

**IDENTIFICAÇÃO DE PROCESSOS GERADORES DE SÉRIES TEMPORAIS  
FINANCEIRAS: USO DE ENTRADAS EXÓGENAS NA MODELAGEM DO IBOVESPA**

Guilherme Keiel<sup>a</sup>, Fernando Augusto Bender<sup>b\*</sup>

a) Universidade Federal do Rio Grande do Sul - UFRGS.

b) Universidade de Caxias do Sul - UCS

| Informações de Submissão  | Resumo   |
|---|--|
| <p>* Autor correspondente (Orientador)<br/>Fernando Augusto Bender, endereço: Rua Francisco Getúlio Vargas, 1130 - Caxias do Sul - RS - CEP: 95070-560.</p> | <p>A análise de séries temporais financeiras é de grande importância no ramo das finanças, possibilitando construir modelos matemáticos capazes de representar ou mesmo prever valores futuros de uma série de ativos. Este trabalho propõe a modelagem do processo gerador da série financeira do Ibovespa por métodos clássicos considerando variáveis explicativas na estrutura do modelo. Foram identificados modelos autorregressivos integradores de média móvel com entrada exógena dada pelo valor do dólar ou do índice Dow Jones e avaliados quantitativamente, comparando-os com os resultados do modelo em que não há o seu uso. Por fim, os modelos são validados utilizando-os para a predição da série, comprovando a maior eficácia do modelo com entrada exógena Dow Jones em relação aos demais.</p> |
| <p><b>Palavras-chave:</b><br/><i>ARIMAX</i>, Séries financeiras. Identificação de sistemas. Previsão macroeconômica.</p>                                    |  |

## 1 INTRODUÇÃO

A análise de séries temporais financeiras lida com a avaliação dos preços ou retornos de ativos durante o tempo. Estes retornos, em geral, são não correlacionados entre si, apresentam agrupamento de volatilidade, possuem distribuição diferente da normal e, em alguns casos, são não lineares (MORETTIN, 2011). A análise e classificação dessas séries de retornos quanto as suas principais características possibilitam entender a dinâmica do mercado de ações a fim de usá-la para uma série de finalidades, entre as quais a de realizar previsões de valores futuros da série.

O livro (BOX; JENKINS; REINSEL, 1994), em sua primeira edição na década de 70, resumia a formulação de séries através de equações matemáticas, popularizando o uso do

modelo autoregressivo integrador de média móvel (do inglês *autoregressiveintegratedmoving-average*, *ARIMA*), sendo empregado para análise macroeconômica em (DHRYMES; PERISTIANI, 1988), por exemplo. A esta representação mais tarde foi adicionado o modelo autoregressivo condicional heteroscedástico generalizado (do inglês *generalizedautocorrelationconditionalheteroscedasticity*, *GARCH*), introduzido por (BOLLERSLEV, 1986), capaz de descrever também séries com volatilidade variante no tempo. Outros autores utilizam ainda estes modelos multivariados, capazes de modelar simultaneamente mais de uma variável endógena do processo e sua interdependência, conforme (ÖLLER, 1985). No entanto, uma parcela menor dos trabalhos leva em consideração influências exógenas no modelo do processo, como usado em (BOS; FRANCES; OOMS, 2002). Este último propicia representações mais fidedignas da dinâmica do preço/retorno de ações em decorrência de fatores externos, como valores de mercados internacionais e, portanto, será objeto de estudo neste trabalho.

Este artigo tem como principal objetivo a avaliação de modelos *ARIMAX* para a descrição do processo gerador da série financeira do índice diário da Bolsa de Valores de São Paulo (Ibovespa), utilizando como entradas exógenas o valor do dólar ou do índice Dow Jones no mesmo período. Entre os objetivos secundários está a comparação da qualidade destes com os modelos *ARIMA*, em que não há entrada exógena, e sua validação pela predição de passos adiante da série.

O trabalho subdividiu-se da seguinte forma: no Capítulo 2 é apresentado o referencial teórico fundamental para a compreensão dos modelos utilizados. O Capítulo 3 descreve a metodologia empregada nos ensaios e o Capítulo 4 apresenta as avaliações e resultados obtidos. Por fim, o Capítulo 5 realiza uma discussão desses resultados e aponta perspectivas de trabalhos futuros.

## **2 REFERENCIAL TEÓRICO**

Neste trabalho serão apresentadas as metodologias básicas para a análise e classificação de séries temporais financeiras, introduzindo os modelos lineares invariantes no tempo popularmente utilizados para a sua representação.

## 2.1 Estacionariedade

A estacionariedade é a base na análise de séries temporais. Uma série  $\{y_t\}_{t=1}^N$  é estritamente estacionária se sua distribuição de probabilidade é invariante sob deslocamento no tempo. Já esta série é dita ser estacionária no sentido amplo se sua média e covariância são invariantes no tempo (TSAY, 2005).

## 2.2 Correlação

Séries temporais são realizações de processos estocásticos, que por sua vez, são resultado de uma coleção de variáveis aleatórias. O coeficiente de correlação  $\rho_{u,y}$  mede a dependência linear entre duas variáveis aleatórias  $U$  e  $Y$ , definido no intervalo  $[-1, 1]$ . O coeficiente de correlação pode ser estimado por

$$\hat{\rho}_{u,y_\ell} = \frac{\sum_{k=1}^{N-\ell} (u_k - \bar{u})(y_{k+\ell} - \bar{y})}{\sqrt{\sum_{k=1}^{N-\ell} (u_k - \bar{u})^2 \sum_{k=1}^{N-\ell} (y_{k+\ell} - \bar{y})^2}} \quad (1)$$

sendo  $\bar{u}$  e  $\bar{y}$  as médias de  $U$  e  $Y$ , respectivamente e  $N$  o número de amostras (TSAY, 2005). A função de correlação cruzada (do inglês *cross-correlation function*, *CCF*) das amostras é uma apresentação sequencial desses coeficientes calculados em cada atraso  $\ell = \pm 1, \pm 2, \dots, \pm N$  e possibilita interpretar o nível de dependência linear entre as duas séries na amostra atual ou em relação a amostras defasadas.

O coeficiente de autocorrelação é uma particularização do coeficiente de correlação (1) medindo a dependência entre valores atuais e passados da mesma série. A função de autocorrelação (do inglês *autocorrelation function*, *ACF*) é uma sequência desses coeficientes obtida para cada atraso  $\ell = 0, 1, 2, \dots, N-1$  pela equação

$$\hat{\rho}_{y_\ell} = \frac{\sum_{k=\ell+1}^N (y_k - \bar{y})(y_{k-\ell} - \bar{y})}{\sum_{k=1}^N (y_k - \bar{y})^2} \quad (2)$$

Estendendo a análise, a função de autocorrelação parcial (do inglês *partial autocorrelation function*, *PACF*) é uma função da *ACF* com a sequência de seu respectivo coeficiente nos  $\ell$  atrasos, obtido como o último parâmetro resultante da estimação

do modelo autoregressivo de ordem  $\ell$  da série (TSAY, 2005). Juntas, a *ACF* e *PACF*, fornecem informações capazes de ajudar na escolha de modelos.

### 2.3 Modelos *ARIMA*

Um modelo autoregressivo de média móvel (do inglês *autoregressivemoving-average*, *ARMA*) expressa a média condicional de  $y$  no instante  $k$  como uma função das observações passadas  $y_{k-1}, \dots, y_{k-p}$  e das inovações do resíduo  $e_{k-1}, \dots, e_{k-q}$  conforme apresenta (BOX; JENKINS; REINSEL, 1994) através da equação

$$y_k = \phi_0 + \sum_{i=1}^p \phi_i y_{k-i} + e_k + \sum_{i=1}^q \theta_i e_{k-i} \quad (3)$$

onde  $p, q \in \mathbb{N}^+$  definem os graus do polinômio auto-regressivo com parâmetros  $\phi$  e do polinômio da média móvel com parâmetros  $\theta$ , respectivamente.

Permitindo que o polinômio autoregressivo possua raízes unitárias, uma característica de processos não-estacionários, o modelo apresentado pode ser estendido para um modelo autoregressivo integrador de média móvel (*ARIMA*) descrito por

$$w_k = \phi_0 + \sum_{i=1}^p \phi_i w_{k-i} + e_k + \sum_{i=1}^q \theta_i e_{k-i} \quad (4)$$

sendo  $w_k = \Delta^d y_k$ , que significa a  $d$ -ésima diferença de  $y_k$ , equivalente a um processo estacionário. Isto implica que a saída do processo não-estacionário  $y_k$  pode ser obtida somando  $d$  vezes o resultado de  $w_k$  modelado por um processo *ARMA*. Estes modelos para séries temporais não estacionárias, considerados desde, são de fundamental importância nos problemas de previsão e controle.

### 2.4 Modelos *ARIMAX*

Partindo de uma análise de sistemas de controle, o modelo *ARIMA* pode ser generalizado para descrever processos não-estacionários que possuam também uma dependência linear de entradas externas. O modelo auto-regressivo integrador de média móvel com entrada exógena (do inglês *autoregressiveintegratedmoving-averagewithexogenous input*, *ARIMAX*) é descrito por

$$w_k = \phi_0 + \sum_{i=1}^p \phi_i w_{k-i} + e_k + \sum_{i=1}^q \theta_i e_{k-i} + \sum_{i=1}^r \beta_i u_{k-i} \quad (5)$$

onde  $\theta \in \mathbb{R}^q$  define o grau do polinômio de ponderação da entrada exógena através dos parâmetros  $\beta$ .

## 2.4 Estimação dos parâmetros

Os parâmetros de modelos da subclasse *ARLX* são estimados buscando a minimização dos quadrados dos resíduos, através do método dos mínimos quadrados (LJUNG, 1999):

$$\hat{\theta} = [\sum_{k=1}^N \varphi_k \varphi_k']^{-1} [\sum_{k=1}^N \varphi_k Y_k] \quad (6)$$

sendo  $\theta$  o vetor dos parâmetros com dimensão  $p + r$ ,  $\varphi$  o vetor de regressão com os valores passados de entradas e saídas relacionadas a cada amostra  $k$  e  $Y$  o vetor com as saídas  $y$  em cada amostra  $k$ .

Quando adotados coeficientes de média móvel no modelo a relação com os erros de predição passados é desconhecida, uma vez que os resíduos não são mensuráveis, portanto faz-se necessário o uso de métodos de busca numérica para estimação de  $\theta$  com dimensão  $p + q + r$ . São utilizados os métodos da Descida ou mínimos quadrados não lineares, buscando minimizar uma função custo dos erros de predição.

## 3 METODOLOGIA

As séries temporais utilizadas estão à disposição no site da Bolsa de Valores de São Paulo, retratando os fechamentos dos índices diários da Ibovespa no período de 03 de Janeiro de 2000 à 19 de Agosto de 2014, totalizando 3537 amostras. Os dados utilizados como a entrada externa no processo, do dólar ou do índice Dow Jones, compreendem o mesmo período. A análise probabilística da série e estimação dos modelos é realizada com o auxílio do *software* Matlab.

Primeiramente a série temporal principal é avaliada com relação à estacionariedade, verificando a necessidade de realizar sua diferenciação a fim de possibilitar a utilização dos modelos abordados neste artigo. Na sequência são realizadas avaliações das funções de

correlação com dados passados da mesma série e com dados das séries consideradas como entradas do processo, a fim de identificar qual a melhor subclasse de modelos a ser utilizada.

A subclasse de modelos escolhida para a representação do processo deve ser estimada para um crescente número de parâmetros, utilizando-se de um critério de desempenho para evitar a sobre-parametrização desnecessária do modelo. A qualidade do ajuste à curva do processo é então comparada com os outros modelos, a fim de se verificar as vantagens ou não na utilização de outras entradas na descrição do processo.

## 4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Nesta seção será aplicada a metodologia descrita para a análise da série temporal do Ibovespa e, a partir disto, obter modelos que descrevam o processo gerador desta série.

### 4.1 Análise da estacionariedade

A inspeção visual da série do Ibovespa, apresentada na Figura 1, bem como as avaliações das médias e variâncias de menores segmentos da série permite concluir que a mesma não é estacionária. A nova série formada pelas diferenças de primeira ordem de  $y$  já possui grau suficiente de estacionariedade, portanto será utilizado  $d = 1$  na determinação dos modelos.

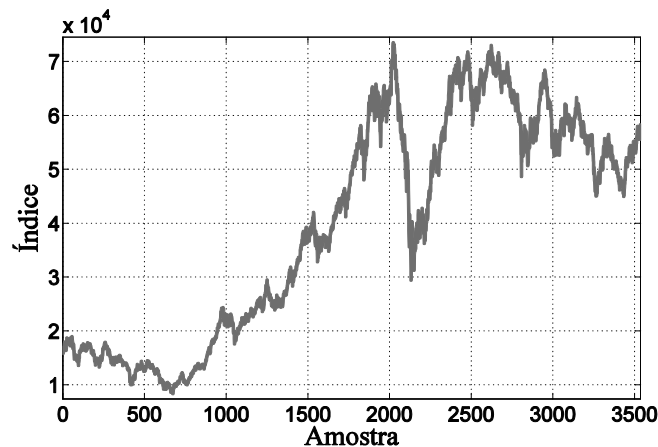


Figura 1: Série do Ibovespa no período 03/01/2000 a 19/08/2014.  
 Fonte: BM&FBOVESPA (2016).

#### 4.2 Análise da correlação

Pelo traçado da *ACF* e *PACF* da série já diferenciada, apresentadas na Figura 2, não se verifica nenhuma correlação significativa entre as amostras e seus 15 primeiros atrasos. Portanto, não é possível definir um comportamento característico do processo e nenhum modelo parece ser predominante, fazendo-se necessário verificar outras influências no modelo, como por exemplo, a influência de entradas exógenas na composição do processo.

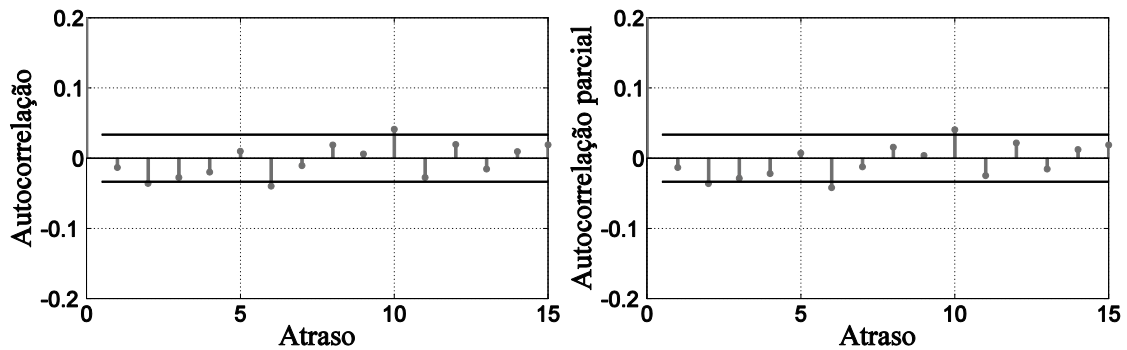


Figura 2: Funções de autocorrelação e autocorrelação parcial da Ibovespa.

Estas entradas, apresentadas na Figura 3, constituem-se das séries com a evolução do valor do dólar e do índice Dow Jones para o mesmo período. A Figura 4 apresenta a *CCF* entre o Ibovespa e as séries do dólar ou do índice Dow Jones, onde é verificada uma correlação elevada entre as amostras atuais das mesmas, sobretudo do Dow Jones, e outra ainda significativa com a primeira amostra passada de ambas.

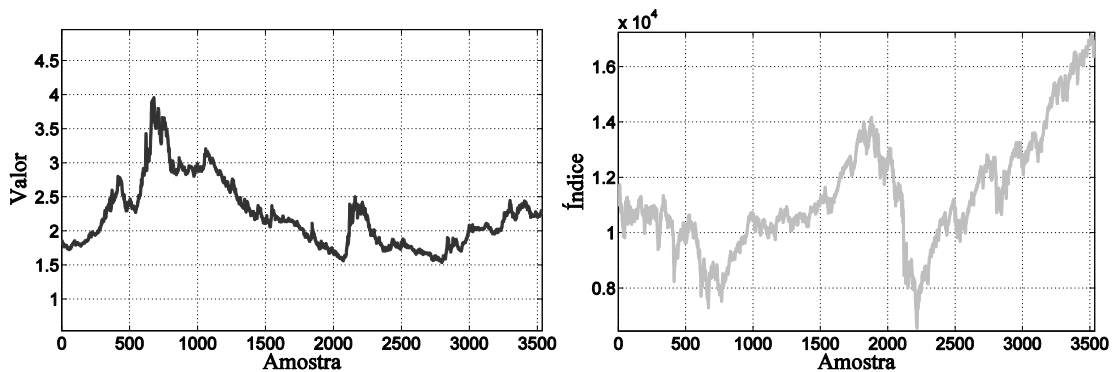


Figura 3: Séries do preço do dólar e do índice Dow Jones no período 03/01/2000 a 19/08/2014.

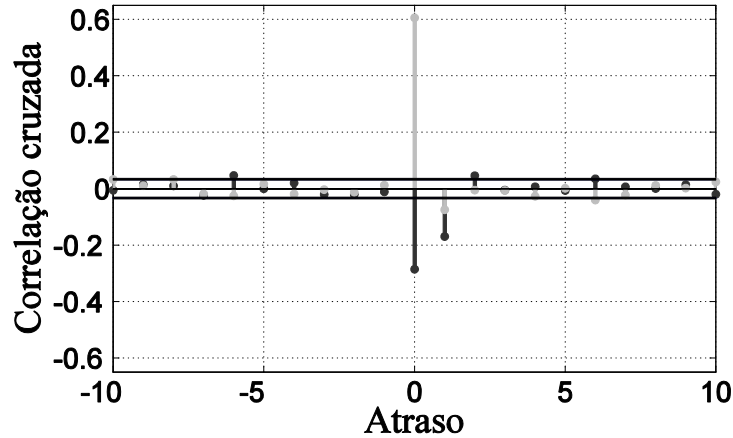


Figura 4: Função de correlação cruzada entre Ibovespa e dólar/Dow Jones.

Através das informações das funções de autocorrelação e da função de correlação cruzada, observa-se uma dependência com rápido decaimento que sugere um modelo com ordem  $r = 2$ , no mínimo. Os resultados até então sugerem um modelo  $ARIMAX(0,1,0,2)$ , que será o ponto de partida para a determinação do modelo.

### 4.3 Identificação do modelo

A partir das características inferidas na análise probabilística da série, é possível realizar a identificação de uma subclasse de modelo adequada. A Tabela 1 descreve a qualidade dos modelos estimados para subclasses  $ARIMA$  com diferentes valores de  $p/q$ . Verifica-se um menor erro absoluto médio (do inglês *mean absolute error*,  $MAE$ ) conforme a ordem do modelo é elevada, o que é esperado uma vez que os coeficientes são determinados minimizando o erro quadrático médio. O comportamento do erro absoluto máximo (do inglês *maximum absolute error*,  $ME$ ) é similar, entretanto seu decréscimo nem sempre é garantido.

Tabela 1–Resíduos dos modelos  $ARIMA$ .

| $p/q$ | $ME$    | $MAE$  | $AIC$   |
|-------|---------|--------|---------|
| 1     | 5024,05 | 504,56 | 13,2167 |
| 2     | 5008,59 | 503,91 | 13,2178 |
| 3     | 5041,27 | 503,90 | 13,2191 |
| 4     | 5144,19 | 503,56 | 13,2189 |
| 5     | 5069,27 | 502,99 | 13,2200 |
| 10    | 5007,44 | 501,54 | 13,2225 |
| 15    | 4890,57 | 501,20 | 13,2341 |
| 20    | 5022,95 | 499,41 | 13,2242 |



A Tabela 2 compara, agora, a qualidade dos modelos para subclasses *ARIMAX*, nos quais são consideradas as séries do dólar do Dow Jones como as entradas  $X^1$  ou  $X^2$  do processo, respectivamente.

Tabela 2–Resíduos dos modelos *ARIMAX*( $p, 1, q, 2$ ).

| $p/q$ | Entrada $X^1$ |            |            | Entrada $X^2$ |            |            |
|-------|---------------|------------|------------|---------------|------------|------------|
|       | <i>ME</i>     | <i>MAE</i> | <i>AIC</i> | <i>ME</i>     | <i>MAE</i> | <i>AIC</i> |
| 0     | 4195,82       | 494,34     | 13,1335    | 4098,74       | 436,59     | 12,7564    |
| 1     | 4377,82       | 493,68     | 13,1281    | 4073,52       | 436,72     | 12,7578    |
| 2     | 4371,61       | 493,47     | 13,1286    | 4113,21       | 436,56     | 12,7565    |
| 3     | 4357,76       | 493,50     | 13,1272    | 4122,71       | 436,27     | 12,7582    |
| 4     | 4402,48       | 493,53     | 13,1284    | 4108,75       | 435,95     | 12,7596    |
| 5     | 4400,68       | 493,52     | 13,1303    | 4109,16       | 435,85     | 12,7608    |
| 10    | 4339,71       | 491,75     | 13,1335    | 4045,87       | 434,52     | 12,7671    |
| 15    | 4273,77       | 491,56     | 13,1414    | 4054,21       | 433,47     | 12,7737    |
| 20    | 4256,61       | 492,25     | 13,1473    | 4043,33       | 432,23     | 12,7811    |

Buscando por uma ponderação entre a qualidade e a complexidade dos modelos, é necessário um critério de informação adicional desencorajando o excessivo número de parâmetros. Para a escolha de modelos com esta parcimônia é aplicado o Critério de Informação de Akaike (do inglês *Akaike's Information Criterion*, *AIC*), o qual determina a qualidade relativa do modelo como uma função de seu custo somado a um peso de seu número total de parâmetros (LJUNG, 1999). Pela Tabela 1 concluímos que o modelo *ARIMA*(1,1,1) parece ser a melhor escolha, enquanto que na Tabela 2 sobressaem-se os modelos *ARIMAX*<sup>1</sup>(3,1,3,2) e *ARIMAX*<sup>2</sup>(2,1,2,2), respectivamente. Portanto, a Tabela 3 sumariza estes modelos selecionados, entre os três casos considerados, e descreve seus respectivos parâmetros:

Tabela 3–Parâmetros dos modelos selecionados.

| Modelo                               | $\phi$    | $\theta$  | $\beta$ |
|--------------------------------------|-----------|-----------|---------|
| <i>ARIMA</i> (1,1,1)                 | 1,0       | 1,0       |         |
|                                      | -0,7259   | -0,7569   | -       |
| <i>ARIMAX</i> <sup>1</sup> (3,1,3,2) | 1,0       | 1,0       | 3,595   |
|                                      | 8,902e-5  | -0,009245 | 0,3596  |
|                                      | -0,008472 | -0,0593   |         |
|                                      | 0,02498   | 0,003611  |         |
| <i>ARIMAX</i> <sup>2</sup> (2,1,2,2) | 1,0       | 1,0       | -9211,0 |
|                                      | -0,3171   | -0,3878   | 3369,0  |
|                                      | -0,1106   | -0,1312   |         |

A Tabela 4 apresenta um comparativo dos resíduos dos modelos para um crescente número de  $r$ , ou seja, do grau do polinômio da entrada exógena. Verifica-se, novamente, que um elevado número de parâmetros não compensa. Uma constatação dessa sobreparametrização é o cancelamento entre os polos e zeros do filtro formado por estes modelos, indicando sua pouca influência na composição da dinâmica do processo.

Tabela 4—Resíduos dos modelos  $ARIMAX(p, 1, q, r)$ .

| $p/q/r$ | Entrada $X^1$ |        |         | Entrada $X^2$ |        |         |
|---------|---------------|--------|---------|---------------|--------|---------|
|         | $ME$          | $MAE$  | $AIC$   | $ME$          | $MAE$  | $AIC$   |
| 1       | 4178,01       | 493,61 | 13,1304 | 4076,33       | 436,71 | 12,7570 |
| 2       | 4371,61       | 493,47 | 13,1286 | 4113,21       | 436,56 | 12,7565 |
| 3       | 4343,47       | 493,61 | 13,1295 | 4132,81       | 436,10 | 12,7591 |
| 4       | 4381,66       | 493,70 | 13,1296 | 4124,36       | 436,22 | 12,7609 |
| 5       | 4421,22       | 493,21 | 13,1300 | 4102,96       | 435,57 | 12,7612 |

#### 4.4 Diagnóstico e validação

Uma análise dos resíduos possibilita checar se os modelos selecionados são apropriados. A Figura 5 apresenta a  $ACF$  e  $CCF$  dos resíduos do modelo  $ARIMAX^2(2, 1, 2, 2)$ , verificando-se que estes são praticamente não-correlacionados entre si e a correlação com a entrada do modelo ainda existe, porém é mínima.

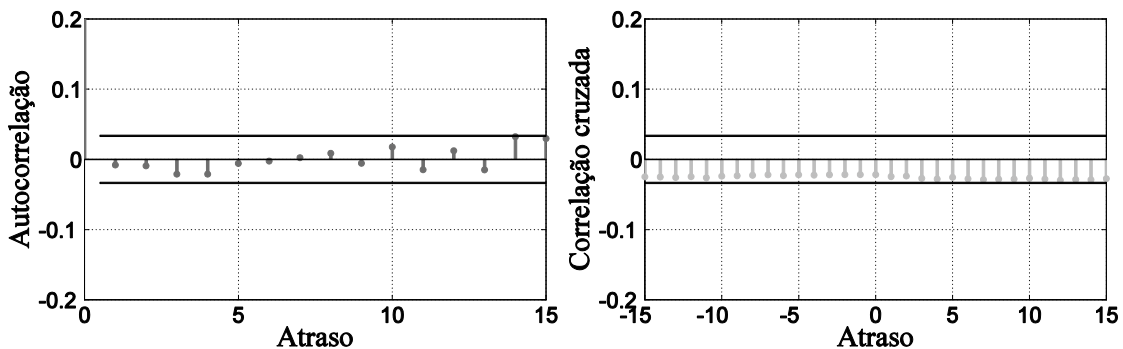


Figura 5: Funções de autocorrelação dos resíduos e da correlação cruzada entre as entradas e resíduos.

Por fim, os modelos identificados são observados quanto à capacidade de se adequarem à curva original do processo, considerando a predição de três até cinco passos futuros da série, comparando-os com momentos específicos da série já conhecidos. As Figuras 6 e 7 apresentam predições de três e cinco passos adiantes da série com esses modelos a partir da amostra 2987. Verifica-se um menor  $MAE$  para o modelo  $ARIMAX^2$ , enquanto que o modelo

sem entrada exógena apresentou, em ambos os casos, o maior erro absoluto.

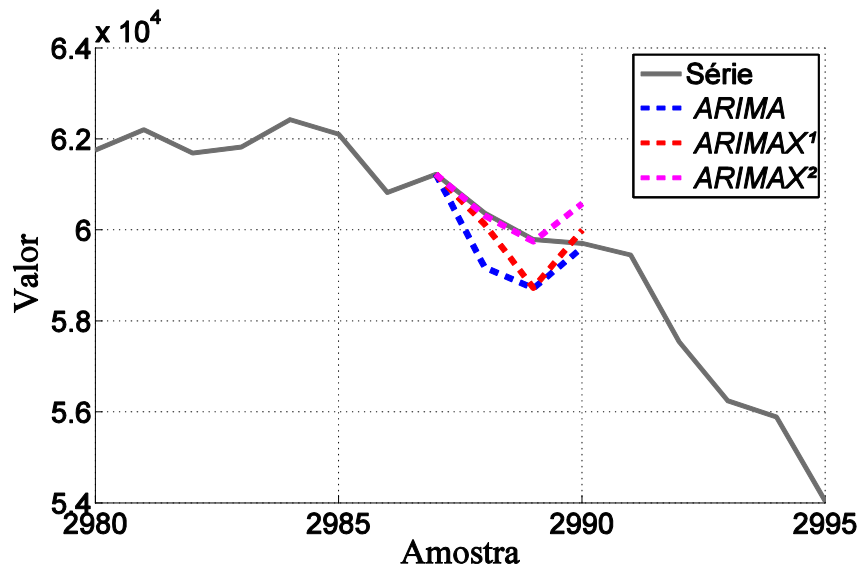


Figura 6: Comparação das predições de três passos adiante dos modelos.

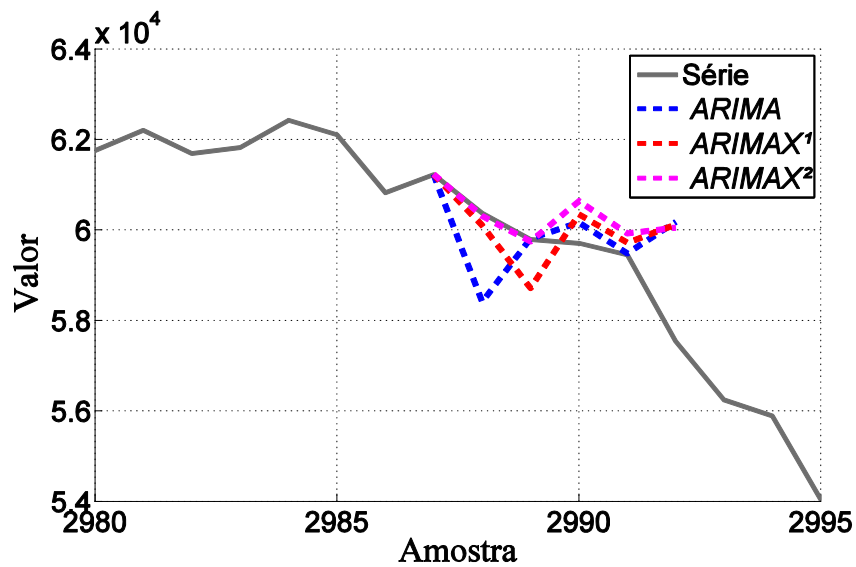


Figura 7: Comparação das predições de cinco passos adiante dos modelos.

## 5 CONSIDERAÇÕES FINAIS

Este trabalho propôs modelar o processo gerador da série financeira do Ibovespa considerando também entradas exógenas na estrutura do modelo. Para isso foram

identificados modelos *ARIMAX* e avaliados quantitativamente, comparando-os com os resultados de modelos *ARIMA*, em que não há uso de variável externa. Por fim, os modelos são validados utilizando-os para a predição da série temporal, comprovando a maior eficácia do modelo *ARIMAX* usando como variável explicativa o índice Dow Jones. Trabalhos futuros na área buscarão a utilização de estruturas de modelo mais elaboradas ou retornos de ativos como variável do processo.

## 6 REFERÊNCIAS

- BM&FBOVESPA. **Séries históricas.** Disponível em: <[http://www.bmfbovespa.com.br/pt\\_br/servicos/market-data/historico/mercado-a-vista/series-historicas/](http://www.bmfbovespa.com.br/pt_br/servicos/market-data/historico/mercado-a-vista/series-historicas/)>. Acesso em: 14 abr. 2016.
- BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. **Journal of Econometrics**. v.31, n.3, p.307-327, 1986.
- BOS, C.;FRANSES, P. H.;OOMS, M. Inflation, forecast intervals and long memory regression models. **International Journal of Forecasting**. v.18, n.2, p.243-264, 2002.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis: forecasting and control**. 3 ed. Prentice-Hall, 1994.
- DHRYMES, P. J.; PERISTIANI, S. C. A comparison of the forecasting performance of WEFA and ARIMA time series methods. **International Journal of Forecasting**, v.4, n.1, p.81-101, 1988.
- LJUNG, L. **System identification: theory for the user**. 1 ed. Prentice Hall PTR, 1999.
- MORETTIN, P. A. **Econometria financeira: um curso em séries temporais financeiras**. 2 ed. Blucher, 2011.
- ÖLLER, L. Macroeconomic forecasting with a vector ARIMA model. **International Journal of Forecasting**, v.1, n.2, p.143-150, 1985.
- TSAY, R. S. **Analysis of financial time series**. 2 ed. Wiley, 2005.