

Juliana dos Santos Lopes

Lucas de Barros Pimenta

PREVISÃO DA SÉRIE DE PREÇOS GLOBAIS DE METANOL ATRAVÉS DOS MODELOS
BOX & JENKINS E REGRESSÃO DINÂMICA

MONOGRAFIA SUBMETIDA À COORDENAÇÃO DE CURSO DE PÓS GRADUAÇÃO EM
MÉTODOS ESTATÍSTICOS COMPUTACIONAIS
DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA
COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A
CONCLUSÃO DO CURSO

Aprovada por:

Prof. Reinaldo Castro Souza, Ph.D

Fernando Luiz Cyrino Oliveira, M.Sc.

JUIZ DE FORA, MG – BRASIL

AGOSTO DE 2011

DEDICATÓRIA

Dedicamos este trabalho às nossas famílias, que sempre nos apoiaram e incentivaram.

AGRADECIMENTO

Agradecemos ao amigo Fernando Cyrino que muito nos auxiliou neste trabalho.

Resumo da Monografia apresentada à Coordenação de Curso de Métodos Estatísticos Computacionais como parte dos requisitos necessários para conclusão do curso

PREVISÃO DA SÉRIE DE PREÇOS GLOBAIS DE METANOL ATRAVÉS DOS MODELOS
BOX & JENKINS E REGRESSÃO DINÂMICA

Juliana dos Santos Lopes

Lucas de Barros Pimenta

Agosto/2011

Orientador: Reinaldo Castro Souza

Co-orientador: Fernando Luiz Cyrino Oliveira

Curso: Pós Graduação em Métodos Estatísticos Computacionais

A previsão auxilia na tomada de decisões dos agentes envolvidos em atividades que necessitam de planejamento e redução de incertezas. A redução de incertezas é especialmente importante no mercado de *commodities*, como é o caso do metanol, pois sendo este um mercado global, está muito sujeito às variações de mercado. Este trabalho tem como proposta determinar um modelo consistente para a previsão de preços de metanol, através da utilização dos modelos Box & Jenkins e Regressão Dinâmica. Estes dois modelos são então comparados através de seus indicadores e o que apresenta melhores resultados é utilizado para a previsão de 12 meses do preço do metanol. Esta previsão é, então, comparada com os valores reais ocorridos no ano de 2010.

Palavras-chaves: previsão, regressão dinâmica, Box & Jenkins, metanol.

Abstract of monograph presented to Statistics Department of UFJF as part of the requirements for graduation

FORECAST OF THE GLOBAL METHANOL PRICES SERIES THROUGH BOX & JENKINS
AND DYNAMIC REGRESSION MODELS

Juliana dos Santos Lopes

Lucas de Barros Pimenta

Agosto/2011

Advisor: Reinaldo Castro Souza

Co-advisor: Fernando Luiz Cyrino Oliveira

Department: Statistics

The forecast helps in making decisions of the agents involved in activities that require planning and reducing uncertainty. The reduction of uncertainty is specially important in the commodities market, such as methanol, because this is a global market and it is very subject to market variations. This monograph aims to determine a consistent model for forecasting prices of methanol, using the Box & Jenkins and dynamic regression models. These two models are then compared through their indicators and the model with best results is used for forecasting 12 months of methanol prices. This forecast is then compared with the real values that occurred in 2010.

Word-keys: forecasting, dynamic regression, Box & jenkins, methanol.

SUMÁRIO

CAPA.....	i
DEDICATÓRIA	ii
AGRADECIMENTO	iii
RESUMO	iiiv
ABSTRACT.....	v
ÍNDICE DE FIGURAS.....	viii
Capítulo I	1
INTRODUÇÃO.....	1
1. CONSIDERAÇÕES INICIAIS	1
2. OBJETIVOS	1
3. JUSTIFICATIVAS.....	2
4. ESCOPO DO TRABALHO	2
5. METODOLOGIA.....	2
Capítulo II	3
REVISÃO BIBLIOGRÁFICA.....	3
1. SÉRIES TEMPORAIS	3
2. MODELAGEM DE SÉRIES TEMPORAIS	3
3. MODELOS ARIMA	5
3.1. MODELO AUTO-REGRESSIVO E DE MÉDIA MÓVEL (ARMA)	6
3.2. MODELO AUTO-REGRESSIVO INTEGRADO E DE MÉDIA MÓVEL (ARIMA) ..	6
3.3. ARIMA Sazonal (SARIMA)	6
3.4. MODELAGEM ARIMA	7
4. MODELO DE REGRESSÃO DINÂMICA	8
Capítulo III	10
DESCRIÇÃO	10
1. DESCRIÇÃO DO PROCESSO PRODUTIVO.....	10
2. DESCRIÇÃO DO PROBLEMA	11
Capítulo IV.....	12
DESENVOLVIMENTO	12
1. ANÁLISE DA SÉRIE	12
1.1. ANÁLISE DA SÉRIE SEM TRANSFORMAÇÃO.....	12
1.2. ANÁLISE DA SÉRIE TRANSFORMADA PARA LN	14
2. MODELAGEM ARIMA.....	15
2.1. ARIMA (1,1,0).....	17

2.2.	ARIMA (2,1,0).....	18
2.3.	ARIMA (1,1,1).....	19
2.4.	ARIMA (1,1,2).....	20
2.5.	ARIMA (0,1,2).....	21
3.	REGRESSÃO DINÂMICA	22
3.1.	VARIÁVEL: CONSTANTE	23
3.2.	VARIÁVEIS: CONSTANTE E ÓLEO COMBUSTÍVEL	24
3.3.	VARIÁVEIS: CONSTANTE, ÓLEO COMBUSTÍVEL E GÁS NATURAL.....	24
3.4.	VARIÁVEIS: CONSTANTE, GÁS NATURAL E METANOL [-1].....	26
3.5.	VARIÁVEIS: CONSTANTE, GÁS NATURAL, METANOL [-1] E PETRÓLEO	27
3.6.	VARIÁVEIS: CONSTANTE, GÁS NATURAL, METANOL [-1], PETRÓLEO E ÓLEO COMBUSTÍVEL.....	29
4.	COMPARAÇÃO ENTRE ARIMA E REGRESSÃO DINÂMICA	30
5.	PREVISÃO.....	31
	Capítulo V.....	33
	CONCLUSÃO	33
	REFERÊNCIAS BIBLIOGRÁFICAS.....	34

ÍNDICE DE FIGURAS

Figura 1 – Modelo Univariado	4
Figura 2 – Modelo Causal.....	4
Figura 3 – Modelo Multivariado	4
Figura 4 – Modelo Box & Jenkins.....	5
Figura 5 – Comportamento das funções FAC e FACP para modelos estacionários	7
Figura 6 – Diagrama de construção do modelo de regressão dinâmica	9
Figura 7 – Processo produtivo de metanol base gás natural	11
Figura 8 – Série de preços de metanol	12
Figura 9 – Histograma dos preços de metanol sem transformação.....	13
Figura 10 – Teste de normalidade da série sem transformação.....	13
Figura 11 – Teste de homocedasticidade da série sem transformação.....	13
Figura 12 – Histograma dos preços de metanol transformada para LN.....	14
Figura 13 – Teste de normalidade da série transformada para LN.....	14
Figura 14 – Teste de homocedasticidade da série transformada para LN.....	15
Figura 15 – FAC da série.....	15
Figura 16 – FACP da série.....	16
Figura 17 – FAC da série com uma diferenciação na parte de tendência	16
Figura 18 – FACP da série com uma diferenciação na parte de tendência	17
Figura 19 – Resultado do modelo ARIMA (1,1,0).....	17
Figura 20 - FAC dos resíduos	18
Figura 21 – Resultado do modelo ARIMA (2,1,0).....	19
Figura 22 – FAC dos resíduos	19
Figura 23 – Resultado do modelo ARIMA (1,1,1).....	20
Figura 24 – FAC dos resíduos	20
Figura 25 – Resultado do modelo ARIMA (1,1,2).....	21
Figura 26 – FAC dos resíduos	21
Figura 27 – Resultado do modelo ARIMA (0,1,2).....	22
Figura 28 – FAC dos resíduos	22
Figura 29 – Resultado do modelo utilizando somente a constante.....	23
Figura 30 – Resultado do modelo utilizando a constante e óleo combustível.....	24
Figura 31 – Resultado do modelo utilizando a constante, óleo combustível e gás natural ..	26
Figura 32 – Resultado do modelo utilizando a constante, gás natural e metanol [-1]	27
Figura 33 – Resultado do modelo utilizando a constante, gás natural, metanol [-1] e petróleo	28

Figura 34 – Resultado do modelo utilizando a constante, gás natural, metanol [-1], petróleo e óleo combustível.....	30
Figura 35 – Comparação entre ARIMA e Regressão Dinâmica	30
Figura 36 – Resultado do ARIMA (1,1,0) sem hold out	31
Figura 37 – Previsão de 12 meses.....	31
Figura 38 – Gráfico da série de preços de metanol com previsão.....	32

Capítulo I INTRODUÇÃO

1. CONSIDERAÇÕES INICIAIS

O preço do metanol é altamente impactado pela sua diversidade de aplicações, pelo nível de produção industrial, pelo valor da energia e pelas mudanças nas condições econômicas. O metanol é uma *commodity* global, sendo um produto com pequeno grau de industrialização, de qualidade uniforme, com cotação e negociação globais. É um líquido incolor, de fórmula molecular CH_3OH . Este produto é obtido principalmente a partir do gás natural. Aproximadamente 85% do metanol produzido no mundo têm origem no gás natural. Apesar disto, o gás de síntese formado pela gaseificação do carvão está cada vez mais sendo utilizado para a geração de metanol, particularmente na China, onde estão proibidos novos projetos de metanol baseados em gás natural.

O metanol é uma das matérias-primas mais consumidas na indústria química. Sua demanda é dividida em duas classes de produtos: produtos químicos (principalmente formaldeído e ácido acético) e produtos energéticos (principalmente biodiesel, éter metil terc-butílico (MTBE) e dimetil éter (DME)).

A indústria de metanol é muito concentrada, sendo que os 10 maiores produtores de metanol participam com cerca de 43% da capacidade instalada mundial e 89% do volume total comercializado (MAXIQUIM, 2008). A Methanex é a maior produtora, com 26% do volume comercializado, tendo uma relevante participação na formação do preço do metanol mundialmente. Desta forma, observa-se que um dos fatores essenciais de competitividade das empresas produtoras de metanol é a escala produtiva.

Esse trabalho visa testar dois modelos de previsão diferentes para prever de preços de metanol no mercado mundial: os univariados, que utilizam somente os dados passados da série temporal para projetar os valores futuros, considerando que a própria série se auto-explica, e os causais, que utilizam variáveis exógenas para explicar a série, ou seja, a série temporal é explicada pelos seus valores passados e pelos valores passados de outras variáveis. Em seguida deve-se fazer a previsão para doze meses e compará-la aos dados reais.

2. OBJETIVOS

Este trabalho visa à análise da série histórica dos preços de metanol com holdout de 12 meses, utilizando dos modelos Box & Jenkins e Regressão Dinâmica, para verificar aquele que apresenta a melhor modelagem. Posteriormente será feita a previsão dos 12 meses seguintes para os preços de metanol e compará-la aos dados reais.

3. JUSTIFICATIVAS

A previsão sempre foi um dos objetivos das análises quantitativas, pois auxilia na tomada de decisões dos agentes envolvidos em atividades que necessitam de planejamento e redução de incertezas. A redução de incertezas é especialmente importante no mercado de *commodities*, visto que, sendo um mercado global, está muito sujeito às variações de mercado.

A estatística é uma ferramenta essencial para a previsão de preços, pois permite avaliar uma quantidade muito grande de informações, levando em consideração as inseguranças que existem no cotidiano das empresas.

As técnicas de previsão auxiliam a tomada de decisões em atividades que necessitam de planejamento e redução da incerteza, tornando assim os possíveis e futuros riscos mais visíveis e conseqüentemente mais gerenciáveis (MAKRIDAKIS et al., 1983).

Desta forma, fazendo a previsão de preços para 12 meses, este trabalho auxilia as empresas, com a utilização desses valores, a preverem seus fluxos de caixa e o lucro que terão em um ano. Além disso, com esta previsão, pode-se analisar se investimentos no setor são viáveis.

4. ESCOPO DO TRABALHO

Neste estudo é realizada a previsão de preços do metanol com base nos dados de preços médios mensais deste produto nos contratos da América do Norte no período de janeiro de 2000 a dezembro de 2009.

5. METODOLOGIA

Este trabalho foi realizado em seis etapas:

- Na primeira etapa foi realizado o estudo do mercado de metanol, as variáveis que o influenciam diretamente e foi decidido qual seria o escopo do estudo;
- A segunda etapa consistiu na coleta de dados da série temporal de preços e nas séries de preços de outros produtos que seriam analisados como variáveis exógenas;
- Na terceira etapa ocorreu a escolha dos modelos de previsão que seriam utilizados e ajuste dos mesmos à série;
- A quarta etapa foi de análise dos modelos obtidos e comparação dos dois para verificação de qual modelo seria o mais adequado para a previsão dos preços;;
- A quinta etapa foi previsão para os 12 meses seguintes com o melhor modelo e comparação das previsões obtidas com os dados reais da série;
- A sexta etapa foi de fechamento do trabalho, com as conclusões obtidas.

Capítulo II

REVISÃO BIBLIOGRÁFICA

1. SÉRIES TEMPORAIS

Para Milone e Angelini (1995), chama-se série temporal a todo conjunto de valores passíveis de ordenação cronológica. Uma característica importante deste tipo de dados é que as observações apresentam dependência serial e, por isso, a ordem das observações é crucial para a análise e modelagem da série.

Uma série temporal discreta é aquela em que as observações são feitas em tempos específicos, por exemplo, as vendas mensais de um determinado produto. Já em uma série temporal contínua, as observações são feitas continuamente no tempo, como o registro de marés no porto de uma cidade.

Segundo Morettin e Toloí (2006), obtida a série temporal, os objetivos da análise podem ser:

- Investigar o mecanismo gerador da série temporal;
- Fazer previsões de valores futuros da série;
- Descrever apenas o comportamento da série;
- Procurar periodicidades relevantes nos dados.

Uma das suposições básicas feitas na análise de séries temporais é que o processo estocástico gerador dos dados seja um processo estacionário. De modo bastante geral, um processo diz-se estacionário se ele oscila ao redor de uma média constante, com uma variância também constante (MORETTIN, 2008). Porém, na prática, a maioria das séries existentes tem alguma forma de não-estacionariedade, que normalmente são tendência e/ou sazonalidade.

2. MODELAGEM DE SÉRIES TEMPORAIS

Um modelo de séries temporais é uma equação ou conjunto de equações que representa a estrutura da série. Esses modelos podem ser classificados em duas classes, de acordo com o número de parâmetros envolvidos, em: modelos paramétricos, para os quais o número de parâmetros é finito, e modelos não paramétricos, em que o número de parâmetros é infinito.

Os métodos quantitativos de modelagem podem ser:

- Univariados: a previsão dos valores futuros (Z_{t+1}) é explicada unicamente pelos valores passados da série (Z_t).

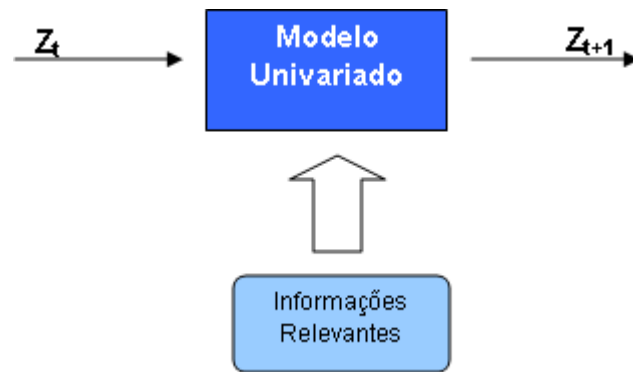


Figura 1 – Modelo Univariado

- Modelos Causais: a previsão dos valores futuros (Z_{t+1}) é explicada por diversas séries que influenciam nesses valores (Z_t, X_t, Y_t).

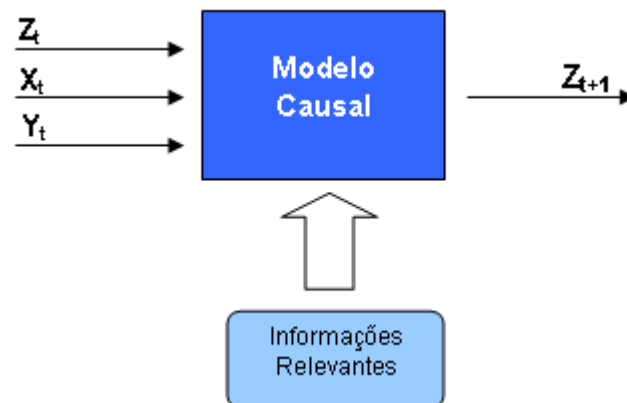


Figura 2 – Modelo Causal

- Modelos Multivariados: o vetor de séries retorna na previsão de outro vetor de séries.

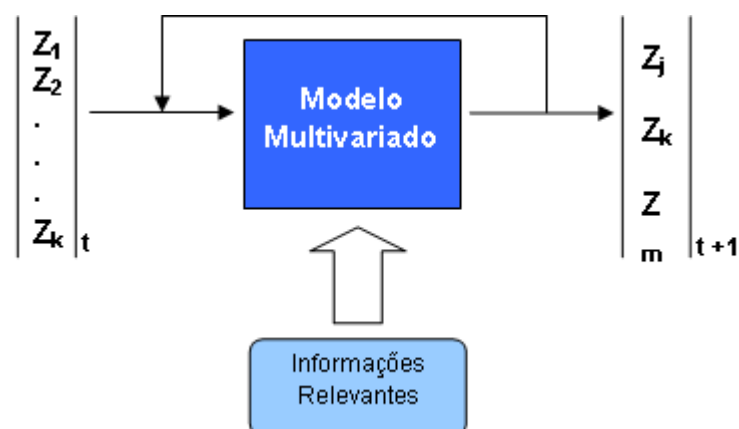


Figura 3 – Modelo Multivariado

3. MODELOS ARIMA

O modelo ARIMA (auto-regressivo integrado de média móvel), também conhecido como Box & Jenkins, é uma metodologia muito usada na análise de modelos paramétricos. Consiste em ajustar modelos auto-regressivos integrados de médias móveis a um conjunto de dados.

De acordo com Morettin e Toloí (2006), a estratégia para a construção do modelo será baseada em um ciclo iterativo, no qual a escolha da estrutura do modelo é baseada nos próprios dados. Os estágios do ciclo iterativo são:

- Uma classe geral de modelos é considerada para a análise;
- Há a identificação de um modelo, com base na análise de autocorrelações, autocorrelações parciais e outros critérios;
- A seguir vem a fase de estimação, na qual os parâmetros do modelo identificado são estimados;
- Finalmente, há a verificação ou diagnóstico do modelo ajustado, através da análise de resíduos, para saber se este é adequado para os fins em vista.

Este ciclo deve ser repetido até a adequação do modelo. Caso o propósito seja a previsão, devem-se comparar os modelos obtidos e escolher o mais ajustado, ou seja, o que tem menor erro de previsão. O objetivo é obter um modelo parcimonioso, isso é, um modelo com pequeno número de parâmetros e com previsões bastante precisas.

O operador retardo (B) é um operador bastante utilizado no modelo ARIMA, que é um operador translação para o passado: $B^m Z_t = Z_{t-m}$.

A aplicação da modelagem de Box & Jenkins leva em consideração alguns pressupostos: as séries devem ser estacionárias de segunda ordem e os resíduos na função de autocorrelação devem ser ruído branco. O modelo Box & Jenkins considera que a passagem de um ruído branco por um filtro linear de memória infinita gera um processo estacionário de segunda ordem (Z_t).



Figura 4 – Modelo Box & Jenkins

Um processo estocástico é estacionário de segunda ordem se apresenta média e variância constantes e a autocovariância só depende do *lag* m .

Um processo estocástico é considerado ruído branco, se além de estacionário de segunda ordem, não apresenta qualquer dependência serial.

3.1. MODELO AUTO-REGRESSIVO E DE MÉDIA MÓVEL (ARMA)

O modelo ARMA (p,q) ajusta-se a séries temporais estacionárias na média e na variância. O valor de “p” indica a ordem da parte auto-regressiva e o valor de “q” indica a ordem para a parte média móvel. Este modelo é definido pela equação:

$$\Phi(B)Z_t = \theta(B)a_t \quad (1)$$

$$\underbrace{(1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p)}_{AR} Z_t = \underbrace{(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)}_{MA} a_t \quad (2)$$

O modelo ARMA (p,q) deve ter seus erros (a_t) distribuídos aleatoriamente, como um ruído branco. A modelagem Box & Jenkins consiste em encontrar a estrutura ARMA (p,q) mais adequada para Z_t .

3.2. MODELO AUTO-REGRESSIVO INTEGRADO E DE MÉDIA MÓVEL (ARIMA)

Na prática, em sua maioria, as séries apresentam-se não estacionárias na média, ou seja, apresentam uma tendência. Desta forma, a modelagem ARMA (p,q) pode ser estendida para atender processos não-estacionários homogêneos. Um processo é não-estacionário homogêneo (Y_t) se ele se torna estacionário após a aplicação de “d” diferenças. Então, para a remoção da tendência da série é necessária a diferenciação da mesma, aplicando sobre ela o operador $(1 - B)$, conforme equações abaixo:

$$Z_t = \nabla^d Y_t \quad (3)$$

$$\nabla^d = (1 - B)^d \quad (4)$$

Com isso, o modelo ARIMA (p,d,q), onde d indica o número de diferenciações de para tornar o modelo estacionário de segunda ordem, fica como na fórmula:

$$(1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p) (1 - B)^d Z_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t \quad (5)$$

3.3. ARIMA Sazonal (SARIMA)

Há casos em que ocorrem simultaneamente tendência e sazonalidade na série. Quando isto acontece, é recomendada a diferenciação da série duas vezes, com o modelo SARIMA (p,d,q) x (P,D,Q)s. Na diferenciação sazonal, aplica-se o operador $(1 - B^s)$, sendo s o número de períodos por ciclo sazonal. Muitas vezes quando essa diferenciação é feita, a série já se torna estacionária, sem a necessidade de diferenciação para eliminar a tendência. Caso contrário, a diferenciação da tendência é feita com o operador $(1 - B)$.

3.4. MODELAGEM ARIMA

Os modelos ARIMA podem possuir termos auto-regressivos (a variável dependente e os termos defasados no tempo) e podem ser integrados quando seus dados originais não são estacionários. Será necessário diferenciar a série original dos dados d vezes até obter uma série estacionária (sua média e variância sejam constantes no tempo). Posteriormente, a série obtida pode ser modelada por um processo ARMA(p,q) (STENGEL, 1986).

Segundo Morettin (2008), no estabelecimento de um modelo ARIMA para uma série temporal há três estágios a considerar: identificação, estimação e diagnóstico.

A identificação tem como objetivo determinar os valores “ p ”, “ d ” e “ q ” do modelo ARIMA (p,d,q). O procedimento de identificação consiste em três partes:

- Verificar se existe necessidade de uma transformação na série original, com o objetivo de estabilizar sua variância;
- Tomar diferenças da série, quantas vezes forem necessárias para se obter uma série estacionária;
- Selecionar o modelo por meio das análises de auto-correlações (FAC) e auto-correlações parciais (FACP) estimadas.

A identificação das ordens “ p ” e “ q ” do modelo ARMA a partir dos gráficos FAC e FACP tem o procedimento a seguir:

Modelo	FAC	FACP
AR (p)	Decaimento gradativo	Decaimento brusco após defasagem p
MA (q)	Decaimento brusco após defasagem q	Decaimento gradativo
ARMA (p,q)	Decaimento gradativo, com onda senoidal amortecida após a defasagem ($q-p$)	Decaimento gradativo, com onda senoidal amortecida após a defasagem ($p-q$)

Figura 5 – Comportamento das funções FAC e FACP para modelos estacionários

Na estimação, devem-se estimar os parâmetros do modelo, selecionar o melhor modelo utilizando critérios adequados (BIC e MAPE) e realizar os testes de sobrefixação (modelos mais elaborados a serem comparados com os mais simples).

O BIC (Critério de Informação Bayesiano) permite comparar diferentes modelos para uma mesma série e escolher o melhor entre eles, que é o modelo que minimiza as variâncias dos resíduos e que sofra uma penalidade se forem incluídos mais parâmetros (parcimônia). Define-se o BIC através da expressão:

$$BIC = \frac{1}{n} \sum_{t=1}^n (x_t - \hat{x}_t)^2 \cdot n^{m/2n} \quad (6)$$

onde m é o número de parâmetros do modelo e n é o número de observações da série temporal.

O MAPE (Erro Médio Absoluto Percentual) indica o valor médio do erro percentual das previsões e é calculado pela fórmula:

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{x_t - \hat{x}_t}{x_t} \right| \times 100 \quad (7)$$

onde n é o número de previsões realizadas, x_t o valor da saída desejada e \hat{x}_t o valor da saída prevista para o índice t .

Na terceira etapa, de diagnóstico, é necessária a realização do teste de Ljung-Box, avaliar a FAC dos resíduos e verificar se os resíduos são ruído branco.

O teste de Ljung-Box é um teste para as autocorrelações dos resíduos estimados, que, apesar de não detectar quebras específicas no comportamento do ruído branco, pode indicar se esses valores são muito altos. Se o modelo for apropriado, a estatística:

$$Q(K) = n(n+2) \sum_{j=1}^K \frac{\hat{r}_j^2}{(n-j)} \quad (8)$$

terá aproximadamente uma distribuição qui-quadrado com $K-p-q$ graus de liberdade. A hipótese de ruído branco para os resíduos é rejeitada para valores grandes de $Q(K)$.

Com todas essas etapas efetuadas, pode-se então, fazer a previsão da série para os períodos futuros.

4. MODELO DE REGRESSÃO DINÂMICA

Em geral, os modelos de regressão linear consideram que os erros gerados pelo modelo possuem média zero, variância constante, distribuição normal e independência (o que implica na inexistência de correlação serial). Na prática, no entanto, os resíduos tendem a apresentar correlação positiva, e erros positivos tendem a ser seguidos por outros também positivos (o mesmo é observado para resíduos negativos). Os modelos de regressão dinâmica estendem os modelos usuais de regressão ao levantarem esta restrição.

Os modelos de regressão dinâmica combinam a dinâmica das séries temporais e o efeito das variáveis explicativas. O termo “regressão dinâmica” não indica que os parâmetros do modelo evoluem no tempo. O termo “dinâmico” está relacionado à estrutura de defasagem da série temporal e das variáveis causais do modelo.

Modelos de regressão dinâmica devem ser usados quando existe uma estrutura de dependência serial entre a variável de interesse e as variáveis causais e, ao mesmo tempo, quando a estrutura de correlação da série dependente indicar que não se pode supor a independência dos erros.

Nos modelos de regressão dinâmica, a variável dependente é explicada por seus valores defasados e pelos valores atuais e passados de variáveis causais ou exógenas. As variáveis exógenas nestes casos são tratadas como séries temporais.

Os modelos de regressão dinâmica podem ser descritos pela equação:

$$\varphi(B) \cdot Y_t = \beta \cdot Z_t + \varepsilon_t \quad (9)$$

onde:

ε_t são erros independentes e identicamente distribuídos com $N(0, \sigma^2)$

Z_t é o vetor de variáveis exógenas no instante t

β é o vetor de coeficientes das variáveis exógenas

$\varphi(B)$ é o polinômio AR(p). Se $\varphi(B)=1$, o modelo se reduz a um modelo de regressão convencional

Y_t é a variável dependente (endógena) no instante t

A estratégia usualmente utilizada para construir um modelo de regressão dinâmica é a estratégia *bottom-up*, ou seja, parte-se de um modelo simples e vai refinando e incluindo novas variáveis até encontrar um modelo apropriado, conforme mostrado no diagrama abaixo.

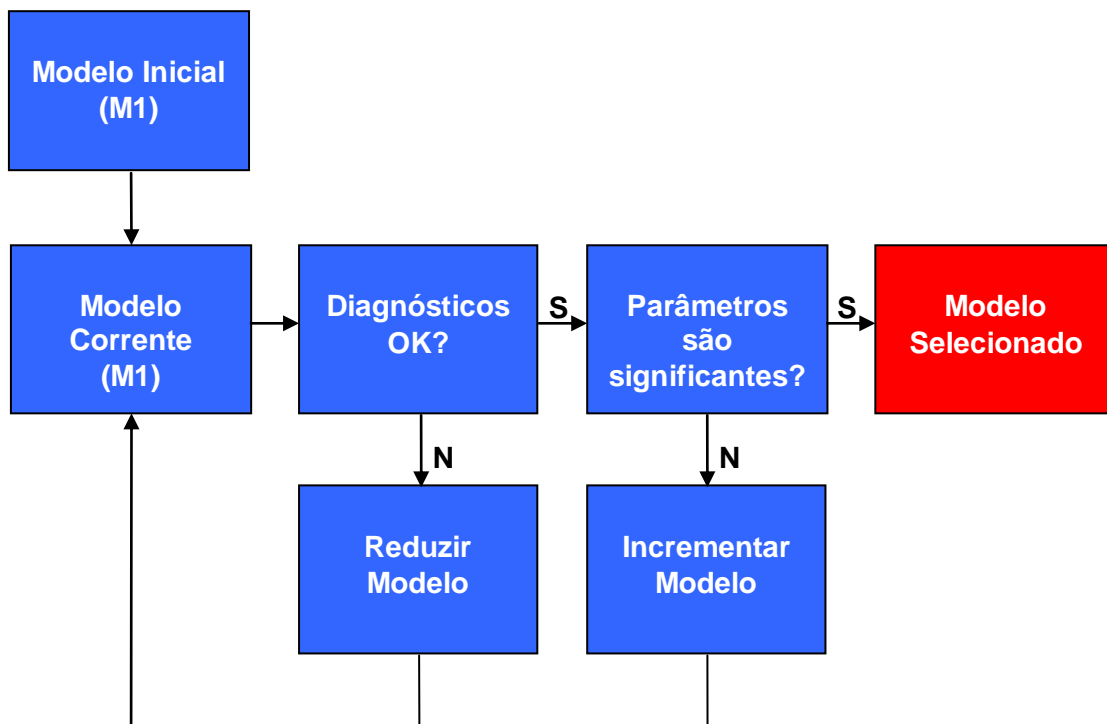


Figura 6 – Diagrama de construção do modelo de regressão dinâmica

Capítulo III

DESCRIÇÃO

1. DESCRIÇÃO DO PROCESSO PRODUTIVO

O metanol (CH_3OH), também chamado de carbinol, hidróxido de metila, monohidroximetano, álcool metílico ou álcool de madeira, é um líquido incolor, volátil e com odor alcoólico levemente adocicado, sendo o mais simples dos alcoóis alifáticos, possuindo apenas um átomo de carbono em sua estrutura molecular. Devido a sua polaridade, o metanol dissolve diversas substâncias inorgânicas, sendo uma das mais importantes matérias-primas químicas no mundo contemporâneo.

A maioria dos processos de produção de metanol é baseada no gás de síntese, obtido do gás natural, como matéria-prima. O processo mais comum envolve a reforma catalítica a vapor do gás natural, resultando em uma mistura de óxidos de carbono e hidrogênio. Esta mistura é comprimida e conduzida à produção de metanol, sendo posteriormente separada de outros subprodutos através da destilação fracionada. As etapas de produção de metanol compreendem a geração de gás de síntese, a síntese do metanol e a destilação dos produtos de reação.

A geração do gás de síntese promove a produção de óxidos de carbono e hidrogênio a partir da reforma a vapor do gás natural com ausência de oxigênio, sobre um leito catalítico de níquel.

A etapa de síntese converte o gás de síntese em metanol cru através da reação de mistura de óxidos de carbono com hidrogênio sobre um leito catalítico de cobre, zinco e alumínio. Esta parte do processo requer uma alimentação adicional de gás carbônico para promover o controle cinético da reação de síntese, aumento a conversão de gás de síntese em metanol.

A etapa seguinte compreende o processo de destilação fracionada dos produtos da reação de metanol, promovendo a separação do produto principal das demais impurezas obtidas no processo anterior, obtendo o metanol em grau químico comercial.

Existe uma grande interação entre as três etapas acima apresentadas devido à extensiva recuperação de calor e reciclo de vapor entre elas. O processo de produção de metanol está demonstrado de forma simplificada abaixo:

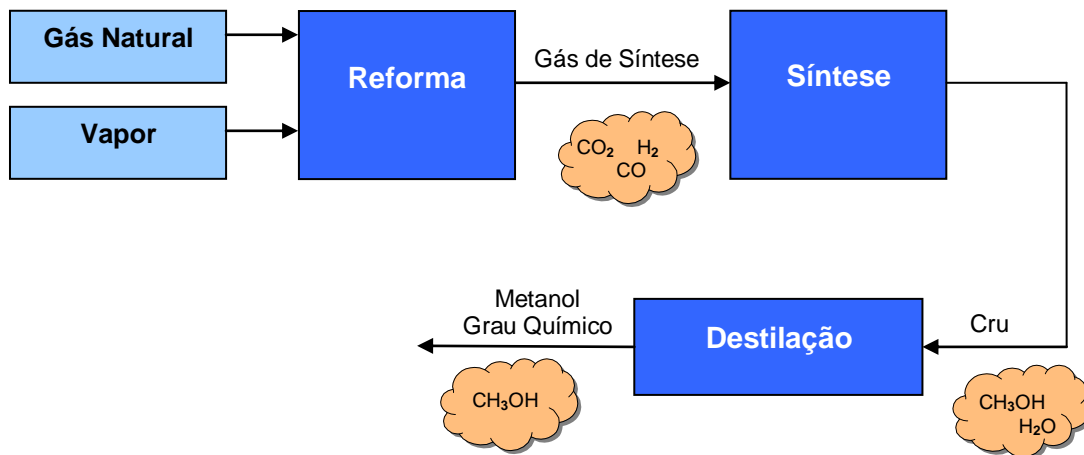


Figura 7 – Processo produtivo de metanol base gás natural

A produção do metanol nos últimos tempos tem se concentrado em regiões onde a matéria-prima tem o custo mais baixo. Na Europa e nos EUA, a construção de novas unidades praticamente não ocorreu a partir dos anos 1980. Os investimentos em metanol passaram a ser feitos na América Central e do Sul, Oriente Médio e Ásia, em megaplantas, pois nessas regiões há grande disponibilidade de gás natural a baixo custo. A produção de megaplantas nessas regiões tem como objetivo ter grande escala de produção e menor custo de produção, para tornarem-se mais competitivas no mercado global. A China, por outro lado, tem optado por construir um maior número de plantas menores, objetivando o abastecimento regional, e sua matéria-prima é o carvão.

O metanol é uma *commodity*, portanto qualquer retração em sua demanda em determinada região ou em função de aplicação específica, provocará excedente, que poderá ser facilmente alocado em outra região ou uso, que esteja em aumento de consumo.

2. DESCRIÇÃO DO PROBLEMA

O problema consiste em analisar os dados de preços mensais do metanol de janeiro de 2000 até dezembro de 2009 através dos modelos Box & Jenkins e Regressão Dinâmica, e verificar qual deles melhor explica a série. Em seguida deve ser feita a previsão de 12 meses (ano de 2010) e compará-la com os dados reais deste ano.

Capítulo IV DESENVOLVIMENTO

1. ANÁLISE DA SÉRIE

O gráfico da série de preços do metanol desde janeiro de 2000 está destacado na figura 8.

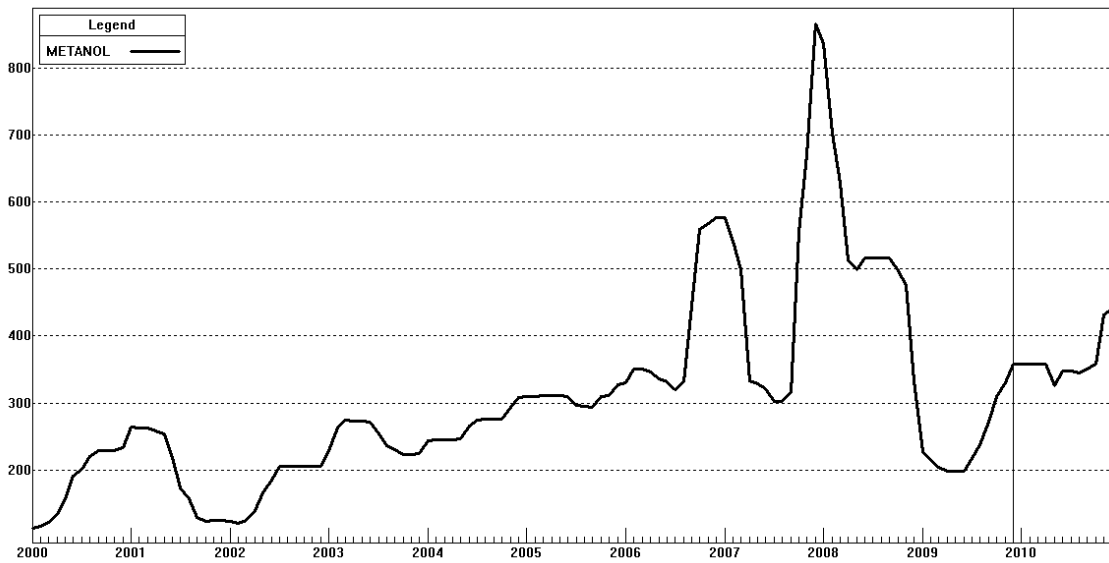


Figura 8 – Série de preços de metanol

1.1. ANÁLISE DA SÉRIE SEM TRANSFORMAÇÃO

Para análise da série de preços do metanol foram feitos os testes de normalidade e homocedasticidade.

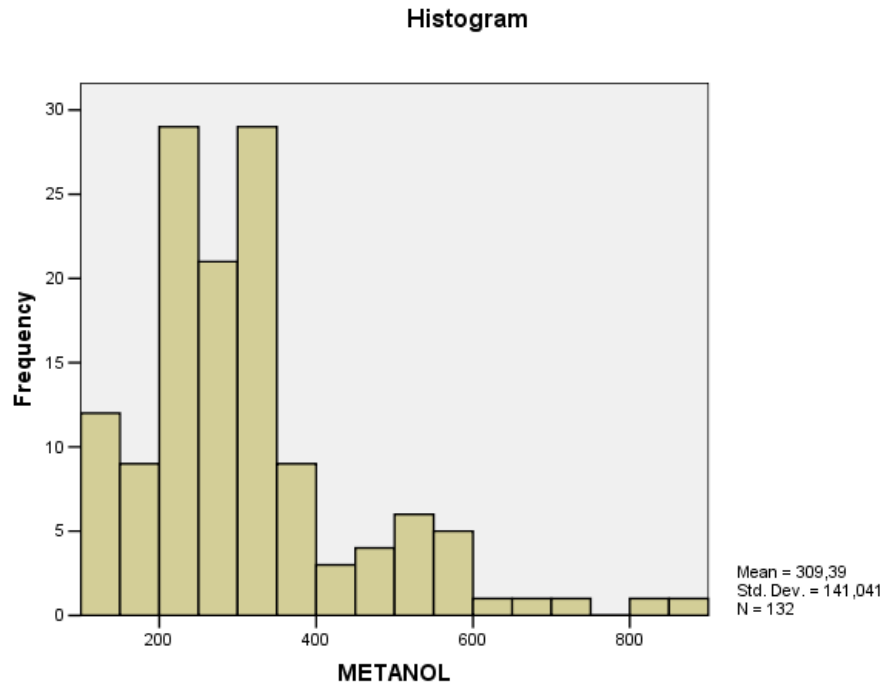


Figura 9 – Histograma dos preços de metanol sem transformação

Tests of Normality

	Kolmogorov -Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
METANOL	,190	132	,000	,883	132	,000

a. Lilliefors Significance Correction

Figura 10 – Teste de normalidade da série sem transformação

Test of Homogeneity of Variance

		Levene Statistic	df 1	df 2	Sig.
METANOL	Based on Mean	35,154	1	130	,000
	Based on Median	14,609	1	130	,000
	Based on Median and with adjusted df	14,609	1	78,133	,000
	Based on trimmed mean	28,908	1	130	,000

Figura 11 – Teste de homocedasticidade da série sem transformação

Através dos resultados do teste de normalidade rejeita-se a hipótese nula (de que os dados são normais), pois o sig = 0. Desta forma, considera-se que os dados não são normais.

Para o teste de homocedasticidade a série foi dividida em dois grupos. Como mostra a Figura 11, sig = 0 indicando que há evidências para se rejeitar a hipótese nula (homocedasticidade).

Com isso, foi necessário transformar a série para que ela se aproximasse a normal e homocedástica, para atender as premissas do ARIMA. Desta forma, a série de preços de metanol foi transformada para logaritmo neperiano (ln).

1.2. ANÁLISE DA SÉRIE TRANSFORMADA PARA LN

Com a série transformada para logaritmo neperiano, foram repetidas as análises de normalidade e homocedasticidade.

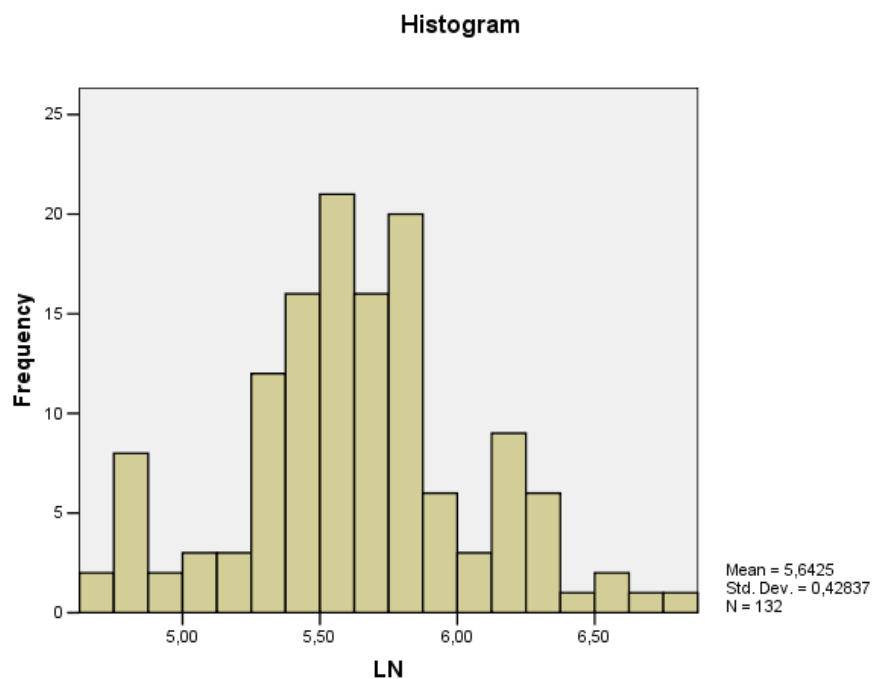


Figura 12 – Histograma dos preços de metanol transformada para LN

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
LN	,114	132	,000	,975	132	,016

a. Lilliefors Significance Correction

Figura 13 – Teste de normalidade da série transformada para LN

Test of Homogeneity of Variance

		Levene Statistic	df 1	df 2	Sig.
LN	Based on Mean	,772	1	130	,381
	Based on Median	,459	1	130	,499
	Based on Median and with adjusted df	,459	1	127,904	,499
	Based on trimmed mean	,800	1	130	,373

Figura 14 – Teste de homocedasticidade da série transformada para LN

No teste de normalidade, com a transformação da série de preços do metanol para logarítmo neperiano, o sig aumentou para 0,016. Já no teste de homocedasticidade baseado na média, o sig subiu para 0,381. Com esses resultados não podemos afirmar que a série é normal e homocedástica, porém, como os resultados apresentados foram melhores do que a série sem transformação, iremos utilizar a série transformada para construir o modelo ARIMA e de regressão dinâmica.

2. MODELAGEM ARIMA

Os pressupostos iniciais para aplicação da modelagem de Box e Jenkins foram verificados no item 1. Após essa verificação é necessária a análise da FAC e FACP da série para a identificação dos parâmetros do modelo ARIMA (p,d,q). Todas as análises foram feitas com um hold out de 12 meses.

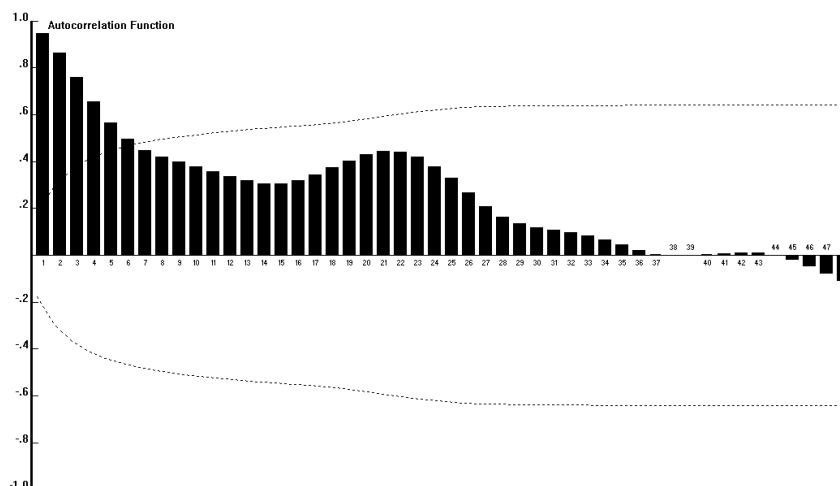


Figura 15 – FAC da série

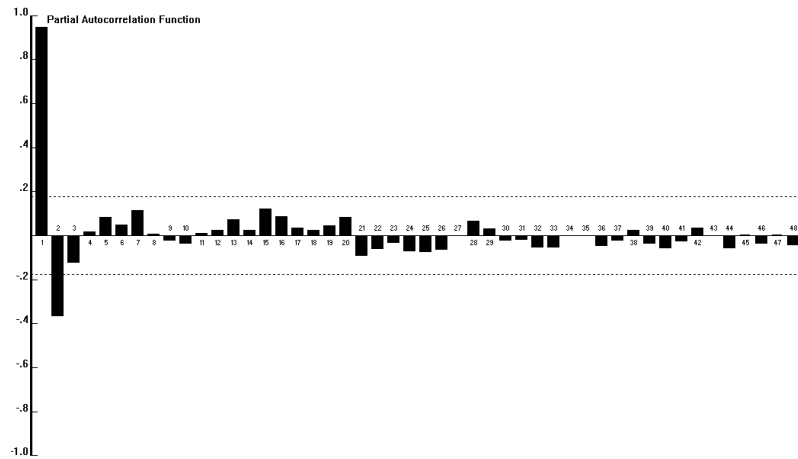


Figura 16 – FACP da série

Analisando os gráficos acima, observa-se a necessidade de diferenciação da série uma vez na parte de tendência, pois a FAC da série decai lentamente. Com esta diferenciação, a FAC e a FACP aparecem da seguinte forma:

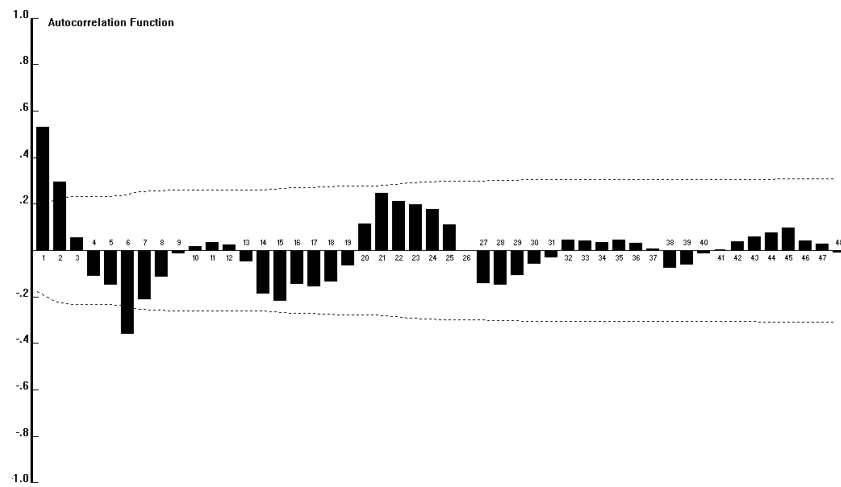


Figura 17 – FAC da série com uma diferenciação na parte de tendência

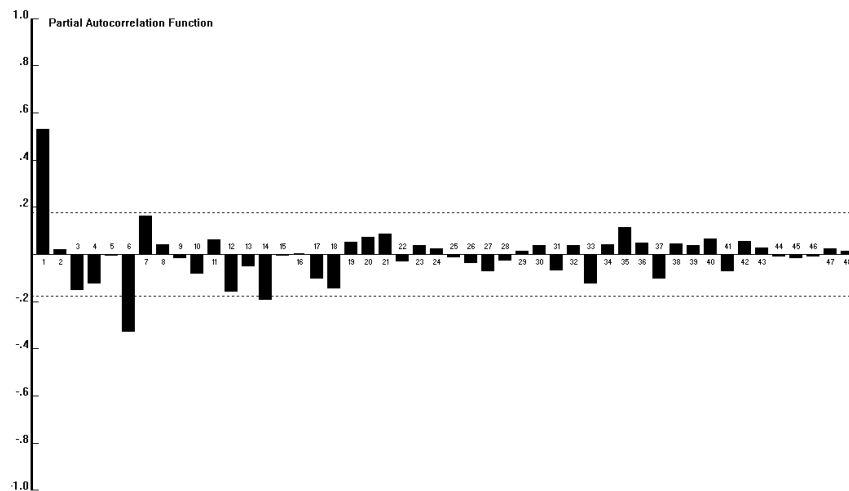


Figura 18 – FACP da série com uma diferenciação na parte de tendência

Após a diferenciação da parte da tendência e análise da FAC e FACP, deve-se então testar os modelos. O primeiro modelo a ser testado é o ARIMA (1,1,0), com holdout de 12.

2.1. ARIMA (1,1,0)

Forecast Model for METANOL
ARIMA(1,1,0) with log transform

Term	Coefficient	Std. Error	t-Statistic	Significance
a[1]	0.5411	0.0773	6.9993	1.0000

Within-Sample Statistics

Sample size	120	Number of parameters	1
Mean	5.617	Standard deviation	0.4406
R-square	0.9526	Adjusted R-square	0.9526
Durbin-Watson	2.013	* Ljung-Box(18)=33.79	P=0.9866
Forecast error	0.09594	BIC	26.81
MAPE	0.05493	RMSE	40.66
MAD	19.53		

Out-of-Sample Rolling Evaluation

H	N	MAD	Cumulative Average	MAPE	Cumulative Average	GMRAE	Cumulative Average
1	12	18.280	18.280	0.048	0.048	1.290	1.160
2	11	27.638	22.756	0.072	0.059	1.417	1.220
3	10	33.183	25.915	0.086	0.067	1.470	1.261
4	9	38.538	28.620	0.099	0.074	1.173	1.237
5	8	39.527	30.365	0.103	0.079	1.140	1.221
6	7	39.501	31.487	0.097	0.081	1.234	1.222
7	6	47.622	33.024	0.115	0.084	1.353	1.234
8	5	42.493	33.720	0.104	0.086	1.310	1.240
9	4	50.072	34.629	0.121	0.088	1.569	1.256
10	3	64.014	35.804	0.152	0.090	3.739	1.312
11	2	60.900	36.456	0.139	0.091	0.724	1.292
12	1	47.300	36.595	0.107	0.092	0.571	1.279

Figura 19 – Resultado do modelo ARIMA (1,1,0)

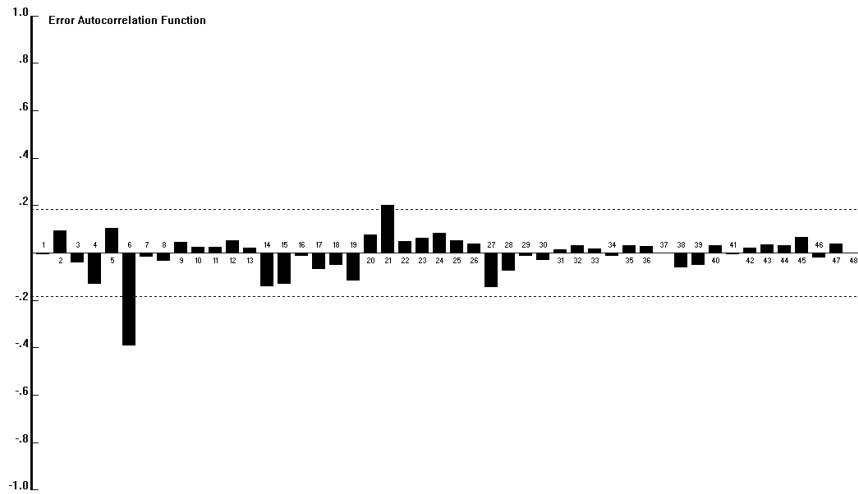


Figura 20 - FAC dos resíduos

Analisando os dados apresentados, o coeficiente testado é significativo. A FAC dos resíduos apresenta um lag significativo (lag 6) que será considerado espúrio, dada a característica dos dados.

2.2. ARIMA (2,1,0)

Forecast Model for METANOL
ARIMA(2,1,0) with log transform

Term	Coefficient	Std. Error	t-Statistic	Significance
a[1]	0.5313	0.0917	5.7916	1.0000
a[2]	0.0173	0.0918	0.1883	0.1490 <-

Try alternative model ARIMA(1,1,0)

Within-Sample Statistics

Sample size	120	Number of parameters	2
Mean	5.617	Standard deviation	0.4406
R-square	0.9526	Adjusted R-square	0.9522
Durbin-Watson	1.988	* Ljung-Box(18)=34.49	P=0.9891
Forecast error	0.09633	BIC	27.35
MAPE	0.05489	RMSE	40.61
MAD	19.5		

Out-of-Sample Rolling Evaluation

H	N	MAD	Cumulative Average	MAPE	Cumulative Average	GMRAE	Cumulative Average
1	12	18.222	18.222	0.048	0.048	1.311	1.171
2	11	27.606	22.710	0.072	0.059	1.357	1.208
3	10	33.240	25.901	0.086	0.067	1.449	1.248
4	9	38.679	28.639	0.100	0.074	1.170	1.226
5	8	39.922	30.444	0.105	0.079	1.160	1.215
6	7	40.105	31.631	0.098	0.081	1.263	1.221
7	6	48.262	33.215	0.117	0.085	1.383	1.236
8	5	43.114	33.943	0.106	0.086	1.360	1.244

9	4	50.528	34.864	0.122	0.088	1.424	1.254
10	3	64.345	36.043	0.154	0.091	3.794	1.311
11	2	59.297	36.647	0.135	0.092	0.699	1.289
12	1	45.183	36.757	0.103	0.092	0.546	1.275

Figura 21 – Resultado do modelo ARIMA (2,1,0)

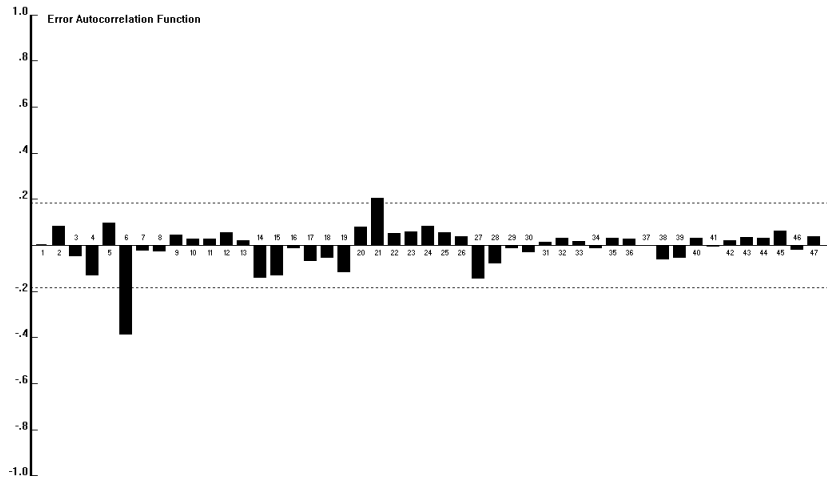


Figura 22 – FAC dos resíduos

Analisando os dados apresentados, como o coeficiente a[2] não é significativo, este modelo será descartado.

2.3. ARIMA (1,1,1)

Forecast Model for METANOL
ARIMA(1,1,1) with log transform

Term	Coefficient	Std. Error	t-Statistic	Significance
a[1]	0.5531	0.1409	3.9250	0.9999
b[1]	0.0166	0.1687	0.0982	0.0780 <-

Insignificant MA terms are harmless.

Within-Sample Statistics

Sample size	120	Number of parameters	2
Mean	5.617	Standard deviation	0.4406
R-square	0.9526	Adjusted R-square	0.9522
Durbin-Watson	2.001	* Ljung-Box(18)=	34.57 P=0.9893
Forecast error	0.09634	BIC	27.35
MAPE	0.0549	RMSE	40.64
MAD	19.51		

Out-of-Sample Rolling Evaluation

H	N	MAD	Cumulative Average	MAPE	Cumulative Average	GMRAE	Cumulative Average
1	12	18.258	18.258	0.048	0.048	1.301	1.166
2	11	27.630	22.740	0.072	0.059	1.388	1.214
3	10	33.228	25.919	0.086	0.067	1.460	1.255

4	9	38.635	28.644	0.099	0.074	1.173	1.232
5	8	39.748	30.420	0.104	0.079	1.151	1.219
6	7	39.832	31.576	0.098	0.081	1.250	1.223
7	6	47.986	33.139	0.116	0.085	1.370	1.236
8	5	42.836	33.852	0.105	0.086	1.334	1.243
9	4	50.219	34.761	0.121	0.088	0.954	1.225
10	3	64.215	35.939	0.153	0.091	3.772	1.281
11	2	60.011	36.565	0.137	0.092	0.710	1.262
12	1	46.095	36.687	0.105	0.092	0.557	1.248

Figura 23 – Resultado do modelo ARIMA (1,1,1)

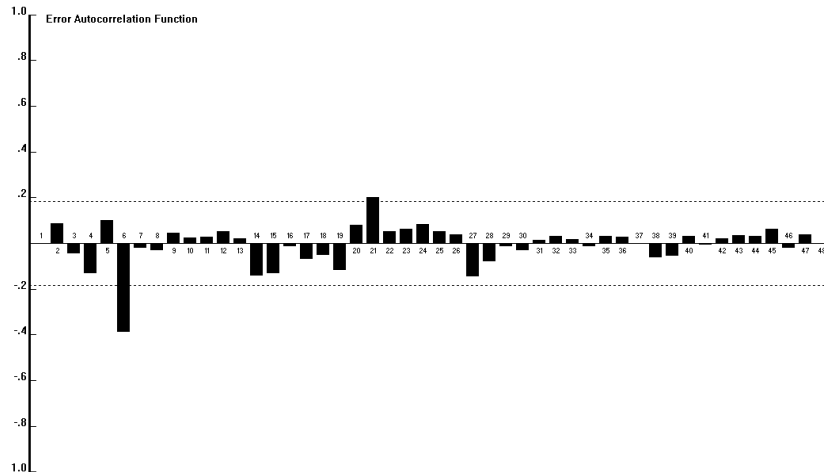


Figura 24 – FAC dos resíduos

Neste caso, o coeficiente $b[1]$ também não se apresentou significativo. Portanto, este modelo será descartado.

2.4. ARIMA (1,1,2)

Forecast Model for METANOL
ARIMA(1,1,2) with log transform

Term	Coefficient	Std. Error	t-Statistic	Significance
a[1]	0.0581	0.2260	0.2572	0.2025 <-
b[1]	-0.4744	0.2111	-2.2478	0.9735
b[2]	-0.3985	0.1151	-3.4629	0.9993

Try alternative model ARIMA(0,1,2)

Within-Sample Statistics

Sample size 120	Number of parameters 3
Mean 5.617	Standard deviation 0.4406
R-square 0.954	Adjusted R-square 0.9533
Durbin-Watson 1.996	** Ljung-Box(18)=37.3 P=0.9952
Forecast error 0.09526	BIC 27.47
MAPE 0.05881	RMSE 37.69
MAD 19.94	

Out-of-Sample Rolling Evaluation

H	N	MAD	Cumulative Average	MAPE	Cumulative Average	GMRAE	Cumulative Average
1	12	19.459	19.459	0.052	0.052	1.872	1.442
2	11	30.023	24.511	0.079	0.065	1.896	1.512
3	10	32.583	26.957	0.083	0.070	1.088	1.362
4	9	37.065	29.123	0.095	0.076	1.053	1.287
5	8	35.714	30.178	0.093	0.078	0.955	1.227
6	7	36.336	30.934	0.088	0.079	0.964	1.191
7	6	43.315	32.113	0.103	0.082	0.928	1.163
8	5	39.386	32.648	0.095	0.083	1.858	1.204
9	4	46.941	33.442	0.111	0.084	1.782	1.231
10	3	61.036	34.546	0.143	0.087	3.100	1.277
11	2	69.818	35.462	0.160	0.089	0.871	1.264
12	1	64.070	35.829	0.145	0.089	0.774	1.256

Figura 25 – Resultado do modelo ARIMA (1,1,2)

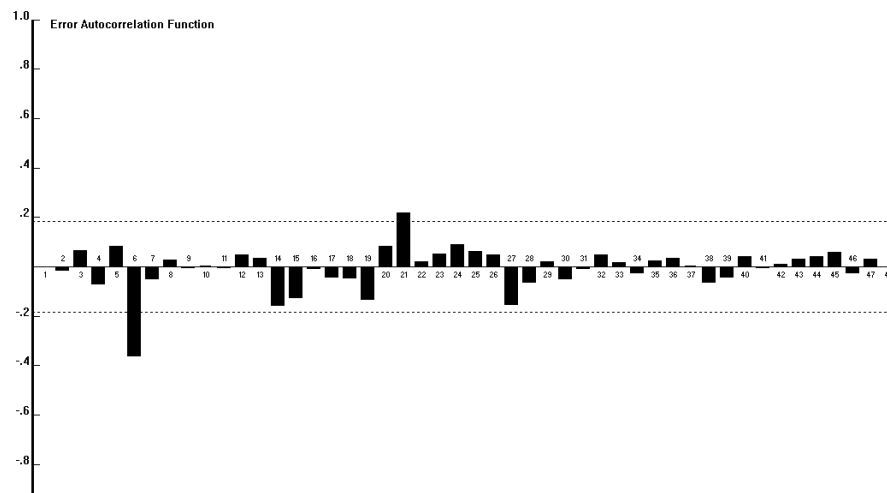


Figura 26 – FAC dos resíduos

Observa-se neste modelo que o coeficiente $a[1]$ não é significativo. Sendo assim, este modelo será descartado.

2.5. ARIMA (0,1,2)

Forecast Model for METANOL
ARIMA(0,1,2) with log transform

Term	Coefficient	Std. Error	t-Statistic	Significance
b[1]	-0.5256	0.0838	-6.2699	1.0000
b[2]	-0.4393	0.0837	-5.2489	1.0000

Within-Sample Statistics

Sample size	120	Number of parameters	2
Mean	5.617	Standard deviation	0.4406
R-square	0.9541	Adjusted R-square	0.9537
Durbin-Watson	1.984	* Ljung-Box (18)=29.44	P=0.9567
Forecast error	0.09482	BIC	26.92
MAPE	0.05938	RMSE	37.63
MAD	20.02		

Out-of-Sample Rolling Evaluation

H	N	MAD	Cumulative Average	MAPE	Cumulative Average	GMRAE	Cumulative Average
1	12	19.551	19.551	0.052	0.052	1.912	1.459
2	11	30.323	24.703	0.080	0.065	1.947	1.535
3	10	32.633	27.106	0.083	0.071	1.245	1.422
4	9	37.051	29.237	0.095	0.076	1.021	1.324
5	8	35.571	30.250	0.093	0.079	0.930	1.251
6	7	36.329	30.997	0.088	0.080	0.997	1.217
7	6	43.330	32.171	0.103	0.082	0.990	1.193
8	5	39.349	32.699	0.095	0.083	1.906	1.235
9	4	46.694	33.477	0.111	0.084	1.645	1.255
10	3	60.889	34.573	0.142	0.087	3.064	1.300
11	2	69.911	35.491	0.160	0.089	0.873	1.287
12	1	64.708	35.866	0.147	0.089	0.782	1.279

Figura 27 – Resultado do modelo ARIMA (0,1,2)

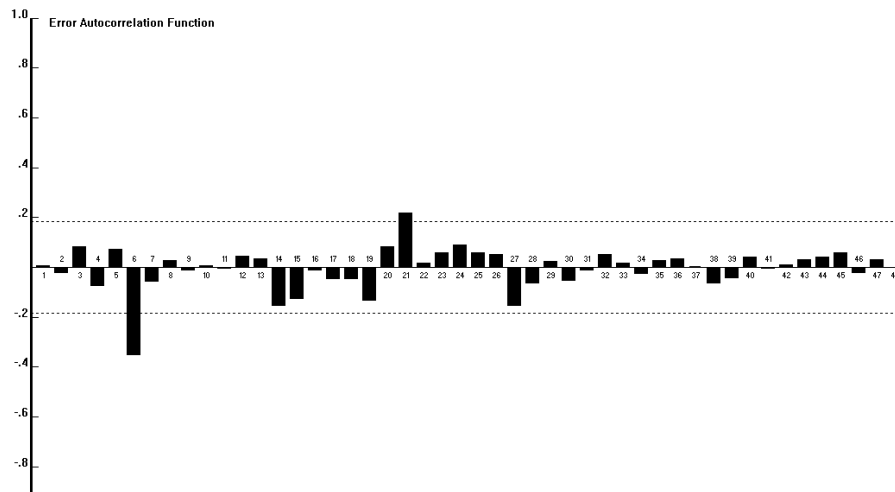


Figura 28 – FAC dos resíduos

Este modelo apresenta os dois coeficientes ($b[1]$, $b[2]$) significativos e a FAC dos resíduos apresenta um lag significativo (lag 6) que será considerado espúrio, dada a característica dos dados.

Comparando todos os modelos apresentados com coeficientes significativos, o modelo ARIMA (1,1,0) foi o que apresentou os menores BIC e MAPE e um R-Square muito alto (0,9526). Por isso, este modelo é o escolhido como o mais adequado para a série.

3. REGRESSÃO DINÂMICA

Para a construção da regressão dinâmica foram colocadas séries causais para analisar se estas séries influenciam no preço do metanol. Foram testadas a série de preços de petróleo, que é a origem do metanol; a série de gás natural, que é a matéria prima principal do metanol; e as séries de preços da gasolina e do óleo combustível, os quais o metanol pode substituir energeticamente. A série de preços do metanol utilizada é a

transformada em ln, conforme transformação feita no item 1.2. Todas as análises foram feitas com hold out de 12 meses.

De início colocou-se em teste somente a série e a constante.

3.1. VARIÁVEL: CONSTANTE

Forecast Model for METANOL with log transform
Regression(1 regressors, 0 lagged errors)

Term	Coefficient	Std. Error	t-Statistic	Significance
_CONST	5.617119	0.040220	139.660501	1.000000

Within-Sample Statistics

Sample size 120	Number of parameters 1
Mean 5.617	Standard deviation 0.4406
R-square 0	Adjusted R-square 0
Durbin-Watson 0.06694	** Ljung-Box(18)=655.7 P=1
Forecast error 0.4406	BIC 123.1
MAPE 0.3485	RMSE 148.6
MAD 100.1	

Out-of-Sample Rolling Evaluation

H	N	MAD	Cumulative Average	MAPE	Cumulative Average	GMRAE	Cumulative Average
1	12	90.072	90.072	0.241	0.241	7.032	3.120
2	11	90.745	90.394	0.242	0.242	5.492	3.274
3	10	91.552	90.745	0.243	0.242	4.921	3.363
4	9	92.539	91.129	0.245	0.243	3.864	3.355
5	8	93.772	91.552	0.246	0.243	3.920	3.440
6	7	99.822	92.568	0.259	0.245	7.462	3.783
7	6	104.231	93.679	0.267	0.247	7.565	4.041
8	5	110.405	94.908	0.278	0.249	8.000	4.249
9	4	120.413	96.325	0.297	0.252	8.700	4.422
10	3	135.320	97.885	0.324	0.255	8.085	4.530
11	2	161.311	99.533	0.370	0.258	2.054	4.438
12	1	165.467	100.378	0.376	0.259	1.999	4.393

Variable specification test battery

GASNATURAL	Chi Square(1)=57.19	Percentile=1.0000	**
GASOLINA	64.85	1.0000	**
OLEOCOMB	72.36	1.0000	**
PETROLEO	70.86	1.0000	**
_TREND	52.10	1.0000	**

Try adding OLEOCOMB to model.

Figura 29 – Resultado do modelo utilizando somente a constante

Este é apenas um modelo preliminar, por isto o R-Square foi zero. De acordo com a análise das variáveis, o óleo combustível é o que apresenta o maior Chi-Square e será incluído no modelo.

3.2. VARIÁVEIS: CONSTANTE E ÓLEO COMBUSTÍVEL

Forecast Model for METANOL with log transform
Regression(2 regressors, 0 lagged errors)

Term	Coefficient	Std. Error	t-Statistic	Significance
OLEOCOMB	0.004709	0.000344	13.677405	1.000000
_CONST	4.962955	0.054023	91.866875	1.000000

Within-Sample Statistics

Sample size	120	Number of parameters	2
Mean	5.617	Standard deviation	0.4406
R-square	0.6132	Adjusted R-square	0.6099
Durbin-Watson	0.2392	** Ljung-Box(18)=206.9	P=1
Forecast error	0.2752	BIC	78.12
MAPE	0.2169	RMSE	105.9
MAD	68.55		

Out-of-Sample Rolling Evaluation

H	N	MAD	Cumulative Average	MAPE	Cumulative Average	GMRAE	Cumulative Average
1	12	22.175	22.175	0.063	0.063	1.643	1.336
2	11	23.054	22.596	0.065	0.064	1.220	1.247
3	10	25.329	23.424	0.072	0.066	1.094	1.192
4	9	25.998	23.976	0.074	0.068	0.859	1.115
5	8	23.746	23.939	0.067	0.068	0.871	1.072
6	7	21.012	23.579	0.058	0.067	1.373	1.105
7	6	21.230	23.356	0.059	0.066	1.307	1.123
8	5	23.183	23.343	0.064	0.066	1.410	1.142
9	4	23.336	23.343	0.063	0.066	1.320	1.151
10	3	21.990	23.289	0.058	0.065	0.918	1.140
11	2	9.087	22.920	0.021	0.064	0.104	1.072
12	1	5.063	22.691	0.011	0.063	0.061	1.033

Variable specification test battery

GASNATURAL	Chi Square(1)=13.36	Percentile=0.9997 **
GASOLINA	2.08	0.8509
PETROLEO	0.20	0.3484
OLEOCOMB[-1]	4.50	0.9661 *
_TREND	2.40	0.8789

Try adding GASNATURAL to model.

Figura 30 – Resultado do modelo utilizando a constante e óleo combustível

A variável óleo combustível apresentou-se significativa. Com isso, o R-Square melhora para 0,6132. Analisando as saídas do software, será acrescentada a variável gás natural.

3.3. VARIÁVEIS: CONSTANTE, ÓLEO COMBUSTÍVEL E GÁS NATURAL

Forecast Model for METANOL with log transform
Regression(3 regressors, 0 lagged errors)

Term	Coefficient	Std. Error	t-Statistic	Significance
GASNATURAL	0.054777	0.014107	3.883000	0.999829
OLEOCOMB	0.003482	0.000453	7.679043	1.000000
_CONST	4.813807	0.063897	75.337133	1.000000

Within-Sample Statistics

Sample size 120	Number of parameters 3
Mean 5.617	Standard deviation 0.4406
R-square 0.6574	Adjusted R-square 0.6515
Durbin-Watson 0.3081	** Ljung-Box(18)=211.5 P=1
Forecast error 0.2601	BIC 75.01
MAPE 0.1912	RMSE 111.4
MAD 65.54	

Out-of-Sample Rolling Evaluation

H	N	MAD	Cumulative Average	MAPE	Cumulative Average	GMRAE	Cumulative Average
1	12	38.205	38.205	0.100	0.100	2.921	1.869
2	11	40.465	39.286	0.106	0.103	2.287	1.848
3	10	41.454	39.943	0.108	0.104	2.049	1.825
4	9	42.928	40.583	0.111	0.106	1.609	1.757
5	8	44.629	41.230	0.115	0.107	1.632	1.736
6	7	49.393	42.233	0.126	0.110	3.405	1.886
7	6	51.614	43.126	0.130	0.111	3.410	1.995
8	5	55.721	44.052	0.138	0.113	3.651	2.086
9	4	63.861	45.153	0.156	0.116	4.308	2.172
10	3	71.168	46.193	0.168	0.118	3.856	2.222
11	2	92.783	47.403	0.213	0.120	1.168	2.185
12	1	78.911	47.807	0.179	0.121	0.953	2.162

Variable specification test battery

GASOLINA	Chi Square(1)=1.34	Percentile=0.7522
PETROLEO	2.31	0.8718
GASNATURAL[-1]	3.00	0.9166
OLEOCOMB[-1]	2.97	0.9151
_TREND	7.57	0.9941 **

Try adding _TREND to model.

Dynamics test battery

METANOL[- 1]	Chi Square(1)=94.05	Percentile=1.0000 **
METANOL[- 2]	57.27	1.0000 **
METANOL[- 3]	24.37	1.0000 **
METANOL[- 4]	5.95	0.9852 *
METANOL[- 5]	1.01	0.6840
METANOL[- 6]	1.77	0.8168
METANOL[- 7]	2.13	0.8559
METANOL[- 8]	1.29	0.7432
METANOL[- 9]	0.93	0.6659
METANOL[-10]	1.42	0.7673
METANOL[-11]	1.74	0.8133
METANOL[-12]	1.41	0.7647
METANOL[-24]	7.85	0.9949 **
_AUTO[- 1]	Chi Square(1)=82.61	Percentile=1.0000 **
_AUTO[- 2]	41.31	1.0000 **
_AUTO[- 3]	11.28	0.9992 **
_AUTO[- 4]	0.83	0.6375
_AUTO[- 5]	4.47	0.9654 *
_AUTO[- 6]	8.92	0.9972 **
_AUTO[- 7]	6.27	0.9877 *

_AUTO[- 8]	2.03	0.8454
_AUTO[- 9]	1.17	0.7207
_AUTO[-10]	3.93	0.9527 *
_AUTO[-11]	5.07	0.9757 *
_AUTO[-12]	3.31	0.9311
_AUTO[-24]	5.52	0.9812 *

Try adding METANOL[-1] to model.

Figura 31 – Resultado do modelo utilizando a constante, óleo combustível e gás natural

Neste caso, as duas variáveis apresentaram-se significativas. Com a análise das variáveis dinâmicas, verifica-se a indicação para incluir a variável metanol defasada em um período. Será incluída esta variável e retirada a variável óleo combustível, na tentativa de explicar o preço do metanol pelo preço de sua matéria prima principal e seu preço no período anterior.

3.4. VARIÁVEIS: CONSTANTE, GÁS NATURAL E METANOL [-1]

Forecast Model for METANOL with log transform
Regression(3 regressors, 0 lagged errors)

Term	Coefficient	Std. Error	t-Statistic	Significance
GASNATURAL	0.037730	0.007796	4.839424	0.999996
_CONST	4.689613	0.040141	116.827834	1.000000
METANOL[-1]	0.002353	0.000124	18.917188	1.000000

Within-Sample Statistics

Sample size 119	Number of parameters 3
Mean 5.625	Standard deviation 0.4348
R-square 0.8716	Adjusted R-square 0.8694
Durbin-Watson 0.5911	** Ljung-Box(18)=98.85 P=1
Forecast error 0.1571	BIC 45.67
MAPE 0.1082	RMSE 67.18
MAD 36.29	

Out-of-Sample Rolling Evaluation

H	N	MAD	Cumulative Average	MAPE	Cumulative Average	GMRAE	Cumulative Average
1	12	234.592	234.592	0.640	0.640	18.974	5.567
2	11	236.243	235.382	0.643	0.641	15.004	6.282
3	10	237.607	236.056	0.645	0.642	13.444	6.759
4	9	238.956	236.677	0.646	0.643	10.556	7.030
5	8	240.077	237.221	0.647	0.644	10.708	7.520
6	7	246.249	238.330	0.653	0.645	19.426	8.450
7	6	250.724	239.510	0.657	0.646	19.244	9.139
8	5	257.575	240.839	0.664	0.647	19.756	9.672
9	4	268.576	242.380	0.676	0.649	20.404	10.081
10	3	283.279	244.016	0.687	0.650	17.536	10.307
11	2	309.500	245.717	0.709	0.652	3.941	10.053
12	1	311.051	246.554	0.706	0.653	3.757	9.927

Variable specification test battery

```

-----
GASOLINA                      Chi Square( 1)=5.40  Percentile=0.9798  *
OLEOCOMB                      4.54                0.9669  *
PETROLEO                      6.75                0.9906  **
GASNATURAL[-1]                2.20                0.8616
_TREND                        11.26               0.9992  **

```

Try adding _TREND to model.

Dynamics test battery

```

-----
METANOL[- 1]                  Chi Square( 1)=56.16  Percentile=1.0000  **
METANOL[- 2]                  4.16                0.9586  *
METANOL[- 3]                  0.29                0.4100
METANOL[- 4]                  0.88                0.6530
METANOL[- 5]                  0.86                0.6461
METANOL[- 6]                  0.70                0.5988
METANOL[- 7]                  0.64                0.5766
METANOL[- 8]                  1.00                0.6820
METANOL[- 9]                  1.36                0.7559
METANOL[-10]                  1.21                0.7295
METANOL[-11]                  0.87                0.6499
METANOL[-12]                  0.60                0.5615
METANOL[-24]                  10.12               0.9985  **

```

```

-----
_AUTO[- 1]                     Chi Square( 1)=59.67  Percentile=1.0000  **
_AUTO[- 2]                     30.88               1.0000  **
_AUTO[- 3]                     9.21                0.9976  **
_AUTO[- 4]                     1.32                0.7490
_AUTO[- 5]                     0.65                0.5798
_AUTO[- 6]                     2.60                0.8930
_AUTO[- 7]                     1.49                0.7770
_AUTO[- 8]                     0.99                0.6807
_AUTO[- 9]                     0.48                0.5117
_AUTO[-10]                     0.65                0.5800
_AUTO[-11]                     0.96                0.6739
_AUTO[-12]                     1.43                0.7683
_AUTO[-24]                     12.88               0.9997  **

```

Try adding _AUTO[-1] to model.

Figura 32 – Resultado do modelo utilizando a constante, gás natural e metanol [-1]

Este modelo melhorou significativamente o resultado. O R-square passou de 0,6574 no modelo anterior para 0,8716. As variáveis são significativas e o BIC e o MAPE reduziram. Será acrescentada a variável petróleo conforme indicado pelo software.

3.5. VARIÁVEIS: CONSTANTE, GÁS NATURAL, METANOL [-1] E PETRÓLEO

Forecast Model for METANOL with log transform
Regression(4 regressors, 0 lagged errors)

Term	Coefficient	Std. Error	t-Statistic	Significance
GASNATURAL	0.029609	0.008179	3.619926	0.999560
PETROLEO	0.002348	0.000876	2.678717	0.991526
_CONST	4.685262	0.039147	119.682927	1.000000
METANOL[-1]	0.002127	0.000148	14.415090	1.000000

Within-Sample Statistics

Sample size 119	Number of parameters 4
Mean 5.625	Standard deviation 0.4348
R-square 0.8792	Adjusted R-square 0.876
Durbin-Watson 0.563	** Ljung-Box(18)=114.3 P=1
Forecast error 0.1531	BIC 45.21
MAPE 0.1125	RMSE 65.38
MAD 37.72	

Out-of-Sample Rolling Evaluation

H	N	MAD	Cumulative Average	MAPE	Cumulative Average	GMRAE	Cumulative Average
1	12	214.305	214.305	0.584	0.584	17.353	5.284
2	11	215.823	215.031	0.587	0.585	13.716	5.926
3	10	217.024	215.635	0.588	0.586	12.290	6.349
4	9	218.396	216.227	0.590	0.587	9.650	6.580
5	8	219.737	216.788	0.592	0.588	9.789	7.012
6	7	225.627	217.874	0.598	0.589	17.780	7.861
7	6	229.832	219.013	0.602	0.590	17.619	8.489
8	5	236.333	220.286	0.609	0.592	18.103	8.975
9	4	246.822	221.760	0.620	0.593	18.729	9.349
10	3	260.683	223.317	0.632	0.595	16.117	9.555
11	2	286.326	224.954	0.656	0.596	3.646	9.319
12	1	287.205	225.752	0.652	0.597	3.469	9.202

Variable specification test battery

GASOLINA	Chi Square(1)=0.13	Percentile=0.2852
OLEOCOMB	7.52	0.9939 **
GASNATURAL[-1]	1.79	0.8193
PETROLEO[-1]	0.89	0.6538
_TREND	5.07	0.9757 *

Try adding OLEOCOMB to model.

Dynamics test battery

METANOL[- 1]	Chi Square(1)=56.78	Percentile=1.0000 **
METANOL[- 2]	3.79	0.9483
METANOL[- 3]	0.72	0.6046
METANOL[- 4]	2.76	0.9036
METANOL[- 5]	3.67	0.9446
METANOL[- 6]	4.10	0.9572 *
METANOL[- 7]	2.30	0.8709
METANOL[- 8]	1.27	0.7401
METANOL[- 9]	0.84	0.6392
METANOL[-10]	0.86	0.6458
METANOL[-11]	1.19	0.7244
METANOL[-12]	1.97	0.8395
METANOL[-24]	6.80	0.9909 **
_AUTO[- 1]	Chi Square(1)=65.00	Percentile=1.0000 **
_AUTO[- 2]	40.42	1.0000 **
_AUTO[- 3]	15.38	0.9999 **
_AUTO[- 4]	3.56	0.9408
_AUTO[- 5]	1.58	0.7913
_AUTO[- 6]	1.62	0.7971
_AUTO[- 7]	1.15	0.7158
_AUTO[- 8]	1.06	0.6978
_AUTO[- 9]	0.65	0.5799
_AUTO[-10]	0.58	0.5535
_AUTO[-11]	0.68	0.5894
_AUTO[-12]	0.92	0.6637
_AUTO[-24]	11.46	0.9993 **

Figura 33 – Resultado do modelo utilizando a constante, gás natural, metanol [-1] e petróleo

Com a adição do petróleo o modelo apresentou-se melhor, com maior R-Square, sendo todas as variáveis significativas. Será acrescentada a variável óleo combustível novamente, conforme indicado nas saídas do software.

3.6. VARIÁVEIS: CONSTANTE, GÁS NATURAL, METANOL [-1], PETRÓLEO E ÓLEO COMBUSTÍVEL

Forecast Model for METANOL with log transform
Regression(5 regressors, 0 lagged errors)

Term	Coefficient	Std. Error	t-Statistic	Significance
GASNATURAL	0.043336	0.009298	4.660880	0.999991
OLEOCOMB	-0.005564	0.001961	-2.837109	0.994611
PETROLEO	0.017010	0.005238	3.247695	0.998471
_CONST	4.610895	0.046163	99.881876	1.000000
METANOL[-1]	0.002187	0.000145	15.105629	1.000000

Within-Sample Statistics

Sample size 119	Number of parameters 5
Mean 5.625	Standard deviation 0.4348
R-square 0.8871	Adjusted R-square 0.8832
Durbin-Watson 0.72	** Ljung-Box(18)=88.52 P=1
Forecast error 0.1486	BIC 44.58
MAPE 0.106	RMSE 67.19
MAD 36.69	

Out-of-Sample Rolling Evaluation

H	N	MAD	Cumulative Average	MAPE	Cumulative Average	GMRAE	Cumulative Average
1	12	216.570	216.570	0.590	0.590	17.880	5.377
2	11	218.757	217.616	0.594	0.592	14.047	6.030
3	10	220.833	218.591	0.599	0.594	12.586	6.464
4	9	223.295	219.599	0.603	0.596	9.882	6.704
5	8	225.207	220.496	0.606	0.598	10.025	7.150
6	7	230.721	221.752	0.611	0.599	18.155	8.016
7	6	235.054	223.019	0.615	0.601	17.989	8.658
8	5	243.325	224.512	0.627	0.603	18.629	9.160
9	4	255.519	226.235	0.643	0.605	19.408	9.550
10	3	268.966	227.944	0.652	0.607	16.636	9.764
11	2	294.883	229.682	0.676	0.609	3.755	9.525
12	1	294.960	230.519	0.670	0.609	3.563	9.406

Variable specification test battery

GASOLINA	Chi Square(1)=0.00	Percentile=0.0411
GASNATURAL[-1]	2.05	0.8482
OLEOCOMB[-1]	0.43	0.4865
PETROLEO[-1]	2.49	0.8854
_TREND	6.86	0.9912 **

Try adding _TREND to model.

Dynamics test battery

METANOL[- 1]	Chi Square(1)=58.75	Percentile=1.0000 **
METANOL[- 2]	8.45	0.9963 **
METANOL[- 3]	0.61	0.5655
METANOL[- 4]	0.75	0.6143
METANOL[- 5]	1.19	0.7255

METANOL[- 6]	1.67	0.8036
METANOL[- 7]	0.97	0.6755
METANOL[- 8]	0.63	0.5732
METANOL[- 9]	0.65	0.5811
METANOL[-10]	0.62	0.5694
METANOL[-11]	0.52	0.5271
METANOL[-12]	0.69	0.5929
METANOL[-24]	8.96	0.9972 **
Chi Square(1)=52.11 Percentile=1.0000 **		
_AUTO[- 1]	28.47	1.0000 **
_AUTO[- 2]	9.34	0.9978 **
_AUTO[- 3]	2.30	0.8703
_AUTO[- 4]	1.40	0.7639
_AUTO[- 5]	2.01	0.8435
_AUTO[- 6]	1.71	0.8095
_AUTO[- 7]	2.40	0.8790
_AUTO[- 8]	1.74	0.8132
_AUTO[- 9]	0.98	0.6773
_AUTO[-10]	0.60	0.5603
_AUTO[-11]	1.14	0.7145
_AUTO[-12]	12.55	0.9996 **

Try adding METANOL[-1] to model.

Figura 34 – Resultado do modelo utilizando a constante, gás natural, metanol [-1], petróleo e óleo combustível

Este modelo apresentou resultados ainda melhores: o maior R-Square (0,8871), o menor MAPE (0,106) e o menor BIC (44,58). Analisando os outputs do software, chega-se a conclusão de que não é necessário acrescentar mais nenhuma variável. Comparando-se todos os modelos, este foi escolhido como o mais adequado por apresentar os melhores resultados.

4. COMPARAÇÃO ENTRE ARIMA E REGRESSÃO DINÂMICA

Após a realização dos dois modelos é necessário compará-los para escolher o mais eficiente na realização da previsão. A tabela comparativa abaixo mostra os principais indicadores da eficiência dos modelos.

Modelo ARIMA (1,1,0)	Modelo Regressão Dinâmica com variáveis: constante, gás natural, metanol [-1], petróleo e óleo combustível
R-Square: 0,9526	R-Square: 0,8871
MAPE: 0,05493	MAPE: 0,106
BIC: 26,81	BIC: 44,58

Figura 35 – Comparação entre ARIMA e Regressão Dinâmica

O melhor modelo é aquele que apresenta o maior R-Square e os menores MAPE e BIC, portanto o modelo ARIMA (1,1,0) é o modelo escolhido para a previsão. Antes, porém, é necessário rodar o método sem o hold out.

```
Forecast Model for METANOL
ARIMA(1,1,0) with log transform

Term          Coefficient  Std. Error  t-Statistic  Significance
-----
a[1]           0.5238      0.0744      7.0366      1.0000

Within-Sample Statistics
-----
Sample size 132          Number of parameters 1
Mean 5.643              Standard deviation 0.4284
R-square 0.9518         Adjusted R-square 0.9518
Durbin-Watson 2.022    ** Ljung-Box(18)=36.18 P=0.9933
Forecast error 0.09401  BIC 26.92
MAPE 0.05413           RMSE 39.51
MAD 19.33
```

Figura 36 – Resultado do ARIMA (1,1,0) sem hold out

5. PREVISÃO

O resultado da previsão realizada através do método ARIMA (1,1,0) é:

```
Forecast Report
Mon Aug 01 22:49:24 2011
Box-Jenkins model for METANOL

Forecasted Values

Date    2.5 Lower    Forecast    97.5 Upper    Actual
-----
2010-01    309.603     373.652     450.951     357.770
2010-02    270.803     382.538     540.375     357.770
2010-03    237.806     387.433     631.206     357.770
2010-04    210.850     390.108     721.765     357.770
2010-05    188.914     391.563     811.597     326.515
2010-06    170.890     392.353     900.815     348.460
2010-07    155.875     392.781     989.747     348.460
2010-08    143.181     393.012    1078.768     345.467
2010-09    132.300     393.138    1168.232     350.788
2010-10    122.858     393.206    1258.446     358.435
2010-11    114.576     393.242    1349.670     432.250
2010-12    107.242     393.262    1442.120     440.563
```

Figura 37 – Previsão de 12 meses

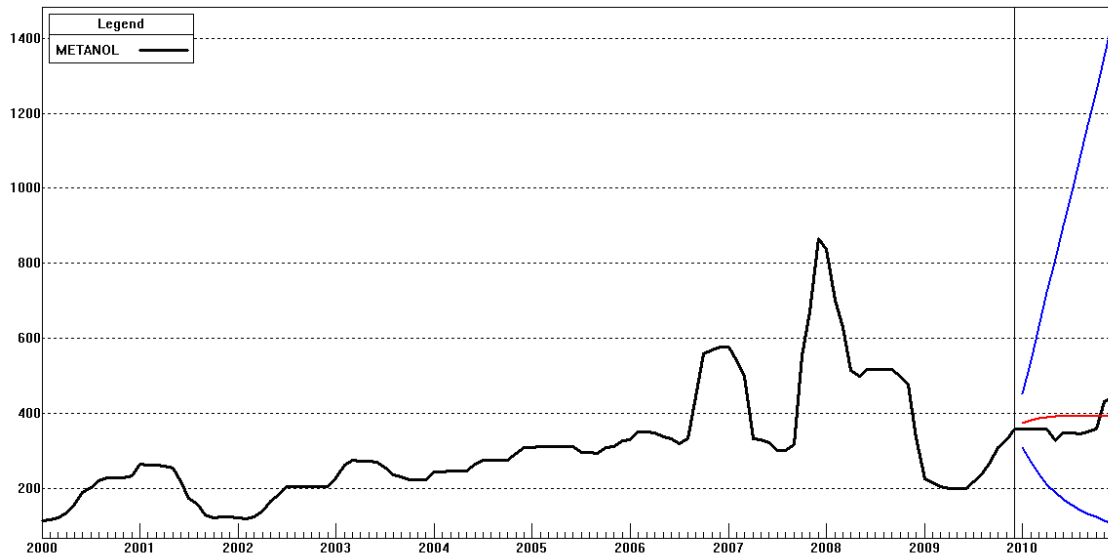


Figura 38 – Gráfico da série de preços de metanol com previsão

Comparando-se os resultados da previsão para 12 meses e o realizado no ano de 2010 e observando-se o gráfico, considera-se a previsão satisfatória, visto que há uma grande dificuldade de previsão para este mercado, que é muito volátil.

Capítulo V CONCLUSÃO

Este trabalho teve como objetivo desenvolver um modelo adequado para a previsão de preços de metanol para 12 meses. Para tanto, utilizou-se a série de preços médios mensais de metanol em um período de 10 anos (de janeiro de 2000 a dezembro de 2009).

O primeiro passo foi transformar a série para logaritmo neperiano para que se aproximasse de uma série normal e homocedástica. Com a série transformada, foram testados os modelos Box & Jenkins e Regressão Dinâmica. No modelo Box & Jenkins analisou-se a série de preços de metanol com base no histórico de dados. Foram testados vários modelos ARIMA. Os indicadores utilizados para compará-los foram o *R-square*, o BIC e o MAPE, sendo que o modelo que apresentou melhor resultado foi o ARIMA (1,1,0). Já na regressão dinâmica incluíram-se outras variáveis além do preço do metanol (preço do petróleo, da gasolina, do óleo combustível e do gás natural, além da série do metanol defasada) para verificar se estas influenciavam nos preços da série estudada. O modelo com melhores indicadores foi o que incluiu as variáveis gás natural, metanol [-1], petróleo, óleo combustível e uma constante.

Após a simulação dos modelos, o ARIMA e a regressão dinâmica com melhores resultados foram comparados para se escolher o mais apropriado. Ambos os modelos apresentaram resultados muito satisfatórios, com altos valores do R-Square e baixos BIC e MAPE, sendo que o ARIMA (1,1,0) apresentou um melhor R-Square (0,9526) e BIC e MAPE menores, e portanto, foi eleito para fazer a previsão de 12 meses.

A previsão para os 12 meses de 2010 utilizando o ARIMA (1,1,0) foi, então, comparada com os dados reais do ano. Pode-se observar que a previsão ficou próxima do realizado, sendo que o resultado obtido foi aceitável, visto que o metanol é uma *commodity* e seus preços são negociados mundialmente, tendo um mercado muito volátil. Com isso, pode-se concluir que o objetivo do trabalho foi alcançado.

Este trabalho demonstrou como é possível utilizar séries reais de preços e adequá-las para modelos matemáticos. Como sugestão de continuação do estudo tem-se o desenvolvimento do modelo para previsão do ano de 2011. Outra proposta é testar modelos mais complexos, como os da família ARCH e GARCH, para verificar se é possível obter modelos com melhores indicadores e previsões ainda mais próximas dos preços reais.

REFERÊNCIAS BIBLIOGRÁFICAS

1. ABIQUIM – Associação Brasileira da Indústria Química. Disponível em: <<http://www.abiquim.org.br>>. Acesso em: março de 2011.
2. CMAI – Chemical Market Associate, Inc.; Disponível em: <<http://www.cmaiglobal.com>>. Acesso em: março de 2011.
3. MAKRIDAKIS, S. G., WHEELWRIGHT, S. C., MCGEE, V. E. **Fundamentals of Quantitative Forecasting**. New York: John Wiley & Sons Inc, 1983.
4. MAXIQUIM. **Análise das Perspectivas de Mercado e Projeções de Preço no Longo Prazo de Metanol no Brasil**. Rio de Janeiro: Maxiquim, 2008.
5. METHANEX; Disponível em: <<http://www.methanex.com>>. Acesso em: março de 2011.
6. MILONE, G., ANGELINI, F. 1ª ed. **Estatística Aplicada**. São Paulo: Atlas, 1995.
7. MORETTIN, P.A. 1ª ed. **Econometria Financeira: um curso em séries temporais financeiras**. São Paulo: Blucher, 2008.
8. MORETTIN, P.A., TOLOI, C.M.C. 2ª ed. **Análise de Séries Temporais**. São Paulo: Blucher, 2006.
9. STENGEL, R. F. 1ª ed. **Stochastic Optimal Control: Theory and Application**. New York: John Wiley & Sons Inc, 1986.