}_{t|q}$: vetor de esperança condicional de x_t em t , dada a informação até o tempo q ;

$\hat{\sigma}_{x_t|q}^2$: matriz de variância condicional de x_t em t , dada a informação até o tempo q ;

S_o : preço à vista do açúcar observado, divulgado pelo CEPEA;

S_f : preço à vista do açúcar previsto pelo modelo;

N : número de observações nas séries de dados;

e_{1t} : erro de previsão do modelo 1 no tempo t ;

e_{2t} : erro de previsão do modelo 2 no tempo t ;

w_t, g_t : medidas de erro de previsão do teste Granger e Newbold (1976);

S_{1ft} : valor previsto do preço do açúcar no tempo t pelo modelo 1;

S_{2ft} : valor previsto do preço do açúcar no tempo t pelo modelo 2;

S_{ot} : valor observado do açúcar à vista no tempo t ;

$\rho_{w,g}$: correlação amostral entre as variáveis w e g .

APÊNDICE I – TESTE DE NORMALIDADE

O teste Jarque-Bera verifica se uma amostra é normalmente distribuída, uma das condições para que os estimadores MQO sejam não enviesados. O teste de normalidade é um teste conjunto de assimetria A_S e curtose K_U , para um número k de parâmetros do modelo, que possui uma distribuição χ^2 com dois graus de liberdade, definido pela seguinte equação:

$$Jarque\ Bera = \frac{N-k}{6} \left(A_S^2 + \frac{(K_U - 3)^2}{4} \right) \quad (1)$$

sendo:

N : número de observações da amostra.

A hipótese nula do teste de normalidade para os resíduos da regressão é a seguinte:

H_0 – os resíduos da regressão possuem distribuição normal.

O objetivo deste teste é a não rejeição da hipótese nula, ou seja, que as séries da amostra tenham distribuição de probabilidade Normal.

APÊNDICE II – TESTE DE CAUSALIDADE DE GRANGER

Um alto índice de correlação não implica que uma variável cause uma outra, essa relação pode ser simplesmente espúria. A abordagem de Granger (1969:424) para essa questão, se x causa y , na verdade é uma tentativa de verificar o quanto do valor contemporâneo de y pode ser explicado por valores passados de x , e verificar se a adição de valores defasados de x ao modelo pode melhorar a explicação de y . Diz-se que x Granger causa y se os coeficientes dos valores defasados de x são estatisticamente significantes. Pode acontecer causalidade nos dois sentidos, ou seja, x Granger causa y e y Granger causa x .

É importante verificar que o fato de x Granger causar y não significa uma relação de causa e efeito, ou seja, significa que x possui informações sobre as variações na variável y , mas não é possível afirmar, em absoluto, que y é uma consequência de x . Enders (2004:283) explica que em um modelo de duas equações com d defasagens, x não Granger causa y se todos os coeficientes defasados de x forem significativamente iguais a zero.

O teste de causalidade de Granger consiste em um teste de estatística F da hipótese conjunta de todos os coeficientes defasados da variável explicativa serem iguais a zero. Os coeficientes das seguintes equações com defasagem devem ser testados:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_d y_{t-d} + \beta_{x,1} x_{t-1} + \beta_{x,2} x_{t-2} + \dots + \beta_{x,d} x_{t-d} + \varepsilon_t \quad (a)$$

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_d x_{t-d} + \beta_{y,1} y_{t-1} + \beta_{y,2} y_{t-2} + \dots + \beta_{y,d} y_{t-d} + \varepsilon_t \quad (b)$$

(2)

As hipóteses nulas do teste de causalidade de Granger são as seguintes:

H_{0a} – x não Granger causa y se $\beta_{x,1} = \beta_{x,2} = \dots = \beta_{x,d} = 0$ na equação (a); e

H_{0b} – y não Granger causa x se $\beta_{y,1} = \beta_{y,2} = \dots = \beta_{y,d} = 0$ na equação (b).

O objetivo dos testes é rejeitar qualquer uma das hipóteses nulas acima com 5% de significância estatística.

Abaixo estão as saídas do *software* Eviews 5.0 para o teste de causalidade de Granger entre o etanol e o açúcar à vista para a amostra de dados semanais:

Pairwise Granger Causality Tests

Date: 03/04/09 Time: 15:02

Sample: 1 343

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
DLOG(ETANOL) does not Granger Cause DLOG(ACUCARBR)	340	0.01658	0.98356
DLOG(ACUCARBR) does not Granger Cause DLOG(ETANOL)		7.29157	0.00079

De acordo com o resultado do teste, rejeitou-se a hipótese nula que tem o açúcar à vista como variável explicativa, ou seja, o retorno do açúcar à vista Granger causa o retorno do etanol.

APÊNDICE III – MATRIZ DE COVARIÂNCIA ROBUSTA DE NEWEY-WEST

Quando existe correlação serial nos resíduos resultantes da regressão MQO, não é possível realizar os testes de hipótese dos coeficientes porque os erros-padrão resultantes da regressão estão errados. Greene (2003:267) sugere a correção da matriz de covariância de Newey e West (1987:704), nos casos de presença de heterocedasticidade e correlação serial nos resíduos. A correção do erro-padrão é obtida pela seguinte equação:

$$S_{e^*} = S_{e_0} + \frac{1}{N} \sum_{j=1}^d \sum_{t=j+1}^N \left(1 - \frac{j}{d+1}\right) e_t e_{t-j} [\mathbf{x}_t \mathbf{x}'_{t-j} + \mathbf{x}_{t-j} \mathbf{x}'_t] \quad (3)$$

sendo:

S_{e^*} : erro-padrão robusto;

S_{e_0} : erro-padrão MQO;

d : defasagem da autocorrelação;

N : número de resíduos;

e_t : resíduos no tempo t , correlacionado com e_{t-d} ;

\mathbf{x}_t : matriz de variáveis explicativas.

APÊNDICE IV – SAÍDAS DO MODELO GARCH

Os testes foram realizados no *software* EVIEWS 5.0 e as saídas do programa estão disponibilizadas nas tabelas abaixo:

Correlograma dos resíduos quadráticos para dados diários

Sample: 3 1631
Included observations: 1629
Q-statistic
probabilities
adjusted for 1
ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
*	*	1	0.132	0.132	28.613	
**	**	2	0.227	0.213	112.91	0.000
***	***	3	0.387	0.357	357.28	0.000
*	*	4	0.177	0.097	408.76	0.000
*		5	0.195	0.051	471.06	0.000
**		6	0.229	0.054	557.09	0.000
**	*	7	0.222	0.108	637.98	0.000
**	*	8	0.250	0.140	740.11	0.000
*		9	0.147	-0.011	775.71	0.000
**	*	10	0.249	0.081	877.14	0.000

Correlograma dos resíduos quadráticos para dados mensais

Sample: 2002:03 2008:06
Included observations: 76
Q-statistic
probabilities
adjusted for 1
ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
**	**	1	0.280	0.280	6.2023	
*	*	2	0.176	0.106	8.6959	0.003
**	*	3	0.217	0.157	12.511	0.002
.	*	4	0.032	-0.084	12.594	0.006
.	.	5	0.049	0.020	12.795	0.012
.	.	6	0.040	-0.002	12.933	0.024
*	*	7	0.106	0.117	13.906	0.031
.	*	8	0.002	-0.072	13.907	0.053
.	.	9	-0.009	-0.020	13.915	0.084
.	.	10	-0.004	-0.034	13.916	0.125

Modelo AR(1) / GARCH (1,1) para dados diários

Dependent Variable: DLOG(DJONES)

Method: ML - ARCH (Marquardt) – Normal distribution

Sample (adjusted): 1/04/2002 6/30/2008

Included observations: 1629 after adjustments

Convergence achieved after 13 iterations

Variance backcast: ON

GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*GARCH(-1)

	Coefficient	Std. Error	z-Statistic	Prob.
AR(1)	-0.065135	0.027436	-2.374024	0.0176
Variance Equation				
C	8.00E-07	1.77E-07	4.533236	0.0000
RESID(-1)^2	0.055444	0.008284	6.693146	0.0000
GARCH(-1)	0.935807	0.009307	100.5469	0.0000
R-squared	0.005046	Mean dependent var	6.73E-05	
Adjusted R-squared	0.003209	S.D. dependent var	0.010058	
S.E. of regression	0.010042	Akaike info criterion	-6.652859	
Sum squared resid	0.163868	Schwarz criterion	-6.639610	
Log likelihood	5422.754	Durbin-Watson stat	2.013699	
Inverted AR Roots	-.07			

Modelo AR(1) / GARCH (1,1) para dados mensais

Dependent Variable: DLOG(DJONES)

Method: ML - ARCH (Marquardt) – Normal distribution

Sample (adjusted): 2002M03 2008M06

Included observations: 76 after adjustments

Convergence achieved after 21 iterations

Variance backcast: ON

GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*GARCH(-1)

	Coefficient	Std. Error	z-Statistic	Prob.
AR(1)	0.010049	0.162664	0.061776	0.9507
Variance Equation				
C	7.54E-05	8.71E-05	0.865978	0.3865
RESID(-1)^2	0.160112	0.167341	0.956797	0.3387
GARCH(-1)	0.788876	0.165432	4.768573	0.0000
R-squared	-0.001197	Mean dependent var		0.001527
Adjusted R-squared	-0.042914	S.D. dependent var		0.037393
S.E. of regression	0.038187	Akaike info criterion		-3.802026
Sum squared resid	0.104994	Schwarz criterion		-3.679356
Log likelihood	148.4770	Durbin-Watson stat		1.851476
Inverted AR Roots	.01			

APÊNDICE V – SAÍDAS DO TESTE DE RAIZ UNITÁRIA

Os testes foram realizados no *software* EVIEWS 5.0, as saídas do programa estão disponibilizadas nas tabelas abaixo.

- Resultados para a variável **acucarbr (açúcar a vista)**

Dados diários

Null Hypothesis: ACUCARBR has a unit root

Exogenous: Constant

Lag Length: 5 (Automatic based on SIC, MAXLAG=24)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.354703	0.6057
Test critical values: 1% level	-3.434169	
5% level	-2.863114	
10% level	-2.567656	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(ACUCARBR) has a unit root

Exogenous: Constant

Lag Length: 4 (Automatic based on SIC, MAXLAG=24)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-11.89335	0.0000
Test critical values: 1% level	-3.434169	
5% level	-2.863114	
10% level	-2.567656	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(ACUCARBR,2) has a unit root

Exogenous: Constant

Lag Length: 6 (Automatic based on SIC, MAXLAG=24)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-24.20926	0.0000
Test critical values: 1% level	-3.434177	
5% level	-2.863117	
10% level	-2.567657	

*MacKinnon (1996) one-sided p-values.

Dados mensais

Null Hypothesis: ACUCARBR has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic based on SIC, MAXLAG=11)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.329365	0.6121
Test critical values: 1% level	-3.517847	
5% level	-2.899619	
10% level	-2.587134	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(ACUCARBR) has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic based on SIC, MAXLAG=11)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-8.203659	0.0000
Test critical values: 1% level	-3.519050	
5% level	-2.900137	
10% level	-2.587409	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(ACUCARBR,2) has a unit root

Exogenous: Constant

Lag Length: 2 (Automatic based on SIC, MAXLAG=11)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-9.359196	0.0000
Test critical values: 1% level	-3.522887	
5% level	-2.901779	
10% level	-2.588280	

*MacKinnon (1996) one-sided p-values.

- Resultados para a variável **petr (petróleo)**:

Dados diários

Null Hypothesis: PETR has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic based on SIC, MAXLAG=24)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	1.721411	0.9997
Test critical values: 1% level	-3.434157	
5% level	-2.863108	
10% level	-2.567653	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(PETR) has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic based on SIC, MAXLAG=24)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-41.95352	0.0000
Test critical values: 1% level	-3.434160	
5% level	-2.863109	
10% level	-2.567653	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(PETR,2) has a unit root

Exogenous: Constant

Lag Length: 12 (Automatic based on SIC, MAXLAG=24)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-20.74349	0.0000
Test critical values: 1% level	-3.434192	
5% level	-2.863124	
10% level	-2.567661	

*MacKinnon (1996) one-sided p-values.

Dados mensais

Null Hypothesis: PETR has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic based on SIC, MAXLAG=11)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	2.291635	1.0000
Test critical values: 1% level	-3.517847	
5% level	-2.899619	
10% level	-2.587134	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(PETR) has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic based on SIC, MAXLAG=11)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-7.715781	0.0000
Test critical values: 1% level	-3.519050	
5% level	-2.900137	
10% level	-2.587409	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(PETR,2) has a unit root

Exogenous: Constant

Lag Length: 2 (Automatic based on SIC, MAXLAG=11)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-9.044390	0.0000
Test critical values: 1% level	-3.522887	
5% level	-2.901779	
10% level	-2.588280	

*MacKinnon (1996) one-sided p-values.

- Resultados para a variável **vol (volatilidade condicional GARCH)**:

Dados diários

Null Hypothesis: VOL has a unit root

Exogenous: Constant

Lag Length: 3 (Automatic based on SIC, MAXLAG=24)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.387805	0.1454
Test critical values: 1% level	-3.434169	
5% level	-2.863114	
10% level	-2.567656	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(VOL) has a unit root

Exogenous: Constant

Lag Length: 2 (Automatic based on SIC, MAXLAG=24)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-21.13570	0.0000
Test critical values: 1% level	-3.434169	
5% level	-2.863114	
10% level	-2.567656	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(VOL,2) has a unit root

Exogenous: Constant

Lag Length: 12 (Automatic based on SIC, MAXLAG=24)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-19.22630	0.0000
Test critical values: 1% level	-3.434197	
5% level	-2.863126	
10% level	-2.567662	

*MacKinnon (1996) one-sided p-values.

Dados mensais

Null Hypothesis: VOL has a unit root
 Exogenous: Constant
 Lag Length: 0 (Automatic based on SIC, MAXLAG=11)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.713640	0.4204
Test critical values: 1% level	-3.519050	
5% level	-2.900137	
10% level	-2.587409	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(VOL) has a unit root
 Exogenous: Constant
 Lag Length: 0 (Automatic based on SIC, MAXLAG=11)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-8.573565	0.0000
Test critical values: 1% level	-3.520307	
5% level	-2.900670	
10% level	-2.587691	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(VOL,2) has a unit root
 Exogenous: Constant
 Lag Length: 5 (Automatic based on SIC, MAXLAG=11)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-5.694785	0.0000
Test critical values: 1% level	-3.528515	
5% level	-2.904198	
10% level	-2.589562	

*MacKinnon (1996) one-sided p-values.

- Resultados para a variável **etanol**:

Dados semanais

Null Hypothesis: ETANOL has a unit root
 Exogenous: Constant
 Lag Length: 1 (Automatic based on SIC, MAXLAG=16)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.435340	0.5652
Test critical values: 1% level	-3.449332	
5% level	-2.869800	
10% level	-2.571239	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(ETANOL) has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic based on SIC, MAXLAG=16)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-11.25376	0.0000
Test critical values: 1% level	-3.449332	
5% level	-2.869800	
10% level	-2.571239	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(ETANOL,2) has a unit root

Exogenous: Constant

Lag Length: 1 (Automatic based on SIC, MAXLAG=16)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-18.26482	0.0000
Test critical values: 1% level	-3.449447	
5% level	-2.869850	
10% level	-2.571266	

*MacKinnon (1996) one-sided p-values.

APÊNDICE VI – SAÍDAS DO TESTE DE COINTEGRAÇÃO

As tabelas abaixo apresentam os resultados obtidos no Eviews 5.0:

Dados diários

Sample(adjusted): 8 1631
 Included observations: 1624 after adjusting endpoints
 Trend assumption: Linear deterministic trend
 Series: ACUCARBR PETR VOL
 Lags interval (in first differences): 1 to 4

Unrestricted Cointegration Rank Test

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	5 Percent Critical Value	1 Percent Critical Value
None	0.007701	20.42152	29.68	35.65
At most 1	0.004110	7.866795	15.41	20.04
At most 2	0.000726	1.179125	3.76	6.65

*(**) denotes rejection of the hypothesis at the 5%(1%) level
 Trace test indicates no cointegration at both 5% and 1% levels

Dados mensais

Sample(adjusted): 2002:05 2008:06
 Included observations: 74 after adjusting endpoints
 Trend assumption: Linear deterministic trend
 Series: ACUCARBR PETR VOL
 Lags interval (in first differences): 1 to 2

Unrestricted Cointegration Rank Test

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	5 Percent Critical Value	1 Percent Critical Value
None	0.166703	20.73162	29.68	35.65
At most 1	0.073005	7.236620	15.41	20.04
At most 2	0.021745	1.626919	3.76	6.65

*(**) denotes rejection of the hypothesis at the 5%(1%) level
 Trace test indicates no cointegration at both 5% and 1% levels

APÊNDICE VII – SAÍDAS DA REGRESSÃO MULTIVARIADA

Saídas do *software* Eviews 5.0, regressões lineares múltiplas MQO:

Dados diários

Dependent Variable: DLOG(ACUCARBR)

Method: Least Squares

Sample(adjusted): 4 1631

Included observations: 1628 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLOG(PETR)	0.027418	0.015050	1.821765	0.0687
D(VOL)	0.550668	0.736953	0.747222	0.4550
COSENO	0.001460	0.000446	3.275772	0.0011
SENO	-0.001100	0.000448	-2.457166	0.0141
R-squared	0.011820	Mean dependent var		0.000233
Adjusted R-squared	0.009994	S.D. dependent var		0.012800
S.E. of regression	0.012736	Akaike info criterion		-5.886304
Sum squared resid	0.263425	Schwarz criterion		-5.873048
Log likelihood	4795.451	Durbin-Watson stat		1.309919

Dados mensais

Dependent Variable: DLOG(ACUCARBR)

Method: Least Squares

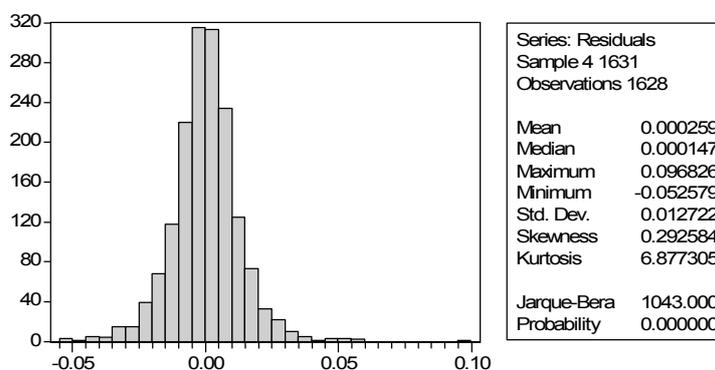
Sample(adjusted): 2002:03 2008:06

Included observations: 76 after adjusting endpoints

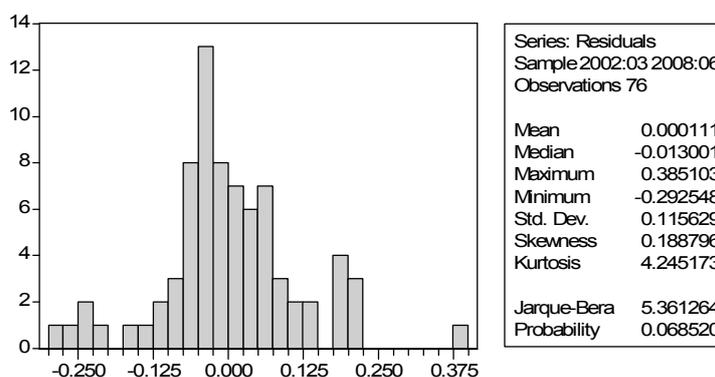
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLOG(PETR(-1))	0.397344	0.160881	2.469801	0.0159
D(VOL)	5.507513	2.999245	1.836300	0.0704
COSENO	0.041362	0.019346	2.137977	0.0359
SENO	-0.008295	0.019389	-0.427822	0.6701
R-squared	0.152484	Mean dependent var		0.007660
Adjusted R-squared	0.117171	S.D. dependent var		0.125601
S.E. of regression	0.118013	Akaike info criterion		-1.384842
Sum squared resid	1.002754	Schwarz criterion		-1.262172
Log likelihood	56.62401	Durbin-Watson stat		2.183614

Testes dos resíduos da regressão – teste de normalidade Jarque-Bera:

Dados diários



Dados mensais



Testes dos resíduos da regressão – correlação serial:

Dados diários

Sample: 4 1631

Included observations: 1628

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
***	***	1	0.344	0.344	193.24	0.000
**	*	2	0.209	0.102	264.30	0.000
**	*	3	0.216	0.134	340.68	0.000
**	*	4	0.223	0.118	422.14	0.000
**	*	5	0.204	0.082	490.48	0.000
*		6	0.134	0.000	519.79	0.000
*		7	0.139	0.042	551.63	0.000
*		8	0.150	0.045	588.28	0.000
*		9	0.136	0.029	618.38	0.000
*		10	0.108	0.008	637.63	0.000

Dados mensais

Sample: 2002:03 2008:06

Included observations: 76

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. * .	. * .	1	-0.093	-0.093	0.6779	0.410
** .	** .	2	-0.225	-0.235	4.7270	0.094
. *	. *	3	0.157	0.117	6.7317	0.081
. **	. **	4	0.257	0.250	12.151	0.016
. * .	. * .	5	-0.167	-0.065	14.473	0.013
. * .	. .	6	-0.101	-0.056	15.329	0.018
. .	. * .	7	0.024	-0.120	15.377	0.031
. *	. *	8	0.119	0.067	16.615	0.034
. * .	. .	9	-0.122	-0.042	17.934	0.036
** .	** .	10	-0.282	-0.273	25.098	0.005

APÊNDICE VIII – SAÍDAS DO FILTRO DE KALMAN

As saídas do filtro de Kalman do *software* Eviews 5.0.

Dados diários

Sspace: MODELO PROPOSTO
 Method: Maximum likelihood (Marquardt)
 Sample: 4 1631
 Included observations: 1628
 User prior mean: SVEC
 User prior variance: SVAR
 Convergence achieved after 1 iteration

	Coefficient
C(1)	0.008398
C(2)	0.116495
C(3)	-6.556159
C(4)	0.000311
C(5)	0.022868
C(6)	1.069168
C(7)	0.001367
C(8)	-0.001836
C(9)	-8.686838
	Final State
SV1	0.031684
SV2	2.662804
Log likelihood	4551.945
Parameters	9
Diffuse priors	2

Sspace: DOISFATORES
 Method: Maximum likelihood (Marquardt)
 Sample: 1 1631
 Included observations: 1631
 User prior mean: SVEC
 User prior variance: SVAR
 Convergence achieved after 1 iteration

	Coefficient
C(1)	0.008339
C(2)	0.111897
C(3)	-6.588269
C(4)	0.000254
C(9)	-8.661802
	Final State
SV1	0.041532
SV2	2.666071
Log likelihood	4548.117
Parameters	5
Diffuse priors	2

Dados mensais

Sspace: MODELO PROPOSTO
 Method: Maximum likelihood (Marquardt)
 Sample: 2002:03 2008:06
 Included observations: 76
 User prior mean: SVEC
 User prior variance: SVAR
 Convergence achieved after 2 iterations

	Coefficient
C(1)	1.112879
C(2)	0.143086
C(3)	-39.29111
C(4)	0.011527
C(5)	0.049383
C(6)	1.187362
C(7)	0.002386
C(8)	-0.035554
C(9)	-4.934495
	Final State
SV1	0.143086
SV2	2.697478
Log likelihood	64.31299
Parameters	9
Diffuse priors	2

Sspace: DOISFATORES
 Method: Maximum likelihood (Marquardt)
 Sample: 2002:01 2008:06
 Included observations: 78
 User prior mean: SVEC
 User prior variance: SVAR
 Convergence achieved after 1 iteration

	Coefficient
C(1)	0.124878
C(2)	0.083884
C(3)	-39.29111
C(4)	0.006749
C(9)	-4.793600
	Final State
SV1	0.083924
SV2	2.680925
Log likelihood	60.32200
Parameters	5
Diffuse priors	2