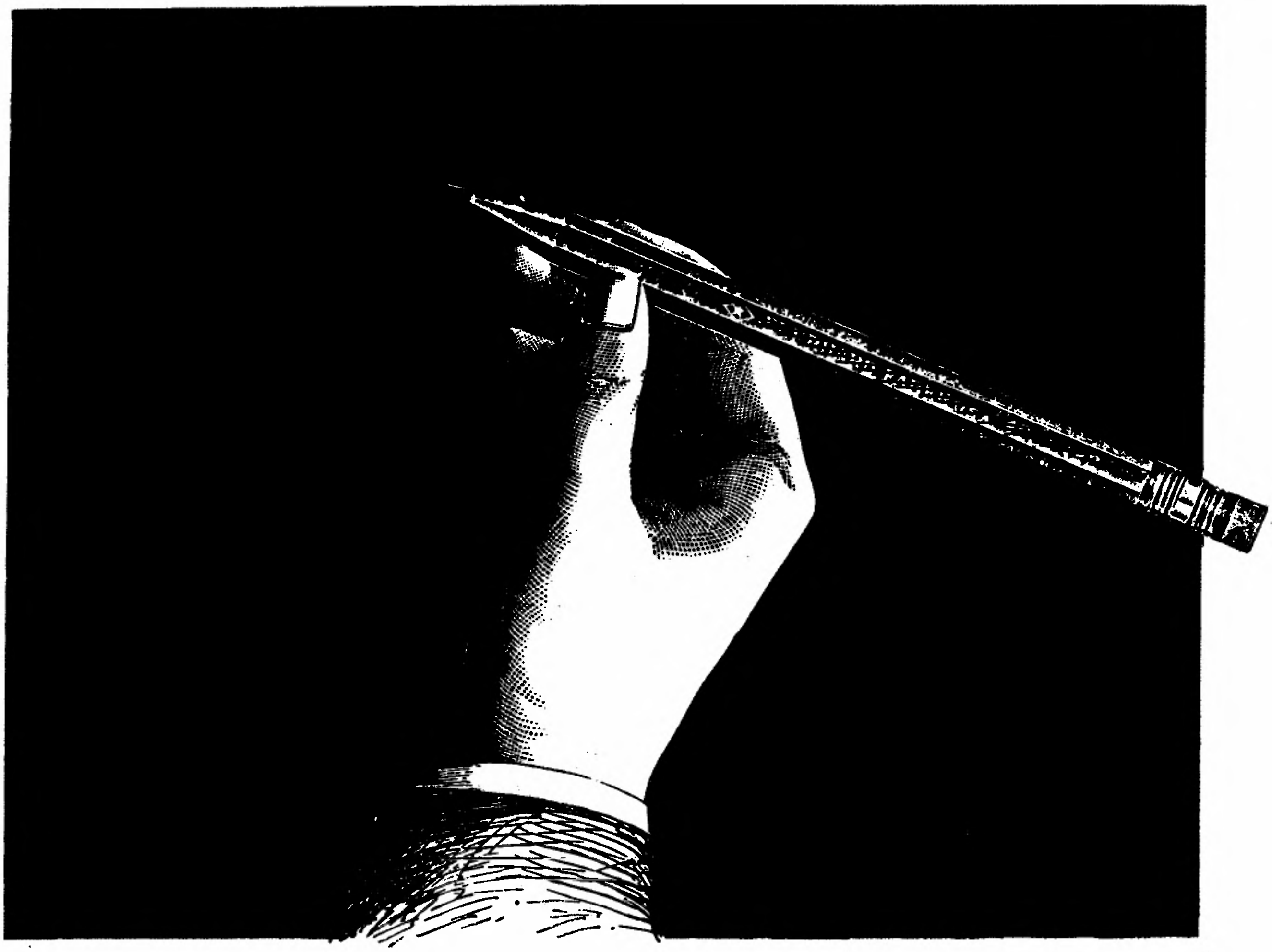
Os Modelos Arima e a Abordagem de Box-Jenkins

Uma Aplicação na Previsão do IBOVESPA a Curtíssimo Prazo



RESUMO: *Este trabalho analisa os modelos auto-regressivo-médias-móveis e os compara com os métodos tradicionais denominados de "análise técnica".*
A abordagem de Box-Jenkins é apresentada em seguida e suas quatro etapas — indentificação, estimação, validação e previsão — são analisadas.
Finalmente, mostra como esta técnica pode ser implementada com excelentes resultados na modelagem e previsão do Índice da Bolsa de Valores de São Paulo (BOVESPA).

PALAVRAS-CHAVE: *ARIMA, Box-Jenkins, previsão, modelagem, processos estocásticos.*

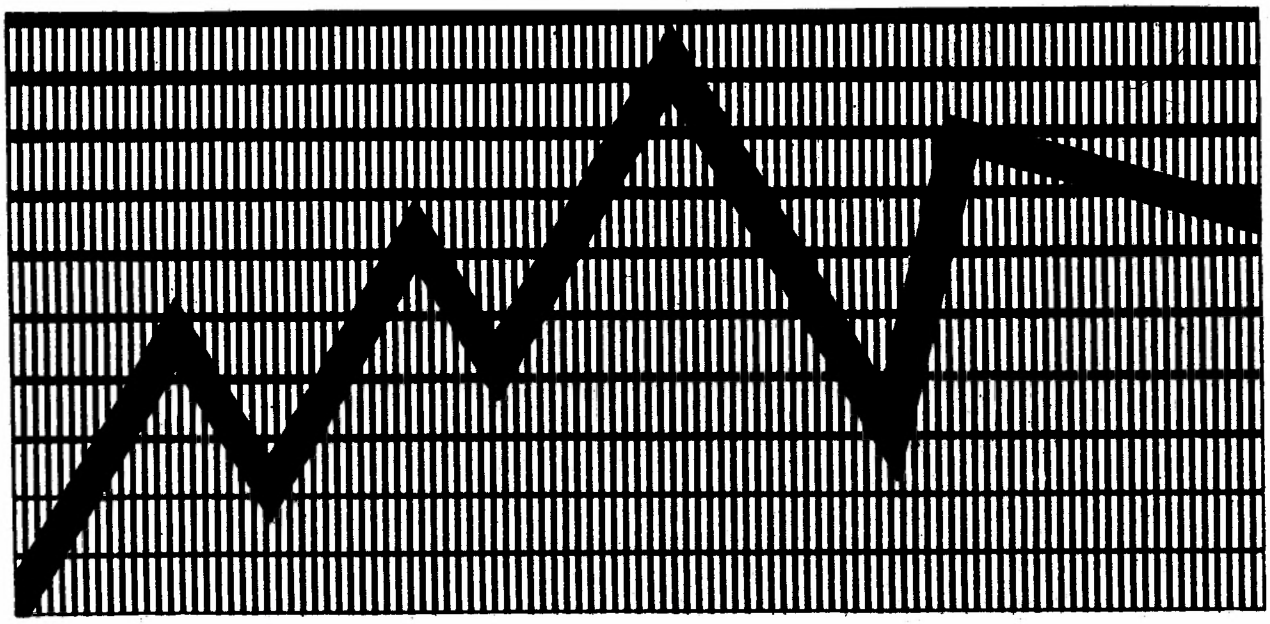
■ **Francisco Carlos Gomes**

Doutorando e Professor no Departamento de Informática e de Métodos Quantitativos aplicados à Administração da EAESP/FGV.

INTRODUÇÃO

No passado, análise técnica foi sinônimo de análise gráfica. Se analisarmos os pressupostos a ela subjacentes, identificados por Edwards e Magee [1], observaremos a presunção de que os

1. EDWARDS, R.D. e MAGEE JR., John. *Technical Analysis of Stock Trends*. Springfield, Stock Trend Service, 1958 (edição revisada).



padrões de comportamento passado se reproduzirão no futuro e, portanto, podem ser empregados com propósitos preditivos (este é o pressuposto da **constância**, subjacente à maioria das técnicas quantitativas de previsão).

Embora o termo análise técnica abranja atualmente algumas ferramentas quantitativas, podemos caracterizá-la como um conjunto de métodos heurísticos, semiquantitativos, que permitem ao analista introduzir as suas **avaliações subjetivas** acerca do padrão de comportamento dos preços e seus **pontos de inflexão** (a antecipação dos *turning points* é um aspecto fundamental dos problemas de previsão).

Esta última característica, a simplicidade e a facilidade de aprendizado e de implementação são os principais responsáveis pela grande popularidade desses métodos.

Contudo, a análise técnica, enquanto um conjunto de métodos de **análise de séries temporais** e/ou **técnicas qualitativas** de previsão, pode ser objeto das seguintes restrições:

a) possui pouca fundamentação teórica e estatística;
b) vários de seus métodos não proporcionam previsões numéricas;
c) não estabelece o **grau de incerteza** associado às respectivas previsões, determinando um intervalo de confiança e respectiva probabilidade de ocorrência;
d) não sistematiza nem calibra as avaliações subjetivas, de sorte a corrigir os vieses associados aos diferentes perfis cognitivos.

Tendo em vista as três primeiras restrições citadas, apresentaremos um método alternativo, que deve ser encarado como uma abordagem **complementar** em lugar de substitutiva da análise técnica.

Os modelos ARIMA foram sistematizados por Box e Jenkins [2]. Esses modelos são robustos do ponto de vista conceitual e estatístico, proporcionam previsões probabilísticas e são de fácil implementação (desde que tenhamos os recursos computacionais adequados). De fato, representam uma generalização dos diversos métodos de análise de séries temporais.

OS MODELOS ARIMA E A ABORDAGEM DE BOX-JENKINS

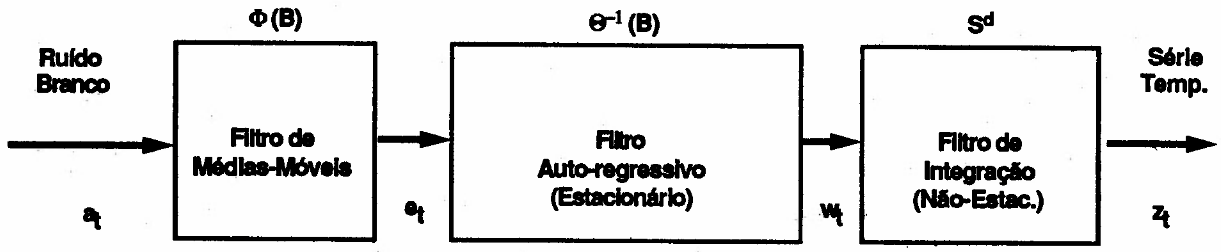
1. Os modelos ARIMA

Tendo em vista os limites desde texto, podemos apresentar sumariamente a forma geral de um modelo ARIMA (p,d,q) como segue:

$$w_t = \mu + \frac{\Theta(B)}{\phi(B)} a_t \qquad (1)$$

2. BOX, George E. P. e JENKINS, Gwilym. *Time Series Analysis: Forecasting and Control*. Oakland, Holden-Day, 1976 (edição revisada). Ver também: MAKRIDAKIS, Spyros e STEVEN, C. Wheelwright. *Forecasting: Methods and Applications*. New York, Wiley, 1978.

FIG. 1 - Diagrama de Blocos dos Modelos Arima



onde:

- t - índice do tempo
- w_t - d'ésima diferença da variável de interesse z_t
- μ - "ponto de referência" do nível do processo
- $\Theta(B)$ - operador de "médias-móveis": $\Theta(B) = (1 - \Theta_1 B^1 - \Theta_2 B^2 - ... - \Theta_q B^q)$
- $\Phi(B)$ - operador auto-regressivo: $\Phi(B) = (1 - \Phi_1 B^1 - \Phi_2 B^2 - ... - \Phi_p B^p)$
- B^p - operador de retrocesso: $B^p z_t = z_{t-p}$
- a_t - "ruído branco" ou erro aleatório

Para maior clareza, o modelo ARIMA (p, d, q) – **diz-se modelo auto-regressivo-médias-móveis integrado de ordem (p, d, q)** – acima estabelecido de forma sintética pode ser expandido, como segue:

$$w_t = \Theta_0 + \Phi_1 w_{t-1} + ... + \Phi_p w_{t-p} + a_t - \Theta_1 a_{t-1} - ... - \Theta_q a_{t-q} \quad (2)$$

onde:

$$\Theta_0 = \mu (1 - \Phi_1 - ... - \Phi_p)$$

De uma forma geral, os modelos ARIMA postulam que as séries temporais (z_t) podem ser representadas por uma seqüência de "choques" aleatórios (a_t) submetidos a três operações de "filtragem" (médias-móveis, auto-regressiva e integração) , como exemplificado na figura 1.

Intuitivamente, podemos afirmar que os modelos ARIMA representam as séries temporais como uma **ponderação** dos próprios valores e/ou erros passados da série.

Um modelo ARIMA (p, d, q) possui p+q+2 parâmetros desconhecidos, que devem ser estimados a partir dos dados, a saber:

a) μ, i.e. o "ponto de referência" do nível do processo;
b) p, parâmetros auto-regressivos $\Phi_1, \Phi_2, ..., \Phi_p$;
c) q, parâmetros médias-móveis $\Theta_1, \Theta_2, ..., \Theta_q$;
d) σ^2, a variância do ruído branco a_t.

2. A Abordagem de Box-Jenkins e o Princípio da Parcimônia

Para a construção dos modelos ARIMA, Box-Jenkins [3] sugeriram as seguintes etapas **iterativas**, como ilustrado na figura 2.

a) Identificação

A identificação de um modelo ARIMA corresponde à determinação:

- **do nível de diferenciação (d), a partir do qual a série se torna estacionária;**
- **da ordem (máxima) dos termos auto-regressivos (p);**
- **da ordem (máxima) dos termos médias-móveis (q).**

Essas determinações são obtidas a partir do exame da **função de autocorrelação e da função de autocorrelação parcial**, que avaliam o padrão de **dependência temporal** da série. Para tanto, basta compararmos as funções de autocorrelação amostrais com os correspondentes **paradigmas** ou funções de autocorrelação "teóricas".

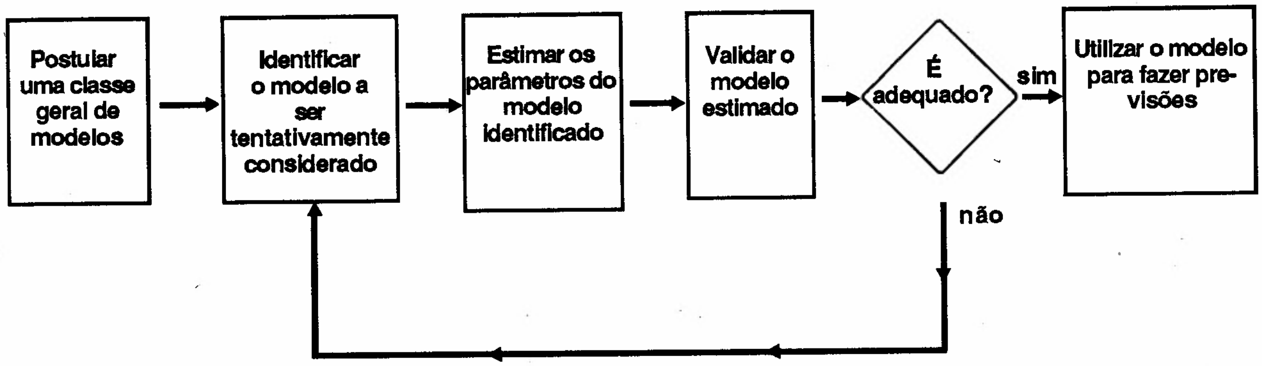
b) Estimação

Uma vez determinada a ordem (p, d, q) do modelo são estimados os seguintes parâmetros:

- μ, o nível do processo;

3. BOX, George, E. P. e JENKINS, Gwilym. Op. cit.

FIG. 2 - Etapas Iterativas da Construção de Modelos de Previsão (Abordagem de Box-Jenkins)



- os parâmetros auto-regressivos $\phi_1, \phi_2, ..., \phi_p$;
- os parâmetros médias-móveis $\Theta_1, \Theta_2, ..., \Theta_q$;
- σ^2, a variância do ruído branco.

Escapa ao escopo deste texto o exame dos complexos algorítmos de estimação. De qualquer forma, existem programas em abundância para executar essa tarefa.

c) Validação

A verificação da adequação do modelo é efetuada em duas dimensões inter-relacionadas, a saber:

- o exame do **grau de ajustamento ou aderência** do modelo – expresso em estatísticas como a variância do erro, o MSE (Mean Square Error) ou o MAPE (Mean Absolute Percentage Error);
- o exame da **aleatoriedade** dos resíduos (erros) do modelo – expresso nas suas funções de autocorrelação.

d) Previsão

Validado o modelo, podemos construir uma **função de previsão,** que além de proporcionar as previsões mais verossímeis dentro do horizonte de planejamento especificado, proporciona também os **limites inferior e superior** do intervalo de confiança associado a um **nível de confidência** (probabilidade) estabelecido pelo analista.

Finalmente, Box e Jenkins [4] sugerem o emprego **parcimonioso** desses modelos, i.e., a sua utilização com um número mínimo de parâmetros, possível em virtude do **teorema da dualidade.**

Tendo em vista que a maioria das séries se torna estacionária após a segunda diferenciação, podemos dizer que, usualmente:

$$p, d, q \leq 2 \qquad (3)$$

UM EXEMPLO : PREVISÃO DO IBOVESPA

A título de exemplo, empregamos uma amostra de 650 observações do Índice de Fechamento da BOVESPA, abrangendo o período de 02/01/85 a 31/08/87.

A figura 3 mostra graficamente que a série considerada foi objeto de várias **mudanças na sua estrutura temporal,** tanto no que diz respeito a suas **tendências** quanto no que concerne a sua **volatilidade**, tornando particularmente difícil a construção de um modelo de previsão.

Empregando-se a abordagem de Box-Jenkins, podemos modelar o IBOVESPA, como segue:

1. Identificação

As figuras 4a e 4b mostram as funções de autocorrelação do IBOVESPA **após a sua diferenciação de primeira ordem**, cujo exame permite concluir que:

a) a série se torna **estacionária** após a primeira diferenciação, portanto d=1.

b) p=1 e q=0, embora não haja uma aderência muito clara aos paradigmas (esses valores foram confirmados através do exame dos modelos alternativos).

---

4. Idem, ibidem.

FIG. 3 - IBOVESPA (Fechamento)

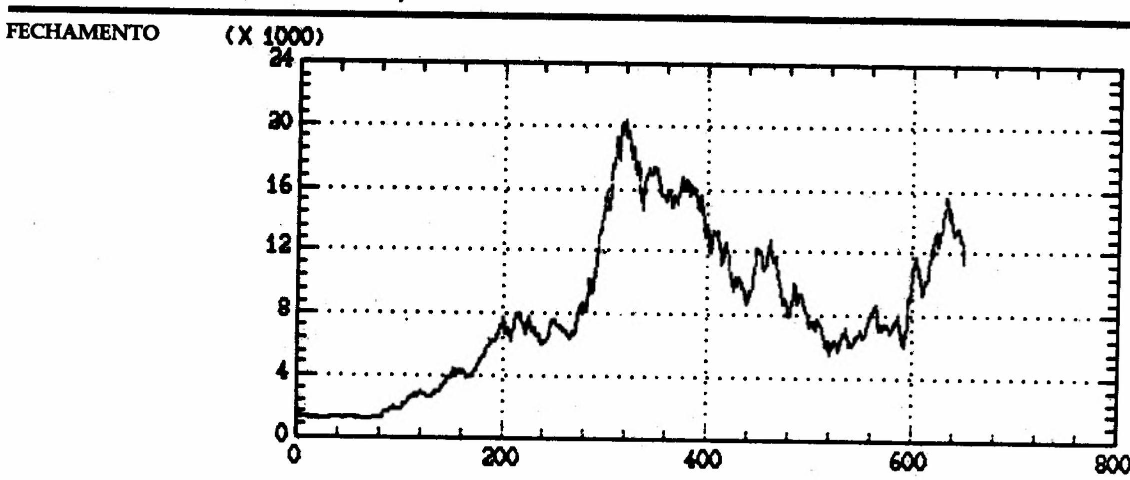


FIG. 4a - Função de Autocorrelação - Primeira Diferença IBOVESPA

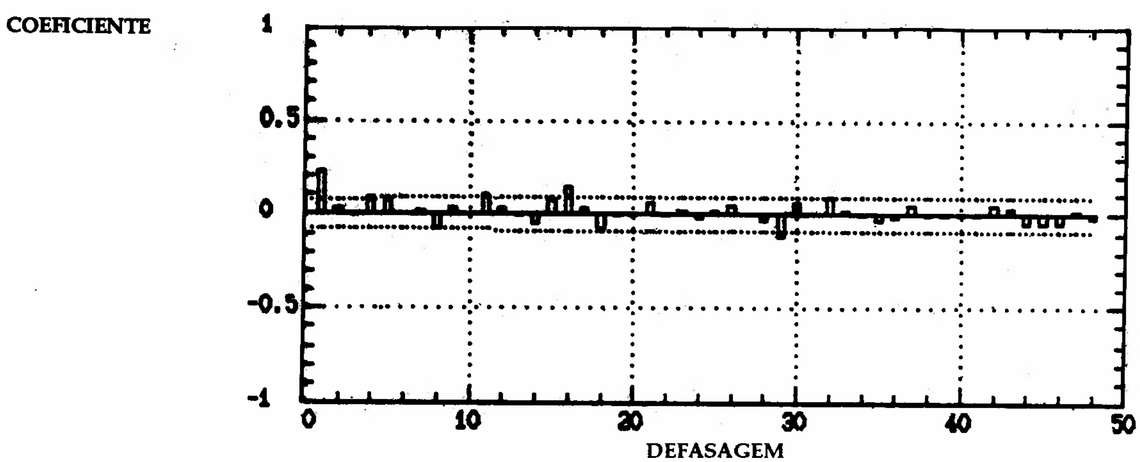
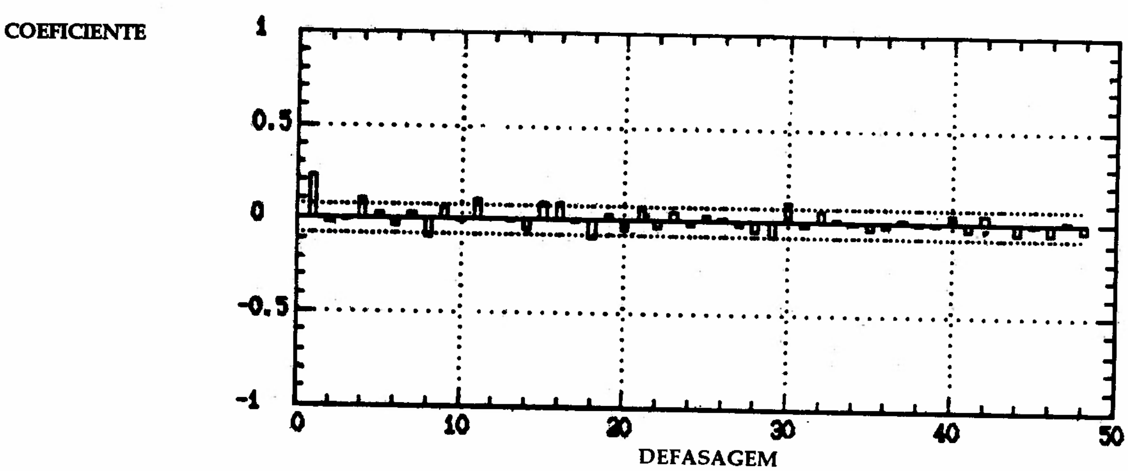


FIG. 4b - Função de Autocorrelação Parcial - IBOVESPA Primeira Diferença



Logo, o IBOVESPA pode ser representado por um modelo ARIMA (1, 1, 0) ou, mais sumariamente, por um modelo auto-regressivo de primeira ordem com um nível de integração [AR (1,1)], como segue:

$$w_t = \Theta_0 + \Phi_1 w_{t-1} + a_t \qquad (4)$$

$$z_t = Sw_t = \sum_{j=0}^{\infty} w_{t-j}$$

onde:

z_t - Índice Bovespa (Fechamento)
w_t - Primeira Diferença de z_t, i. e, $w_t = z_t - z_{t-1}$
μ - "Ponto de Referência" do Nível do Processo
Θ_0 - Constante, i.e., $\Theta_0 = \mu (1 - \Phi_1)$
a_t - Ruído Branco
S - Operador de Integração

2. Estimação

As estimativas obtidas dos parâmetros do modelo identificado foram as seguintes:

IBOVESPA FECHAMENTO - MODELO ARIMA (1, 1, 0)				
Parâmetro	Notação	Estimativa	Erro Padrão	Prob.
Auto-regressivo	Φ_1	0,23584	0,0384	0,0000
Nível do processo	μ	15,23413	0,87854	0,3800
Constante	Θ_0	11,88789		
Variância	σ^2	114.648		

A tabela acima mostra que obtivemos uma estimativa extremamente confiável do termo auto-regressivo de primeira ordem (ϕ_1), já que o seu erro padrão é bastante pequeno (notem que esse termo se mostra estatisticamente diferente de zero a um nível de probabilidade de aproximadamente 100%).

Por outro lado, a estimativa do nível do processo (μ) não é tão confiável, pois possui um erro padrão mais elevado (notem que esse termo se mostra estatisticamente diferente de zero a um nível de probabilidade de apenas 62%) [5].

Finalmente, a estimativa da variância do erro aleatório (σ^2) é relativamente pequena, mostrando – como veremos – a excelente aderência do modelo (notem que a estimativa do desvio padrão (σ), i.e., o erro-padrão, é extremamente pequena, $\sigma = 338{,}60$).

Portanto, o modelo estimado para o Índice Bovespa (Fechamento) foi o seguinte:

$$\hat{w}_t = 11{,}88789 + 0{,}23584 w_{t-1} \qquad (5)$$

$$\hat{z}_t = \sum_{j=0}^{\infty} \hat{w}_{t-j}$$

A figura 5 apresenta os valores estimados do Índice Bovespa (Fechamento) pelo modelo ARIMA (1,1,0) acima.

3. Validação

A figura 6 apresenta os resíduos (erros) do modelo estimado.

a) Aderência

Um sumário das estatísticas básicas dos resíduos (erros) do modelo é apresentado abaixo:

Média: 0,14
Desvio-Padrão: 338,60
Mínimo: -1.264,69 (7,10%; 16/04/86)
Máximo: 1.997,81 (19,67%; 04/03/86)
MAPE [6]: 2,67%

Constatamos, portanto, que o modelo apresentou uma excelente aderência, com um erro percentual médio da ordem de apenas 2,67%.

b) Aleatoriedade dos Resíduos

A figura 7 mostra a função de autocorrelação dos resíduos (erros).

Alguns poucos coeficientes se encontram além da região de aceitação, embora apresentem uma magnitude relativamente pequena. Provavelmente, a crescente volatilidade (variância) da série tenha provocado alguns vieses de estimação, indicando uma provável necessidade de transformação (logarítmica ou raiz quadrada)

5. Neste caso, o procedimento mais adequado seria a estimação de um modelo sem constante; entretanto, manteremos o modelo inalterado, tendo em vista os propósitos ilustrativos deste artigo.

6. O MAPE (Mean Absolute Percentage Error), ou erro percentual médio, é definido como:

$$MAPE = \frac{1}{n} \sum \left| \frac{Z_t - \hat{Z}_t}{Z_t} \right| 100$$

FIG. 5 - IBOVESPA : Estimativas do Modelo Arima (1,1, 0)

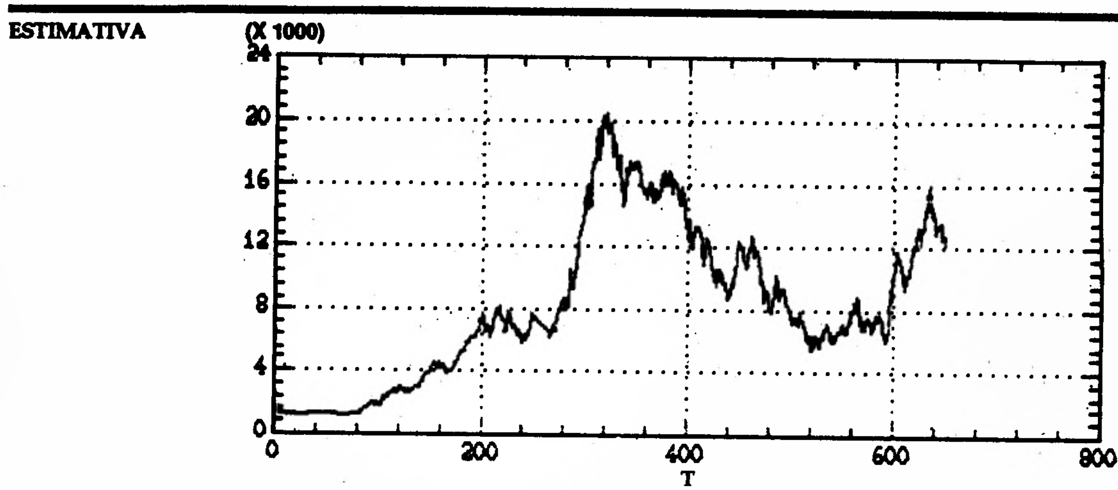


FIG. 6 - IBOVESPA : Erros do Modelo Arima (1,1, 0)

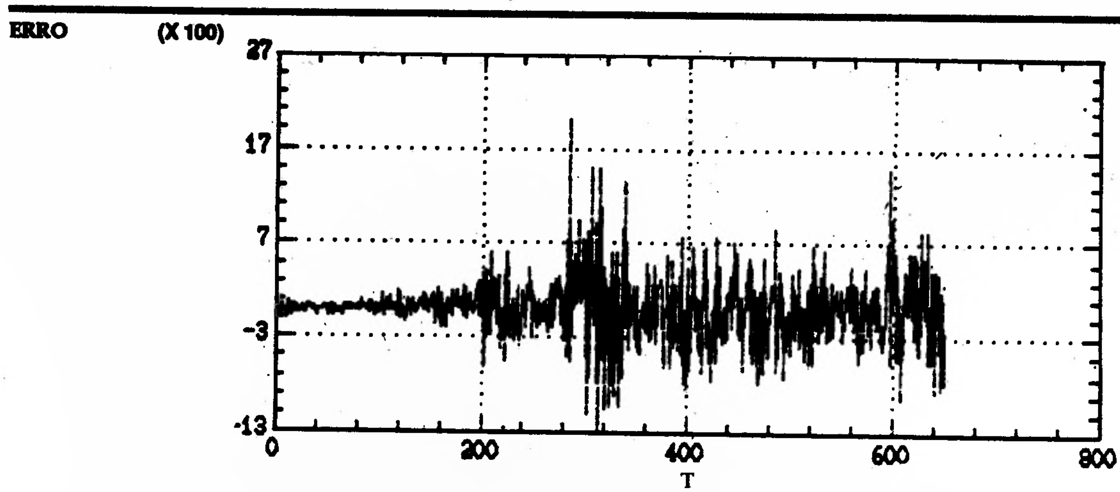
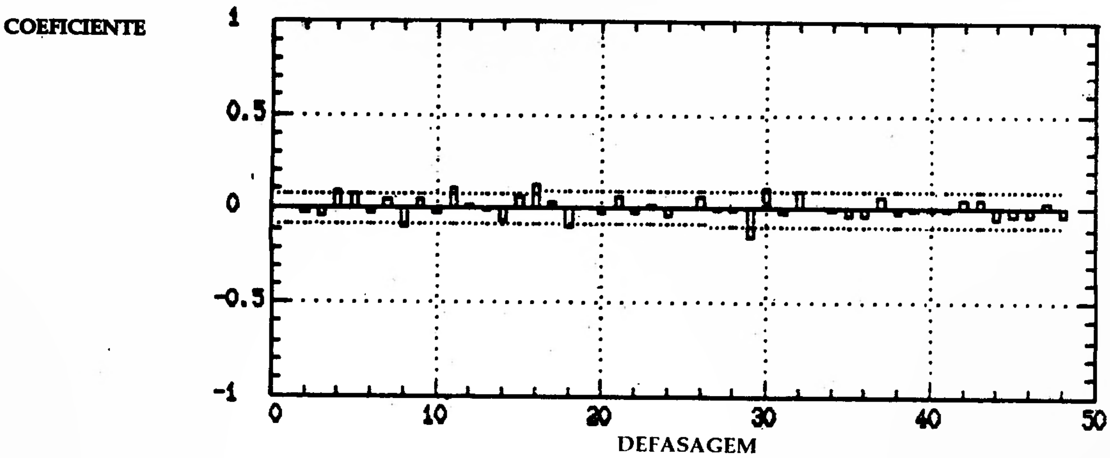


FIG. 7 -Função de Autocorrelação dos Erros do Modelo Arima (1,1, 0)



dos dados para lhes estabilizar a variância. Entretanto, dado o escopo deste texto, a ótima aderência do modelo e a pequena magnitude desses coeficientes, não estimaremos o modelo transformado e consideraremos os resíduos como razoavelmente aleatórios.

4. Previsões

Na figura 8, encontramos a seguinte função de previsão do Índice Bovespa:

$$\hat{w}_{650+m} = 11{,}88789 + 0{,}23584\, w_{649+m}$$

$$\hat{z}_{650+m} = 11.476{,}76 + \sum_{j=1}^{m} w_{650+j} \qquad (6)$$

onde:
m - horizonte de planejamento (m = 1, 2, 3, ...)

Ao lado, apresentamos as previsões do IBOVESPA para os cinco dias úteis subseqüentes à última observação, bem como os respectivos intervalos de confiança associados a um nível de probabilidade de 95% (por simplicidade deixamos de apresentar a expressão descritiva desses limites probabilísticos).

t	z_t	limite inferior	limite superior
651	11.286,3	10.622,6	11.949,9
652	11.253,0	10.197,9	12.308,0
653	11.256,7	9.897,4	12.616,0
654	11.269,3	9.657,7	12.880,9
655	11.283,9	9.453,4	13.114,3

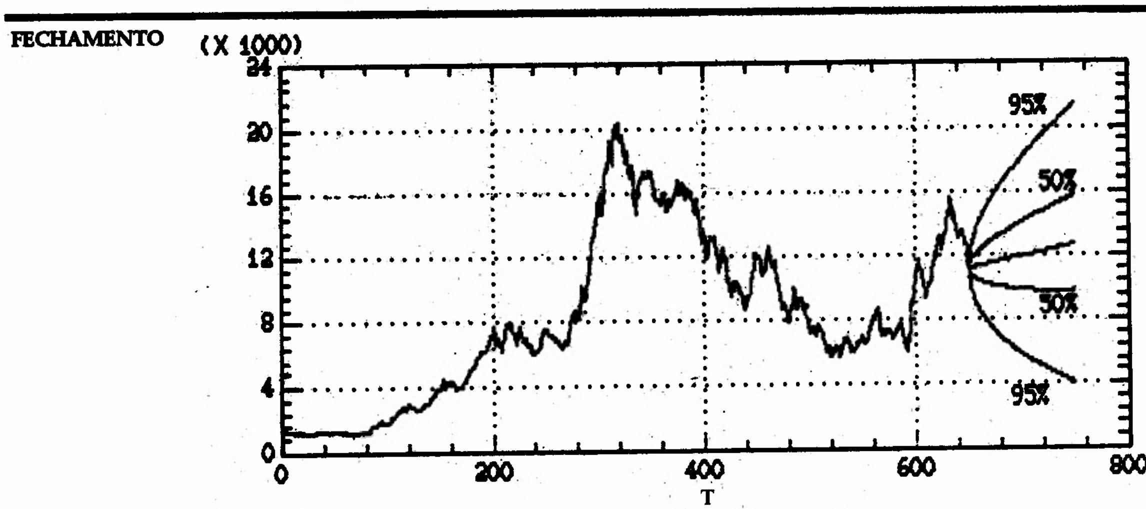
É importante notar que essas previsões são efetuadas no instante t = 650. Tendo em vista o caráter auto-ajustável do método, obteríamos uma maior precisão se essas previsões fossem efetuadas passo a passo (i.e., previsão para o próximo período → observação → previsão para o período subseqüente → etc.).

CONCLUSÕES

Apresentamos uma abordagem alternativa e complementar aos métodos tradicionais de análise técnica, que proporciona previsões quantitativas e probabilísticas robustas do ponto de vista conceitual e estatístico.

No experimento efetuado, constatamos que essa abordagem apresentou resultados extremamente satisfatórios, em que pese à instabilidade e à volatilidade da série empregada. ❑

FIG. 8 - IBOVESPA : Função de Previsão do Modelo Arima (1,1,0)



ABSTRACT: *This paper reviews the Autorregressive - Moving Average Models and compares them with the traditional methods known by the term "technical analysis".*

The Box-Jenkins approach is presented next and its four steps – Identification, Estimation, Checking and Forecasting – are reviewed.

Finally, it is shown how this technique is easily implemented with satisfactory results in the modelling and forecasting of the Index of the Bolsa de Valores de São Paulo (BOVESPA).

KEY WORDS: *Arima, Box-Jenkins, prediction, modelling, stochastic processes.*