XVII SEMEAD
Seminários em Administração
outubro de 2014
ISSN 2177-3866

PREVISÃO DE PREÇOS DE AÇÕES POR MEIO DE MODELOS DE REGRESSÃO COM DADOS EM PAINEL APLICADOS A DADOS DE ALTA FREQUÊNCIA

PEDRO HENRIQUE FILIPINI DOS SANTOS
Universidade Federal de São Paulo - UNIFESP
pedrofilipini@hotmail.com

ROSÂNGELA TOLEDO KULCSAR
Universidade Federal de São Paulo - UNIFESP
rosangela.kulcsar@unifesp.br

MAURI APARECIDO DE OLIVEIRA
USP - Universidade de São Paulo
mauriao@usp.br

À Prof.ª Dr.ª Rosângela Toledo Kulcsar, minha orientadora, por todo o apoio no desenvolvimento deste trabalho.
Ao Prof. Dr. Mauri Aparecido de Oliveira, pelo incentivo, pelos ensinamentos e pelas reuniões que tanto ajudaram no desenvolvimento deste trabalho.
Aos professores da EPPEN – UNIFESP que tanto me ensinam durante minha graduação.
À CNPq pelo financiamento, sem o qual esse trabalho jamais teria acontecido.
Aos amigos Andreas, Eliandra, Felipe e Wesley, que tanto me ensinaram e me ajudaram no período de desenvolvimento deste trabalho.
À minha família pelo apoio incondicional, sem o qual eu jamais chegaria onde estou hoje.

ÁREA TEMÁTICA: FINANÇAS

PREVISÃO DE PREÇOS DE AÇÕES POR MEIO DE MODELOS DE REGRESSÃO COM DADOS EM PAINEL APLICADOS A DADOS DE ALTA FREQUÊNCIA

RESUMO

Nos últimos anos, o número de negociações ligadas às atividades da área financeira, especialmente em ações e mercados monetários, tem crescido consideravelmente. Mais pessoas estão envolvidas em negociações com instrumentos financeiros pelo desenvolvimento dos sistemas de informação, especialmente na modalidade de comércio eletrônico.

O objetivo deste trabalho é realizar previsões de algumas das ações mais negociadas na BM&FBOVESPA utilizando modelos de regressão com dados em painel. Todas as ações escolhidas foram PN por possuírem maior liquidez em relação às ON. Foi selecionada uma série de cada ação, as quais foram obtidas através da BM&FBOVESPA. Cada série foi analisada num período de vinte dias úteis, sendo que cada série continha todas as negociações realizadas no dia, o valor da ação negociada e o volume de ações negociadas naquele preço. Os modelos serão estimados a partir de 75% da amostra e as previsões serão feitas para os 25% restantes. Por fim, utilizando os critérios RMSE e TIC, os resultados das previsões serão comparados com outros modelos de previsão, com o intuito de descobrir se o modelo de regressão com dados em painel é eficiente em relação aos outros modelos apresentados.

Palavras-chave: Dados em Painel, Econometria, Previsão.

ABSTRACT

In recent years, the number of negotiations concerning the activities of the financial sector, especially in stocks and money markets, have grown considerably. More people are involved in negotiations with financial instruments due to the development of information systems, especially in the form of e-commerce.

The objective of this work is to make predictions of some of the most traded shares on the BM & FBOVESPA by using regression models with panel data. All actions were chosen PN because of the greater liquidity in relation to ON. One set of each share, which were obtained through the BM & FBOVESPA was selected. Each series was analyzed over a period of twenty days, with each set containing all trades carried out on the value of shares traded and the volume of shares traded at that price. The models are estimated from 75% of the sample and the predictions will be made for the remaining 25%. Finally, using the RMSE and TIC criteria, the results of the predictions are compared with other prediction models, in order to discover whether the regression model with panel data is efficient compared to the other models presented.

Keywords: Panel Data, Econometrics, Forecast.

1 INTRODUÇÃO

Com o desenvolvimento dos sistemas de informação, mais pessoas estão envolvidas em negociações com instrumentos financeiros, especialmente na modalidade de comércio eletrônico. Os mercados financeiros têm também se tornado uma fonte de grande quantidade de dados que exigem análise cada vez mais rápida. Para a compreensão e a previsão futura da evolução do mercado, a utilização e construção de modelos e *softwares* econométricos têm demandado maior atenção e investimentos, tanto do setor acadêmico, quanto das empresas financeiras. Com a alta frequência dos dados existem agora formas de responder algumas das questões com relação a, por exemplo, volatilidade dos preços dos ativos, mas também surgem novas questões.

No Brasil, a BM&FBOVESPA oferece o Sistema de Acesso Direto ao Mercado com uma opção de conexão chamada de *co-location*. Especificamente, a *co-location* destina-se a aplicações de negociação em alta-frequência. Na negociação via *co-location*, as ordens de compra e venda do cliente são geradas por programas de computador (*Automated Trading System* - ATS), instalados em servidores hospedados no centro de processamento de dados da BM&FBOVESPA. Vale ressaltar que o equipamento hospedado na Bolsa pode ser acessado remotamente pelo participante para fins de monitoração, configuração de parâmetros, manutenção e etc.

Os mercados financeiros são a fonte dos principais dados de série temporais utilizados em econometria. A forma original de preços de mercado é de dados *tick-by-tick*: cada "*tick*" é uma unidade lógica de informação, como uma cota ou um preço de transação. Por natureza, estes dados são irregularmente espaçados no tempo. Mercados líquidos geram centenas ou milhares de *ticks* por dia útil.

Compreendendo a importância dos dados fornecidos pelos mercados financeiros, torna-se relevante a criação de novas estratégias através de sua análise. Assim, são procurados novos meios de se verificar quais ativos estão propensos a gerar retornos aos investidores, para que esses possuam um cenário mais embasado em seu processo de tomada de decisão, principalmente através de técnicas estatísticas e econométricas, utilizando-se de modelos de previsão.

O objetivo deste trabalho é realizar previsões de séries de ações de algumas das empresas mais negociadas na BM&FBOVESPA, sendo elas o Banco Bradesco (BBDC4), o Banco Itaú (ITSA4), a Petrobrás (PETR4) e a Vale (VALE5), utilizando modelos de regressão com dados em painel. Todas as ações escolhidas foram PN por possuírem maior liquidez em relação às ON. Foi selecionada uma série de cada empresa, as quais foram obtidas através da BM&FBOVESPA. Cada série foi analisada num período de vinte dias úteis, entre 6 de Agosto de 2012 e 31 de Agosto de 2012, sendo que cada série continha todas as negociações realizadas no dia, o valor da ação negociada e o volume de ações negociadas naquele preço.

Tendo em vista que o objetivo é estimar modelos através de um painel equilibrado e com dados igualmente espaçados, os dados foram tratados. A amostra utilizada possuía o preço da ação ao ser negociada, além do horário de sua negociação e o volume negociado naquele valor de vendas à vista. Considerando que foram observadas as negociações que estavam registradas entre as 10:00:00h e as 16:55:00h, que são respectivamente o horário de início e horário de fechamento das negociações na Bolsa, foi feita uma média do preço de cada ação a cada período de cinco minutos, além da soma do volume negociado no mesmo período, totalizando uma amostra igualmente espaçada de 83 valores de preços e volumes para cada uma das ações em cada um dos dias observados.

Os modelos foram estimados a partir de 75% da amostra e as previsões foram feitas para os 25% restantes. Por fim, utilizando os critérios RMSE e TIC, os resultados das

previsões foram comparados com outros modelos de previsão, com o intuito de descobrir se o modelo de regressão com dados em painel é eficiente em relação aos outros modelos apresentados. Assim, além dos modelos de regressão com dados em painel, foram estimados quatro modelos ARIMA-GARCH e quatro modelos NAIVE, sendo um deles para cada uma das ações observadas.

2 REVISÃO DE LITERATURA

2.1 Tipos de Dados

Conforme salienta Gujarati (2004, p. 25) existem três tipos de dados que podem ser avaliados empiricamente: séries temporais, dados em corte transversal e dados combinados. Os dados em séries temporais são um conjunto de observações dos valores que uma variável assume ao longo do tempo. Dados em corte transversal são dados em que uma ou mais variáveis foram coletados em um mesmo ponto no tempo. Já os dados combinados possuem características tanto de séries temporais quanto de corte transversal. Os dados combinados, ou dados em painel, serão o principal foco desse estudo.

2.2 Dados em Painel

Segundo Gujarati (2004, p. 513), os dados em painel, também conhecidos como dados combinados, são compostos por uma mesma unidade de corte transversal, a qual é acompanhada ao longo do tempo, o que significa que tais dados possuem uma dimensão espacial e uma dimensão temporal. Embora existam algumas variações sutis entre os tipos de dados de painel, essencialmente, todos tratam basicamente do movimento no tempo de unidades de corte transversal. Os modelos de regressão embasados nesses termos são conhecidos como modelos de regressão com dados em painel.

Algumas das vantagens dos dados em painel incluem: o controle individual da heterogeneidade; a geração de dados mais informativos, além de maior variabilidade e menor colinearidade entre variáveis devido à combinação entre séries temporais e dados de corte transversal, gerando maior eficiência e número de graus de liberdade; a facilidade para análise das dinâmicas de ajuste; a identificação de efeitos que não são detectados em dados que são compostos somente por corte transversal ou séries temporais; permite a construção de modelos comportamentais mais complexos; o viés resultante da agregação de empresas ou indivíduos pode ser reduzido ou eliminado; testes de raiz unitária possuem distribuição assintótica padrão (BALTAGI, 2005, p. 4-7).

Entretanto, os dados em painel também possuem suas limitações: problemas no formato e coleta de dados; distorção da medição dos erros; problemas de seletividade; dimensão de séries temporais curtas; dependência de cortes transversais (BALTAGI, 2005, p. 7-9).

Os dados em painel também podem ser divididos entre painéis equilibrados e desequilibrados. No caso, serão tratados apenas painéis equilibrados. Gujarati (2004, p. 516) define um painel equilibrado como um painel onde cada unidade de corte transversal possui o mesmo número de observações em séries temporais.

No mais, serão abordados dois tipos de modelos de dados em painel, sendo o primeiro o modelo de efeitos fixos e o segundo o modelo de efeitos aleatórios.

2.2.1 Modelo de Efeitos Fixos

O modelo de efeitos fixos recebe esse nome devido ao fato de que, embora o intercepto possa diferir entre indivíduos, cada intercepto individual não se altera ao longo do tempo.

É recomendável que se tome certa cautela quanto aos modelos de efeitos fixos. Gujarati (2004, p. 521-522) expressa alguns problemas frequentes, como a perda de graus de liberdade quando existe a inclusão de muitas variáveis *dummy*, a existência de multicolinearidade, a presença de variáveis que não mudam ao longo do tempo e a necessidade de verificar se o termo de erro segue uma distribuição normal.

2.2.2 Modelo de Efeitos Aleatórios

A abordagem sugerida pelo modelo de efeitos aleatórios é a de que a inclusão de variáveis binárias no modelo de efeitos fixos representam uma falta de conhecimento com relação ao modelo que seria considerado perfeito, sugerindo, assim, que deixamos de incluir variáveis relevantes no modelo. Para compensar tal problema, o modelo de efeitos aleatórios expressa essa falta de conhecimento através dos termos de erro (GUJARATI, 2004, p. 521-523). Vale lembrar que o termo de erro não é diretamente observável.

2.3 Processos Estocásticos Estacionários

Tal processo estacionário também é conhecido como 'fracamente estacionário', ou 'estacionário em covariâncias', ou 'estacionário de segunda ordem', ou 'processo estocástico em sentido amplo'. Uma série é estacionária no sentido estrito quando todos os momentos de sua distribuição de probabilidade não variam ao longo do tempo. Entretanto, se existir distribuição normal, o processo estocástico fracamente estacionário é estritamente estacionário, visto que o processo estocástico normal é plenamente definido por seus dois momentos, a média e a variância (GUJARATI, 2004, p. 639).

Segundo Gujarati (2004, p.639) a estacionariedade fraca possui as seguintes propriedades:

Média: $E(Y_t) = \mu$

Variância: $\text{var}(Y_t) = E(Y_t - \mu)^2 = \sigma^2$

Covariância: $\gamma_k = E[(Y_t - \mu)(Y_{t+k} - \mu)]$

2.4 Raiz Unitária

Koop (2000, p. 147) trata do conceito de raiz unitária através do exemplo de um AR(*p*), tal como:

$$Y_t = \theta_1 Y_{t-1} + ... + \theta_p Y_{t-p} + \varepsilon_t .$$

Ao se tratar de raízes unitárias, convém que se subtraia Y_{t-1} do modelo de ambos os lados, de modo que através de algumas simplificações, podemos obter:

$$\Delta Y_t = \rho Y_{t-1} + \gamma_1 \Delta Y_{t-1} + ... + \gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t .$$

De forma que $\rho, \gamma_1, ..., \gamma_{p-1}$ são funções simples de $\theta_1, ..., \theta_p$. Por exemplo, $\rho = \theta_1 + ... + \theta_p - 1$. Sendo assim, o modelo é idêntico ao modelo AR(*p*), de modo que a equação ainda está na forma de um modelo de regressão de forma que se $\rho = 0$, então a série

temporal AR(*p*) possui uma raiz unitária e se $-2 < \rho < 0$, então a série é estacionária. Assim, podemos dizer que a não-estacionariedade e a presença de raiz unitária podem ter seu significado tido como o mesmo. É possível remover raízes unitárias através da diferenciação, de modo que apenas a partir da primeira diferença estacionária de ΔY que a série é apropriada para ser devidamente analisada, caso contrário, sua variância irá para o infinito à medida que o tempo passa.

2.5 Teste Dickey-Fuller

Conforme explica Sartoris (2003, p. 368), o teste Dickey-Fuller (DF) é utilizada para a identificação da existência de uma raiz unitária em um AR(1). O teste consiste basicamente em testar a hipótese nula de que $\rho = 0$, que caso seja rejeitada, significa que a série temporal não possui raiz unitária. Ele é feito computando-se como se fosse uma estatística *t* em um teste comum numa regressão qualquer, entretanto, os valores para o teste DF são diferentes, visto que chegaram a valores-limites por meio de simulações, dessa forma, criando a estatística τ. O teste DF é aplicado no seguinte modelo, que conforme demonstrado anteriormente, equivale a um AR(1), neste caso, com intercepto e tendência determinística:

$$\Delta Y_t = \alpha + \beta t + \rho Y_{t-1} + \varepsilon_t .$$

Para um processo AR(*p*), utiliza-se o chamado teste Dickey-Fuller Aumentado (ADF), que é dado pela expressão abaixo, neste caso, com intercepto e tendência determinística:

$$\Delta Y_t = \alpha + \beta t + \rho Y_{t-1} + \sum_{i=2}^{p} \sigma_i \Delta Y_{t-i+1} + \varepsilon_t .$$

Vale lembrar que a série temporal pode apresentar mais de uma raiz unitária, ou seja, é necessária mais de uma diferença para que ela se torne estacionária.

2.6 Processos Autorregressivos de Médias Móveis – ARMA(*p*,*q*)

É uma combinação dos processos autorregressivos de ordem *p*, onde os modelos são explicados através dos valores defasados de *Y* mais um choque aleatório ε_t, com os processos de médias móveis de ordem *q*, onde os modelos são explicados através da combinação do choque presente com choques passados, de forma que um ARMA(*p*,*q*) pode ser demonstrado como:

$$Y_t = \theta_1 Y_{t-1} + ... + \theta_p Y_{t-p} + \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \varphi_2 \varepsilon_{t-2} - ... - \varphi_q \varepsilon_{t-q} .$$

De forma que:

$$Y_t - \theta_1 Y_{t-1} + ... - \theta_p Y_{t-p} = \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \varphi_2 \varepsilon_{t-2} - ... - \varphi_q \varepsilon_{t-q} .$$

Assim, a forma simplificada de um ARMA(*p*,*q*), sendo B um operador retroativo, pode ser dada por:

$$\Theta_p(B) Y_t = \Phi_q(B) \varepsilon_t .$$

Onde:

$$\Theta_p(B) \equiv 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_p B^p .$$

$$\Phi_q(B) \equiv 1 + \varphi_1 B - \varphi_2 B^2 - ... - \varphi_q B^q .$$

De forma que o processo sempre será estacionário se as raízes de $\Phi_p(B) = 0$ forem, em módulo, maiores que um, e sempre será invertível se as raízes de $\Theta_p(B) = 0$ forem, em módulo, maiores que um, conforme salienta Oliveira (2007, p. 12).

2.7 Processos Autorregressivos Integrados de Médias Móveis – ARIMA(*p*,*d*,*q*)

Sartoris (2003, p. 349) diz que algo que deve ser notado é que os processos apresentados anteriormente podem ser não-estacionários, neste caso, podemos definir uma variável Z_t como sendo a primeira diferença de Y_t, chamada integrada de ordem 1, ou I(1), tal como no exemplo:

$$Z_t = Y_t - Y_{t-1} = \Delta Y_t .$$

Caso a primeira diferença não seja o suficiente para tornar o processo estacionário, tomamos *d* diferenças de Y_t até que o processo torne-se estacionário, chamadas de I(*d*), assim:

$$Z_t = \Delta^d Y_t .$$

De acordo com Sartoris (2003, p. 349) um processo ARIMA(*p*,*d*,*q*) refere-se a um Y_t integrado de ordem *d*, e sua *d*-ésima diferença segue um processo combinado autorregressivo (de ordem *p*) e de médias móveis (de ordem *q*). Logo, o processo para Y_t será dado por:

$$\Delta^d Y_t = \theta_1 \Delta^d Y_{t-1} + \theta_2 \Delta^d Y_{t-2} + ... + \theta_p \Delta^d Y_{t-p} + \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \varphi_2 \varepsilon_{t-2} - ... - \varphi_q \varepsilon_{t-q} .$$

2.8 Modelo Autorregressivo de Heterocedasticidade Condicional – ARCH

O modelo ARCH foi inicialmente proposto por Engle (1982). Segundo Gujarati (2004), a presença de uma alta volatilidade sugere que a variância da série temporal varia ao longo do tempo, ou seja, alta volatilidade sugere a existência de heterocedasticidade, de forma que essa heterocedasticidade apresentada em diferentes períodos também pode estar correlacionada, podendo ser modelada através do ARCH.

Um modelo ARCH pode ser demonstrado da seguinte forma (OLIVEIRA, 2004, p. 15):

$$\varepsilon_t = \eta_t \sqrt{h_t} .$$

Sendo:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + ... + \alpha_r \varepsilon_{t-r}^2 .$$

Em que η_t é i.i.d. (0,1), $\alpha_0 > 0, \alpha_i \geq 0, i > 0$ e $\eta_t \sim N(0,1)$.

Possuindo as seguintes propriedades:

(i) ε_t tem média zero

(ii) ε_t tem variância condicional dada por $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$.

(iii) ε_t tem variância incondicional dada por $\sigma^2 = \dfrac{\alpha_0}{1-\alpha_1}$.

(iv) ε_t tem auto-covariâncias zero.

2.9 Modelo Autorregressivo de Heterocedasticidade Condicional Generalizado – GARCH (*r*,*s*)

Criado por Bollerslev (1986) é uma variação do modelo ARCH criado por Engle (1982). O GARCH diz que a variância condicional dos erros, em certo momento, depende não apenas do termo dos erros quadrados do período anterior, mas também da variância condicional do período anterior.

Um modelo GARCH(*r*,*s*) pode ser definido da seguinte forma (OLIVEIRA, 2007, p.17):

$$\varepsilon_t = \eta_t \sqrt{h_t}$$

Sendo:

$$h_t = \alpha_0 + \sum_{i=1}^{s} \alpha_1 \varepsilon_{t-1}^2 + \sum_{j=1}^{r} \beta_j h_{t-j}$$

Em que η_t segue uma distribuição t de Student ou normal e é i.i.d. (0,1), $\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0, \sum_{i=1}^{q} (\alpha_i + \beta_j), q = \max(r, s)$.

2.10 Critérios de Informação

São critérios utilizados para definir qual o modelo mais adequado. Serão tratados apenas dois critérios de informação, sendo o primeiro conhecido como Akaike (CIA) e o segundo conhecido como Schwarz (CIS), sendo que quanto menor o valor calculado, melhor o modelo (SARTORIS, 2003, p. 272). Os critérios de informação são dados pelas seguintes fórmulas:

$$CIA = 1 + \ln 2\pi + \ln \frac{\sum \hat{\mu}_i^2}{n} + \frac{2k}{n}.$$

$$CIS = 1 + \ln 2\pi + \ln \frac{\sum \hat{\mu}_i^2}{n} + \frac{k \ln n}{n}.$$

2.11 Teste Breusch-Godfrey (TESTE LM)

Conforme exemplifica Gujarati (2004), o teste de autocorrelação desenvolvido por Breusch e Godfrey, possui a seguinte hipótese nula:

$$H_0 : \theta_1 = \theta_2 = ... = \theta_\theta = 0.$$

Ou seja, a hipótese nula é que não existe correlação serial de qualquer ordem.
Dada a seguinte equação para exemplificação:

$$Y_t = \beta_1 + \beta_2 + X_t + \mu_t.$$

Sendo μ_t um AR(p), tal que:

$$\mu_t = \theta_1 \mu_{t-1} + \theta_2 \mu_{t-2} + ... + \theta_\theta \mu_{t-\theta} + \varepsilon_t.$$

Em seguida, calcula-se uma regressão auxiliar, tal como:

$$\hat{\mu}_t = \alpha_1 + \alpha_2 X_t + \hat{\theta}_1 \hat{\mu}_{t-1} + \hat{\theta}_2 \hat{\mu}_{t-2} + ... + \hat{\theta}_\theta \hat{\mu}_{t-\theta} + \varepsilon_t.$$

Na qual encontra-se o R^2. Caso a amostra seja grande, é dado que:

$$(n - p)R^2 \sim \chi_p^2.$$

De forma que se $(n - p)R^2$ exceder o valor crítico de χ_p^2, H_0 é rejeitada.

2.12 Tipos de Previsão

Gujarati (2004) explica sobre dois tipos de previsão. O primeiro tipo é a chamada previsão estática. Essa previsão é feita levando em consideração os valores que ocorreram no passado. O segundo tipo é a chamada previsão dinâmica. Essa previsão é feita levando em

consideração previsões dos valores passados, ou seja, ela leva em conta os erros cometidos nas previsões passadas.

2.13 Métodos para Avaliação de Previsões

Basicamente serão utilizados dois métodos de avaliação de previsão. O primeiro é chamado de Raiz do Erro Quadrado Médio (RMSE) e o segundo de Coeficiente de Desigualdade de Theil (TIC). Oliveira (2007, p. 20) salienta que quanto menor o valor do RMSE, menor é o erro obtido pelo modelo na realização da previsão e os valores de TIC variam entre 0 e 1, sendo que zero é um ajuste perfeito. O RMSE e o TIC são dados por:

$$RMSE = \sqrt{\sum_{t=T+1}^{T+h} \frac{(\hat{y}_t - y_t)^2}{h}} .$$

$$TIC = \frac{\sqrt{\sum_{t=T+1}^{T+h} \frac{(\hat{y}_t - y_t)^2}{h}}}{\sqrt{\sum_{t=T+1}^{T+h} \frac{\hat{y}_t^2}{h}} + \sqrt{\sum_{t=T+1}^{T+h} \frac{y_t^2}{h}}} .$$

2.14 Dados de Alta Frequência

Uma negociação de alta frequência é aquela que possui uma maior rotatividade de capital e rápidas respostas computadorizadas. Basicamente, o que diferencia as negociações de alta frequência das negociações de baixa frequência, é que a primeira acontece em questão de milionésimos de segundo, em que respostas rápidas podem vir a mudar as condições de mercado, de forma que ocorre um alto número de negociações com baixos ganhos médios (ALDRIDGE, 2009, p.1).

A partir da descrição feita pelo Banco Central do Brasil, pode-se dizer que uma negociação *intraday* nada mais é do que uma negociação de títulos de uma empresa realizada no mesmo dia, ou seja, a compra e a venda da mesma quantidade de títulos de uma empresa através da mesma corretora e do mesmo agente de compensação num mesmo dia.

3 METODOLOGIA

Com relação aos dados, os mesmos tiveram que passar por um tratamento, tendo em vista que foram observados vinte dias úteis, que compreenderam o período entre 6 de Agosto de 2012 e 31 de Agosto de 2012, contendo todas as negociações de cada uma das quatro ações observadas, sendo que cada dia possuía um número de negociações que variava entre cerca de 6 mil até 35 mil, dependendo do dia e da ação observada. As quatro ações observadas foram BBDC4, ITSA4, PETR4 e VALE5. Todas as ações escolhidas foram PN por possuírem maior liquidez em relação às ON e os dados foram obtidos através de um pedido à BM&FBOVESPA.

Tendo em vista que o objetivo é estimar modelos através de um painel equilibrado e com dados igualmente espaçados, os dados foram tratados. A amostra utilizada possuía o preço da ação ao ser negociada, além do horário de sua negociação e o volume negociado naquele valor de vendas à vista. Tendo em vista que foram observadas as negociações que estavam registradas entre as 10:00:00h e as 16:55:00h, que são respectivamente o horário de

início e horário de fechamento das negociações na Bolsa, foi feita uma média do preço de cada ação a cada período de cinco minutos, além da soma do volume negociado no mesmo período, totalizando uma amostra igualmente espaçada de 83 valores de preços e volumes para cada uma das ações em cada um dos dias observados.

Entretanto, foram necessários dois ajustes. O primeiro foi realizado tendo em vista que em sete das oitenta amostras, sendo cada amostra tida como um dos vinte dias de cada uma das quatro ações, não existiam registros de negociações anteriores a 10:05:00h, sendo todas elas lançadas em um único momento após as 10:05:00h. Tal comportamento ocorreu pois tais valores eram provenientes do leilão de abertura, que ocorre das 09:45:00h até as 10:00:00h, podendo ser prorrogado. Nessas sete amostras, o leilão tal lançamento foi feito após as 10:05:00h, fazendo com que o primeiro valor da amostra fosse dado como vazio. Para ajustar esse problema, a média do preço entre as 10:00:00h e as 10:05:00h foi tomada como a mesma média entre 10:05:00h e 10:10:00h, visto que as negociações que deveriam estar presentes no primeiro valor da amostra, estavam presentes no segundo, enquanto que o volume total entre as 10:05:00h e as 10:10:00h foi dividido pela metade, de forma que metade transformou-se na primeira soma de volume negociado da amostra e a segunda metade transformou-se na segunda soma de volume negociado na amostra.

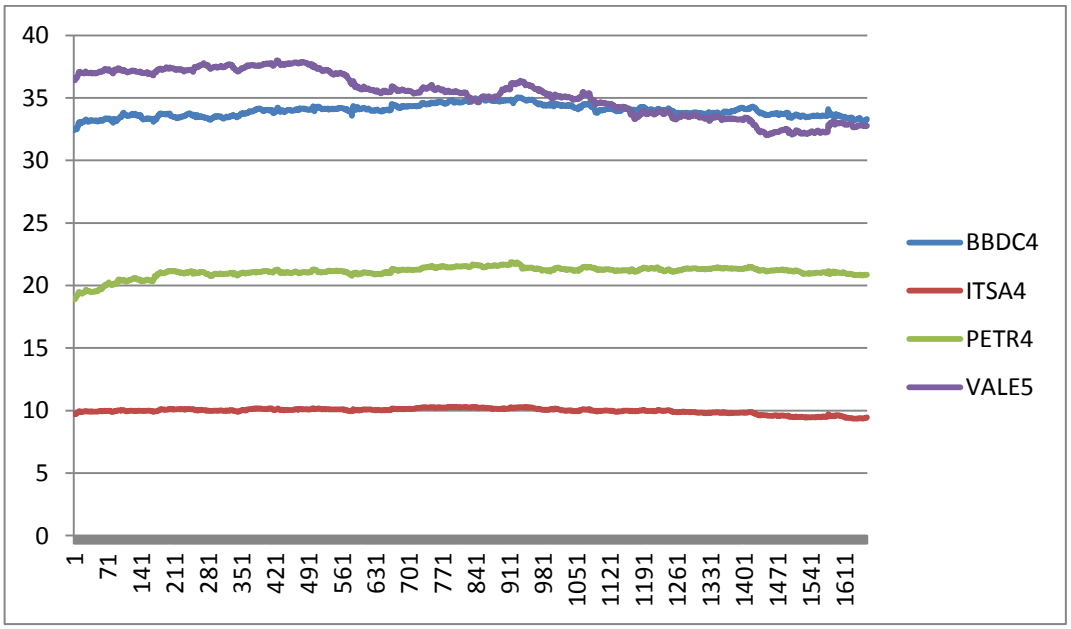
O segundo ajuste foi feito através da retirada de um elemento da amostra, tendo em vista que nenhuma das amostras possuía negociações entre 16:50:00h e 16:55:00h do dia 31 de Agosto de 2012. Assim, tal elemento foi retirado da amostra, que em seu total possui 1659 valores para cada uma das quatro ações, totalizando 6636 valores.

Por fim, os modelos foram utilizados usando 1245 valores de cada ação, compreendendo cerca de 75% da amostra. Os outros 414 valores, que equivalem a cerca de 25% da amostra, foram utilizados para previsão. Foram feitas apenas previsões estáticas em todos os modelos, sendo que a variável que está sendo prevista é o preço da ação.

4 RESULTADOS

4.1 Análise da Estatística Descritiva

FIGURA 1: Média do Preço das Ações em Períodos de 5 Minutos



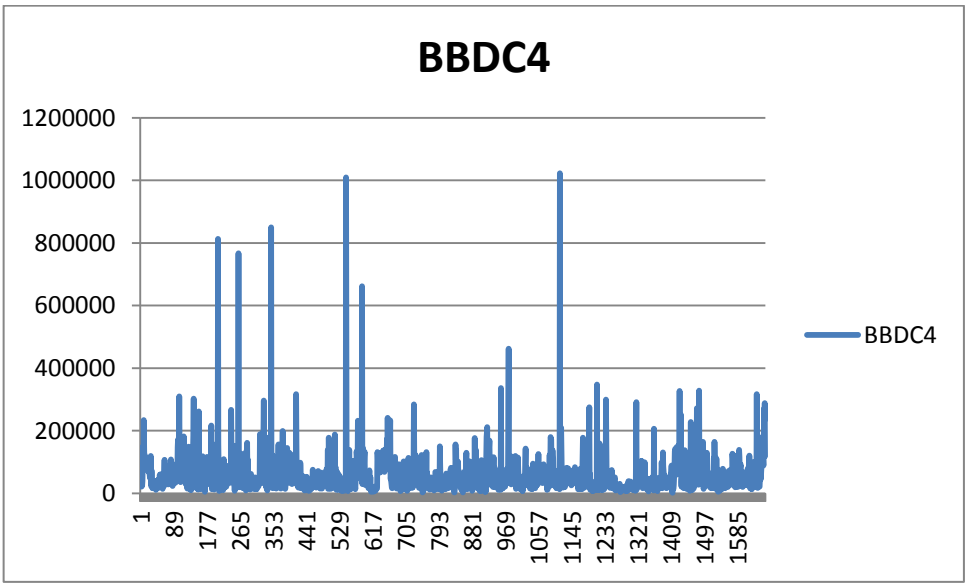
Fonte: Elaboração Própria

TABELA 1: Estatística Descritiva da Média do Preço das Ações em Períodos de 5 Minutos

Ação	Média	Mínimo	Máximo	Desvio Padrão
BBDC4	33,96	32,42	35,04	0,47
ITSA4	9,97	9,34	10,31	0,22
PETR4	21,09	18,89	21,87	0,44
VALE5	35,34	32,00	37,99	1,79

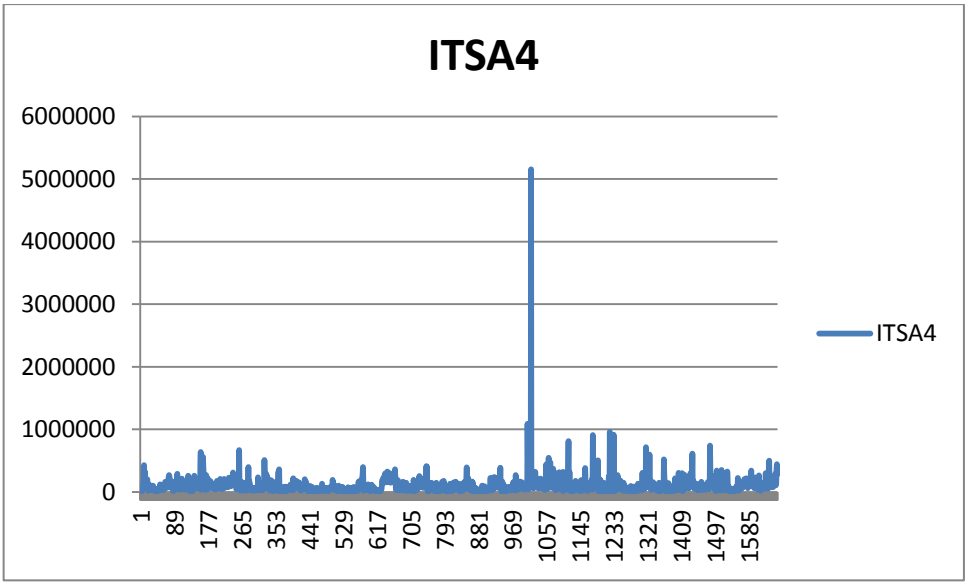
Fonte: Elaboração Própria

FIGURA 2: Soma do Volume da Ação BBDC4 em Períodos de 5 Minutos



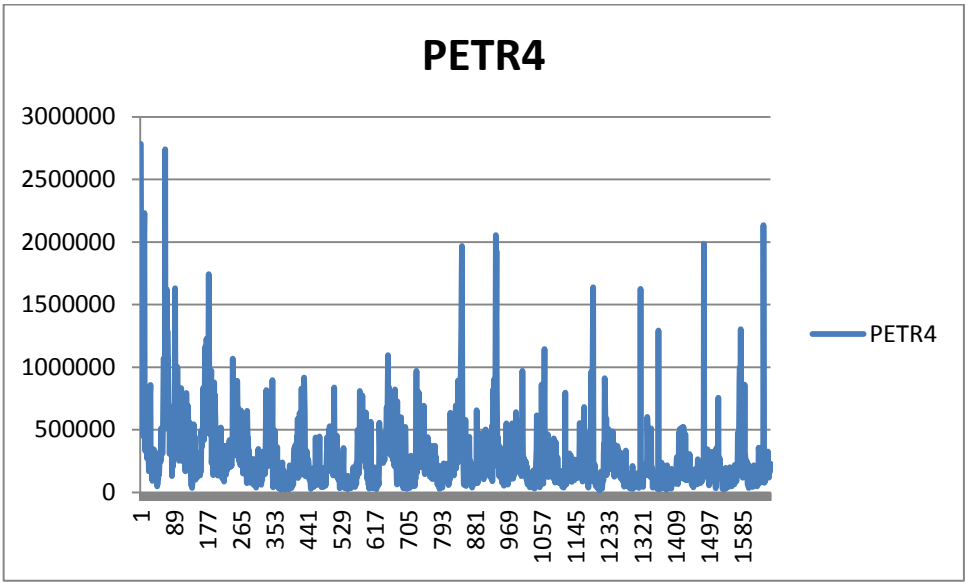
Fonte: Elaboração Própria

FIGURA 3: Soma do Volume da Ação ITSA4 em Períodos de 5 Minutos



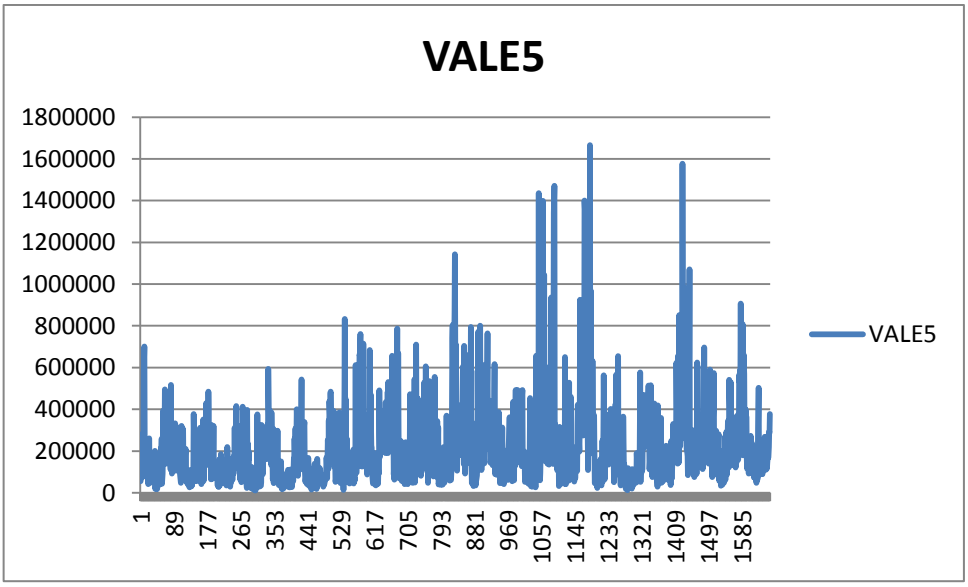
Fonte: Elaboração Própria

FIGURA 4: Soma do Volume da Ação PETR4 em Períodos de 5 Minutos



Fonte: Elaboração Própria

FIGURA 5: Soma do Volume da Ação VALE5 em Períodos de 5 Minutos



Fonte: Elaboração Própria

TABELA 2: Estatística Descritiva da Soma do Volume das Ações em Períodos de 5 Minutos

Ação	Média	Mínimo	Máximo	Desvio Padrão
BBDC4	56361,12	2200	1023000	68381,66
ITSA4	99823,69	2100	5152300	158982,21
PETR4	271956,60	15800	2783050	284020,40
VALE5	208807,17	10100	1665500	187062,43

Fonte: Elaboração Própria

4.2 Estimativas dos Modelos

4.2.1 Modelos de Regressão com Dados em Painel

Conforme salienta Gujarati (2004, p. 524), quando o número de dados de séries temporais é grande e o número de unidades de corte transversal é pequeno, a diferença entre os valores estimados por meio do modelo de efeitos fixos e de efeitos aleatórios será pequena, de forma que a escolha se baseará na conveniência computacional.

Foram testadas diversas combinações de modelos utilizando até três defasagens em efeitos fixos e aleatórios. O modelo conveniente que veio a ser aceito foi o modelo de efeitos aleatórios defasados em três momentos em ambas as variáveis. Tal comportamento pode ser justificado pelo tempo para tomada de decisão do investidor.

O modelo estimado foi o seguinte:

TABELA 3: Estimação do Modelo de Regressão com Dados em Painel de Efeitos Aleatórios

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.013472	0.010885	1.237726	0.2159
VALUE(-3)	0.999438	0.000392	2550.223	0.0000
VOLUME(-3)	1.11E-08	5.67E-09	1.964503	0.0495
Effects Specification				
			S.D.	Rho
Cross-section random			0.008190	0.0102
Idiosyncratic random			0.080580	0.9898
Weighted Statistics				
R-squared	0.999236	Mean dependent var		6.812460
Adjusted R-squared	0.999236	S.D. dependent var		2.920421
S.E. of regression	0.080718	Sum squared resid		32.34869
F-statistic	3248522.	Durbin-Watson stat		0.632862
Prob(F-statistic)	0.000000			
Unweighted Statistics				
R-squared	0.999941	Mean dependent var		25.33579
Sum squared resid	32.52530	Durbin-Watson stat		0.629426
Random Effects				
BBDC4	0.008475			
ITSA4	-0.007596			
PETR4	0.000564			
VALE5	-0.001443			

Fonte: Adaptado do EViews®

4.2.2 Modelos de Regressão ARIMA-GARCH

Foram estimados quatro modelos ARIMA-GARCH, sendo um deles para cada uma das ações observadas. Para que os modelos ARIMA fossem estimados com maior precisão, foi feita a primeira diferença logarítmica de cada uma das séries, garantindo, assim, a sua estacionariedade.

Já para a estimação de um modelo GARCH, é preciso realizar o Teste LM. Infelizmente, apenas a BBDC4 ARMA(3,3) apresentou um p-valor abaixo de 5%, sendo assim, o único modelo estimado por GARCH.

Os modelos foram escolhidos através, primeiramente, dos critérios de informação Akaike e Schwarz, sendo utilizado o critério da parcimônia para o caso de modelos muito próximos. Os modelos escolhidos foram:

TABELA 4: Modelo Estimado para a Ação BBDC4 – ARMA(3,3) GARCH(3,1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	2.60E-05	2.57E-05	1.013078	0.3110
AR(3)	0.713892	0.107287	6.654037	0.0000
MA(3)	-0.772078	0.098494	-7.838805	0.0000
Variance Equation				
C	1.02E-06	7.94E-08	12.78582	0.0000
RESID(-1)^2	0.324160	0.035996	9.005351	0.0000
RESID(-2)^2	0.110023	0.033577	3.276701	0.0011
RESID(-3)^2	0.297611	0.038295	7.771440	0.0000
GARCH(-1)	-0.011538	0.048781	-0.236523	0.8130
R-squared	0.012202	Mean dependent var		3.85E-05
Adjusted R-squared	0.010606	S.D. dependent var		0.001570
S.E. of regression	0.001561	Akaike info criterion		-10.27808
Sum squared resid	0.003018	Schwarz criterion		-10.24505
Log likelihood	6385.546	Hannan-Quinn criter.		-10.26565
Durbin-Watson stat	1.851990			
Inverted AR Roots	.89	-.45-.77i	-.45+.77i	
Inverted MA Roots	.92	-.46-.79i	-.46+.79i	

Fonte: Adaptado do EViews®

TABELA 5: Modelo Estimado para a Ação ITSA4 – ARMA(3,1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.41E-05	4.77E-05	0.505368	0.6134
AR(3)	-0.071999	0.028324	-2.542010	0.0111
MA(1)	0.176076	0.027982	6.292427	0.0000
R-squared	0.032944	Mean dependent var		2.42E-05
Adjusted R-squared	0.031382	S.D. dependent var		0.001558
S.E. of regression	0.001533	Akaike info criterion		-10.12081
Sum squared resid	0.002909	Schwarz criterion		-10.10843
Log likelihood	6282.964	Hannan-Quinn criter.		-10.11615
F-statistic	21.08727	Durbin-Watson stat		2.019029
Prob(F-statistic)	0.000000			
Inverted AR Roots	.21+.36i	.21-.36i	-.42	
Inverted MA Roots	-.18			

Fonte: Adaptado do EViews®

TABELA 6: Modelo Estimado para a Ação PETR4 – ARMA(3,1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.31E-05	5.20E-05	1.789741	0.0737
AR(3)	-0.069999	0.028236	-2.479052	0.0133
MA(1)	0.206835	0.027830	7.432122	0.0000
R-squared	0.045707	Mean dependent var		9.31E-05
Adjusted R-squared	0.044165	S.D. dependent var		0.001662
S.E. of regression	0.001625	Akaike info criterion		-10.00420
Sum squared resid	0.003269	Schwarz criterion		-9.991816
Log likelihood	6210.607	Hannan-Quinn criter.		-9.999544
F-statistic	29.64772	Durbin-Watson stat		2.000363
Prob(F-statistic)	0.000000			
Inverted AR Roots	.21+.36i	.21-.36i	-.41	
Inverted MA Roots	-.21			

Fonte: Adaptado do EViews®

TABELA 7: Modelo Estimado para a Ação VALE5 – ARMA(1,2)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5.88E-05	5.04E-05	-1.166120	0.2438
AR(1)	0.161588	0.027936	5.784193	0.0000
MA(2)	-0.064046	0.001324	-48.37236	0.0000
R-squared	0.027907	Mean dependent var		-5.73E-05
Adjusted R-squared	0.026339	S.D. dependent var		0.001613
S.E. of regression	0.001592	Akaike info criterion		-10.04539
Sum squared resid	0.003142	Schwarz criterion		-10.03302
Log likelihood	6246.208	Hannan-Quinn criter.		-10.04073
F-statistic	17.79927	Durbin-Watson stat		1.993629
Prob(F-statistic)	0.000000			
Inverted AR Roots	.16			
Inverted MA Roots	.25	-.25		

Fonte: Adaptado do EViews®

4.3 Avaliação da Capacidade Preditiva dos Modelos

Após todas as previsões serem realizadas e transformadas de volta no nível (no caso das previsões feitas através dos modelos ARIMA-GARCH), para avaliar sua capacidade preditiva serão usados dois métodos. A Raiz do Erro Quadrado Médio, conhecido como RMSE, e o Coeficiente de Desigualdade de Theil, conhecido como TIC. Em ambos os casos, quando mais próximos de zero, mais próximos da realidade estão os modelos. Em negrito estão os melhores modelos, sendo que o modelo NAIVE, o qual prevê o valor atual como sendo o valor imediatamente anterior, foi utilizado como comparação, de forma que, se o NAIVE superar o modelo em questão, o modelo é considerado ineficiente.

TABELA 8: Avaliação da Capacidade Preditiva dos Modelos Estimados por meio do RMSE

RMSE				
Modelo	BBDC4	ITSA4	PETR4	VALE5
ARIMA-GARCH	0,05516	0,01808	0,02587	**0,04978**
PAINEL	0,09128	0,03086	0,04468	0,10565
NAIVE	**0,05488**	**0,01786**	**0,02562**	0,05898

Fonte: Elaboração Própria

TABELA 9: Avaliação da Capacidade Preditiva dos Modelos Estimados por meio do TIC

TIC				
Modelo	BBDC4	ITSA4	PETR4	VALE5
ARIMA-GARCH	0,00082	0,00094	0,00061	**0,00076**
PAINEL	0,00135	0,00160	0,00105	0,00161
NAIVE	**0,00081**	**0,00092**	**0,00060**	0,00090

Fonte: Elaboração Própria

Como pode ser visto, o modelo em painel foi considerado o modelo mais ineficiente em todos os casos, sendo seguido pelo ARIMA-GARCH. De modo geral, o NAIVE conseguiu fazer previsões mais eficientes do que todos os modelos, com exceção da previsão para a ação VALE5, onde o modelo ARMA(1,2) se mostrou como modelo mais eficiente.

5 CONCLUSÃO

Os dados de previsão das ações BBDC4, ITSA4, PETR4 e VALE5 foram analisados com os mesmos critérios para todos os modelos em questão. Apenas um dos modelos de previsão, o qual previa o preço da ação VALE5 através de um ARMA(1,2), se mostrou satisfatório. Para as outras ações, o melhor modelo de previsão foi o NAIVE. Portanto, nenhuma outra previsão foi considerada eficiente em comparação ao método de previsão NAIVE, o qual prevê o valor atual como sendo o valor imediatamente anterior.

Dessa forma, conclui-se que para esta amostra de série temporal, não poderão ser feitas previsões precisas através dos modelos de regressão com dados em painel, pois estes não obtiveram os valores esperados no critério TIC ou RMSE em nenhuma das amostras, sendo assim, com exceção do modelo ARMA(1,2) para a ação VALE5, nenhum dos modelos atingiu as expectativas.

Por fim, é necessária uma análise mais aprofundada da série, através de diferentes modelos ou de uma amostra diferente, de forma que a série possa ser compreendida de maneira mais satisfatória, mas, inicialmente, provou-se que os modelos de regressão com dados em painel não são eficientes para a previsão do preço de ações e que a utilização do volume negociado como variável independente não causou uma melhora visível no modelo, uma vez que, em todos os casos, as previsões feitas a partir dos modelos de regressão com dados em painel foram as mais ineficientes.

REFERÊNCIAS BIBLIOGRÁFICAS

ALDRIDGE, Irene. ***High-Frequency Trading****: A Practical Guide to Algorithmic Strategies and Trading Systems.* New York: Wiley, 2009.

BALTAGI, Badi. ***Econometric Analysis of Panel Data***. Chichester: John Wiley & Sons, 2005, p. 4-9, 3.a ed.

BANCO CENTRAL DO BRASIL. Disponível em:< http://www.bcb.gov.br.>. Acesso em: 02 abr. 2014.

BM&FBOVESPA. **Horários de Negociação no Mercado de Ações**. Disponível em: <http://www.bmfbovespa.com.br/pt-br/regulacao/horarios-de-negociacao/acoes.aspx?Idioma=pt-br>. Acesso em: 02 abr. 2014.

BOLLERSLEV, T., ***Generalized Autoregressive Conditional Heteroskedasticity***. Journal of Econometrics, 31, 1986, p. 307-327.

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

ENDERS, Walter. ***Applied Econometric Time Series***. New York: Wiley, 1995.

ENGLE, R. F. ***Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U. K. Inflation***. Econometrica 50, 1982, p. 987-1008.

GUJARATI, Damodar. **Econometria Básica**. Rio de Janeiro: Elsevier, 2006, 4.a ed.

KOOP, Gary. ***Analysis of Economic Data***. Nova York: John Wiley & Sons, 2000.

OLIVEIRA, Mauri Aparecido de. **Aplicação de Redes Neurais Artificiais na Análise de Séries Temporais Econômico-Financeiras**. São Paulo:[s.n.], 2007.

SARTORIS, Alexandre. **Estatística e Introdução à Econometria**. São Paulo: Saraiva, 2003.

WOOLDRIDGE, Jeffrey M. ***Introductory econometrics***. South-Westem College Publishing, 2000.