

Rafael Varago de Aguiar

# Predição de Séries Temporais - Uma Abordagem Bio-inspirada

Trabalho de Graduação submetido à  
Universidade Federal do ABC como  
parte dos requisitos para a obtenção  
do grau de Engenheiro de Informação.

Santo André - São Paulo

2017

Rafael Varago de Aguiar

# Predição de Séries Temporais - Uma Abordagem Bio-inspirada

Trabalho de Graduação submetido à Universidade Federal do ABC como parte dos requisitos para a Obtenção do grau de Engenheiro de Informação.

**Orientador: Prof. Dr. Ivan Roberto**

**Santana Casella**

Centro de Engenharia, Modelagem e Ciências  
Sociais Aplicadas.

Santo André, Novembro de 2017

### **Ficha Catalográfica**

Aguiar, Rafael Varago de

Predição de Séries Temporais - Uma Abordagem Bio-inspirada. Santo André, 2017. 77 p.

Trabalho de Graduação - Universidade Federal do ABC. Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas.

1. Séries Temporais 2. IBOVESPA 3. ARIMA 4. ANN I. Universidade Federal do ABC. II. Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas.



# Resumo

O estudo e a análise de séries temporais representam uma preocupação crescente para a sociedade moderna, isto justifica-se pelo fato de que muitos processos reais, como a disseminação de doenças, o tráfego de canal em comunicações móveis, índices de mercado de ações etc, são analisados com as ferramentas de séries temporais. Entre as possíveis aplicações para a análise de séries temporais, enfatizamos o problema de antecipar eventos futuros, mitigando os efeitos adversos de sua incerteza inerente. Portanto, métodos para a predição das séries temporais têm um grande valor para o desenvolvimento tecnológico, social e econômico das nações. Entre estas técnicas, as abordagens estatísticas e bio-inspiradas mostraram ótimos resultados no problema de predição de séries temporais.

Desta forma, este trabalho descreve a análise e predição do mais importante índice do mercado de ações brasileiro, o IBOVESPA, utilizando técnicas estatísticas baseadas na abordagem clássica do modelo linear *Autoregressive Integrated Moving Average* e modelo bio-inspirado não-linear baseado na *Artificial Neural Network*. Em complemento, para cada modelo proposto, é feito um estudo da otimização de seus parâmetros livres mais relevantes, considerando as métricas de desempenho empregadas para avaliar seu efeito nas predições realizadas.

Ao longo do desenvolvimento deste trabalho, foram aplicados diversos conceitos aprendidos nas disciplinas do curso de Engenharia da Informação da UFABC.

# Abstract

Time series study and analysis represent a growing concern in the modern society, it is justified by the fact that many real-life processes, such as disease spreading, channel traffic in mobile communications, stock market indices etc, are analyzed with time series framework. Among the the possible applications for time series analysis, we emphasis the problem of anticipating future events, mitigating the adverse effects of its embedded uncertainty. Therefore, methods for time series prediction has a great value for technological, social and economical development of the nations. In the midst of these techniques, the statistical and the bio-inspired approaches have showed great results in the problem of time series prediction.

Thereby, this work describes the analysis and prediction of the most important Brazilian stock market index, the IBOVESPA, using statistical techniques based on the classical approach of the linear Autoregressive Integrated Moving Average model and the bio-inspired non-linear model based on the Artificial Neural Network. In addition, for each proposed model, a study of the optimization of its most relevant free parameters is done, considering the performance metrics used to evaluate its effect on the predictions made.

Throughout the development of this work, several concepts learned in the disciplines of the UFABC Information Engineering course were applied.

# Lista de Figuras

2.1	Fluxo de Recursos no Mercado Financeiro. . . . .	8
2.2	Comportamento do Erro de Predição $e(n)$ . . . . .	22
2.3	Série Temporal Obtida pelo Modelo AR(3). . . . .	23
2.4	Série Temporal Obtida pelo Modelo MA(3). . . . .	24
2.5	Série Temporal Obtida pelo Modelo ARMA(3, 3). . . . .	26
2.6	Séries Temporais Obtidas pelo Modelo I(1). . . . .	27
2.7	Série Temporal Obtida pelo Modelo ARIMA(3, 1, 3). . . . .	28
2.8	Representação do Neurônio Fisiológico. Fonte: (HAYKIN, 2008). . . . .	30
2.9	Exemplo de Arquitetura de ANN em Camadas. . . . .	31
4.1	Autocorrelação da Série definida pela Equação 3.1. . . . .	50
4.2	Autocorrelação Parcial da Série definida pela Equação 3.1. . . . .	51
4.3	Comportamento do Modelo ARIMA para a Série Definida pela Equação 3.1. . . . .	53
4.4	Autocorrelação para Dados Reais do IBOVESPA. . . . .	54
4.5	Autocorrelação Parcial para Dados Reais do IBOVESPA. . . . .	55
4.6	Comportamento do Modelo ARIMA para Dados do IBOVESPA. . . . .	57
4.7	Comportamento da MLP Otimizada para a Série definida pela Equação 3.1. . . . .	59
4.8	Configuração da MLP Simples Otimizada para Dados do IBOVESPA. . . . .	64
4.9	Comportamento da MLP Simples Otimizada para Dados do IBOVESPA. . . . .	64
4.10	Configuração da MLP Exógena Otimizada para Dados do IBOVESPA. . . . .	69
4.11	Comportamento da MLP Exógena Otimizada para Dados do IBOVESPA. . . . .	70

# Lista de Tabelas

4.1	Desempenho do Modelo ARIMA para a Série definida pela Equação 3.1. . .	52
4.2	Desempenho do Modelo ARIMA para Dados do IBOVESPA. . . . .	56
4.3	Desempenho da MLP Simples para a Série definida pela Equação 3.1. . . .	58
4.4	Critérios de Parada da MLP Simples para a Série definida pela Equação 3.1.	58
4.5	Desempenho da MLP Simples para Dados do IBOVESPA e $N_L = 1$ . . . . .	61
4.6	Critérios de Parada da MLP Simples para Dados do IBOVESPA e $N_L = 1$ .	61
4.7	Desempenho da MLP Simples para Dados do IBOVESPA e $N_L = 2$ . . . . .	62
4.8	Critérios de Parada da MLP Simples para Dados do IBOVESPA e $N_L = 2$ .	63
4.9	Desempenho da MLP Exógena para Dados do IBOVESPA e $N_L = 1$ . . . . .	66
4.10	Critérios de Parada da MLP Exógena para Dados do IBOVESPA e $N_L = 1$ .	67
4.11	Desempenho da MLP Exógena para Diferentes Configurações. . . . .	68
4.12	Critérios de Parada da MLP Exógena para Dados do IBOVESPA e $N_L = 2$ .	69
4.13	Desempenho dos Diferentes Modelos para Dados do IBOVESPA. . . . .	69



## Lista de Abreviaturas

<i>AIC</i>	<i>Akaike Information Criterion</i>
<i>AICC</i>	<i>Akaike Information Criterion Corrected</i>
<i>ANN</i>	<i>Artificial Neural Network</i>
<i>ARIMA</i>	<i>Autoregressive Integrated Moving Average</i>
<i>BC</i>	Banco Central
<i>CDI</i>	Certificado de Depósito Interfinanceiro
<i>CETIP</i>	Central de Custódia e de Liquidação Financeira de Títulos
<i>FTDNN</i>	<i>Focused Time Lagged Feedforward</i>
<i>GRNN</i>	<i>Generalized Regression Neural Network</i>
<i>IBGE</i>	Instituto Brasileiro de Geografia e Estatística
<i>IBOVESPA</i>	Índice de Ações da Bolsa de Valores do Estado de São Paulo
<i>IPCA</i>	Índice Nacional de Preços ao Consumidor Amplo
<i>MCP</i>	<i>McCulloch and Pitts</i>
<i>MLP</i>	<i>Multilayer Perceptron</i>
<i>MPE</i>	<i>Mean Percent Error</i>
<i>MSE</i>	<i>Mean Square Error</i>
<i>NAR</i>	<i>Nonlinear Autoregressive</i>
<i>NARX</i>	<i>Nonlinear Autoregressive with External Input</i>
<i>RBF</i>	<i>Radial Basis Function</i>
<i>RNN</i>	<i>Recurrent Neural Network</i>
<i>SIR</i>	<i>Schwarz Information Criterion</i>
<i>TDL</i>	<i>Tapped Delay Line</i>
<i>VAR</i>	Variância



## Lista de Símbolos

$R_{juros}$	Taxa de juros
$R_{inflacao}$	Taxa de inflação
$R_{cambio}$	Taxa de câmbio
$p$	Ordem do modelo AR(p)
$q$	Ordem do modelo MA(q)
$d$	Ordem do modelo I(d)
$\phi_i$	Coefficiente de auto regressão $i$
$\theta_i$	Coefficiente de média móvel $i$
$R_{ag}$	Receita do agente econômico
$G_{ag}$	Gasto do agente econômico
$\mu_X(n)$	Média da variável aleatória $X$ no instante $n$
$\sigma_X^2(n)$	Variância da variável aleatória $X$ no instante $n$
$\rho_X^2(n_1, n_2)$	Autocorrelação da variável aleatória $X$ entre os instantes $n_1$ e $n_2$
$N_{obs}$	Número de observações
$L^i[\cdot]$	Operador de atraso $i$ amostras
$N_{term}$	Número de terminais no neurônio
$\vec{w}$	Vetor de pesos sinápticos
$\lambda$	Termo de <i>bias</i>
$\beta_x$	Centro da função de ativação RBF
$\zeta_x$	Grau de espalhamento da função gaussiana
$N_p$	Número de parâmetros
$N_W$	Número de amostras
$N_L$	Número de camadas
$H_l$	Número de neurônios na camada $l$

# Sumário

<b>Lista de Figuras</b>	<b>vii</b>
<b>Lista de Tabelas</b>	<b>viii</b>
<b>Lista de Abreviaturas e Siglas</b>	<b>ix</b>
<b>Lista de Símbolos</b>	<b>xi</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	2
1.2 Objetivos . . . . .	2
1.3 Organização do Texto . . . . .	3
<b>2 Conceitos Fundamentais</b>	<b>4</b>
2.1 Mercado Financeiro . . . . .	7
2.1.1 Agentes Econômicos e Fluxo Monetário . . . . .	7
2.1.1.1 Mercado Monetário . . . . .	9
2.1.1.2 Mercado de Crédito . . . . .	10
2.1.1.3 Mercado de Câmbio . . . . .	11
2.1.1.4 Mercado de Capitais . . . . .	11
2.1.2 Variáveis Macroeconômicas . . . . .	12
2.1.2.1 Taxa de Juros . . . . .	12
2.1.2.2 Inflação . . . . .	13
2.1.2.3 Taxa de Câmbio . . . . .	14
2.1.2.4 Índice de Ações . . . . .	14
2.2 Teoria de Dow . . . . .	16

---

2.2.1	Movimentação do Mercado . . . . .	16
2.2.2	Tendência do Volume . . . . .	16
2.2.3	Fases da Tendência Primária de Alta . . . . .	16
2.2.4	Fases da Tendência Primária de Baixa . . . . .	17
2.2.5	Relação entre Médias e Informações . . . . .	17
2.2.6	Confirmação Conjunta de Médias . . . . .	17
2.2.7	Indefinição da Tendência . . . . .	18
2.2.8	Aproximação da Média pelo Preço de Fechamento . . . . .	18
2.2.9	Continuidade da Tendência . . . . .	18
2.3	Séries Temporais . . . . .	18
2.3.1	Processos Estocásticos e Séries Temporais . . . . .	19
2.3.1.1	Estacionariedade e Ergodicidade . . . . .	20
2.3.2	Decomposição de Séries Temporais . . . . .	21
2.3.3	Modelo ARIMA para Séries Temporais . . . . .	21
2.3.3.1	Modelo AR . . . . .	23
2.3.3.2	Modelo MA . . . . .	24
2.3.3.3	Modelo ARMA . . . . .	25
2.3.3.4	Modelo I . . . . .	25
2.3.3.5	Modelo ARIMA . . . . .	27
2.3.3.6	Metodologia Box-Jenkins . . . . .	28
2.4	Inteligência Artificial Conexionista . . . . .	29
2.4.1	Neurônio Artificial . . . . .	31
2.4.1.1	Algoritmo de Treinamento . . . . .	33
2.4.2	Arquitetura <i>Multilayer Perceptron</i> . . . . .	35
2.4.3	Arquitetura <i>Radial Basis Function</i> . . . . .	36
2.4.4	Arquitetura <i>Generalized Regression Neural Network</i> . . . . .	37
2.4.5	Arquitetura <i>Focused Time Lagged Feedforward</i> . . . . .	37
2.4.6	Arquitetura <i>Recurrent Neural Network</i> . . . . .	38
2.5	CrITÉRIOS de Desempenho em Predições . . . . .	39
<b>3</b>	<b>Modelos Propostos</b> . . . . .	<b>42</b>
3.1	Descrição do Problema . . . . .	43
3.2	Modelo ARIMA . . . . .	44

---

3.3	Modelo ANN . . . . .	44
3.3.1	Modelo MLP Simples . . . . .	47
3.3.2	Modelo MLP Exógena . . . . .	47
<b>4</b>	<b>Análise dos Resultados</b>	<b>49</b>
4.1	Resultados do Modelo ARIMA . . . . .	50
4.1.1	Série Temporal de Teste . . . . .	50
4.1.2	Dados Reais do IBOVESPA . . . . .	53
4.2	Resultados do Modelo MLP Simples . . . . .	56
4.2.1	Série Temporal de Teste . . . . .	56
4.2.2	Dados Reais do IBOVESPA . . . . .	60
4.3	Resultados do Modelo MLP Exógeno . . . . .	65
4.4	Comparação entre os Modelos . . . . .	69
<b>5</b>	<b>Conclusão</b>	<b>72</b>
5.1	Trabalhos Futuros . . . . .	73
	<b>Referências Bibliográficas</b>	<b>77</b>
	<b>Apêndices</b>	<b>77</b>

# Capítulo 1

## Introdução

Atualmente, a utilização de técnicas de Processamento Digital de Sinais e Computação Evolutiva abarcam diferentes contextos, tais como alocação de recursos em comunicações, navegação de robôs móveis, análise de investimentos econômicos etc (MONTGOMERY; JENNINGS, 2008). Entre as aplicações destas técnicas, destaca-se a predição de valores em uma série temporal.

O problema de predição de valores, *i.e.* estudo de séries temporais, representa um importante desafio para a sociedade moderna, pois são diversas as situações onde o processo de tomada de decisão é demasiadamente influenciado por estimativas da probabilidade de ocorrência de determinados eventos futuros (MORETTIN; TOLOI, 2006).

Assim, pode-se definir o problema de predição considerando a série temporal desejada, a partir desta, deseja-se estimar seu valor para os instantes futuros ainda não-observados, com base no conjunto de observações passadas e presente.

De particular importância a este problema, ressalta-se a necessidade de dispor de predições referentes aos indicadores econômicos, em especial, destaca-se o valor do índice de ações (ROQUE, 2009). A relevância da predição na área econômica resulta da imprescindível demanda por escolhas entre os investimentos disponíveis, sendo que o resultado final atingido em uma determinada data futura só pode, no máximo, ser estimado, *i.e.* aproximado, no presente.

Neste contexto, o presente trabalho tem por objetivo representar o Índice de Ações da Bolsa de Valores do Estado de São Paulo (IBOVESPA) mediante uma série temporal e desenvolver um modelo preditivo (BMFBOVESPA, 2016), baseado em Processamento Digital

de Sinais aliado à Computação Evolutiva e à Teoria de Informação. O IBOVESPA surge na economia brasileira como o principal índice de bolsa de valores e, conseqüentemente, como essencial medidor de atividade econômica no Brasil. Entre as técnicas estudadas ao longo desta obra, em particular, destaca-se o modelo estatístico clássico denominado *Autoregressive Moving Average* (ARIMA) (BOX et al., 2015) e a técnica bio-inspirada da Inteligência Artificial Conexionista denominada *Artificial Neural Network* (ANN) (HAYKIN, 2008).

## 1.1 Motivação

A incerteza em relação ao futuro tem despertado interesse da sociedade desde a antiguidade (MORETTIN; TOLOI, 2006), pois decisões tomadas no presente exercem influência em seus efeitos no futuro, por exemplo, a decisão acerca do plantio de grãos é embasada na previsão de uma série de parâmetros futuros como índice pluviométrico durante o plantio, preço do grão na data de colheita etc.

Outra aplicação a ressaltar nesta obra compreende a política macroeconômica exercida pelo governo, em que a previsão da inflação afeta diretamente a taxa de juros praticada pelo Banco Central e, conseqüentemente, toda a economia.

Por fim, destaca-se a previsão do IBOVESPA, que é um importante insumo no processo de tomada de decisão acerca da alocação de investimentos por parte da população, gestão de negócios por parte dos empresários e avaliação da regulação e incentivo econômico aos diversos setores da economia por parte do governo.

## 1.2 Objetivos

O objetivo geral do presente trabalho corresponde ao projeto de modelos para predição do IBOVESPA implementados em Matlab.

Além disso, tem-se os seguintes objetivos específicos:

1. Estudo de métodos de processamento de sinais, tais como ARIMA;
2. Estudo de técnicas de computação evolutiva, tais como ANN;
3. Implementação de algoritmos em Matlab;

4. Comparação das técnicas de predição mediante diferentes métricas objetivas;
5. Aplicação dos conceitos aprendidos nas disciplinas do curso de Engenharia de Informação, em particular: Sinais Aleatórios, Processamento Digital de Sinais, Comunicações Multimídia, Teoria de Informação e Códigos, Redes de Computadores etc.

### 1.3 Organização do Texto

O presente trabalho visa estudar e analisar diferentes aspectos envolvidos na predição de séries temporais. Neste sentido, este texto foi organizado da seguinte forma. No Capítulo 2 introduz-se conceitos fundamentais associados à predição de valores, incluindo a caracterização estatística de uma série temporal, bem como métodos estatísticos e bio-inspirados aplicados ao problema de predição. No Capítulo 3 apresenta-se os diferentes modelos de predição utilizados. Já no Capítulo 4 discute-se os resultados obtidos por cada modelo. Por fim, o Capítulo 5 encerra esta obra apresentando as conclusões obtidas.

# Capítulo 2

## Conceitos Fundamentais

*We have the duty of formulating, of summarizing, and of communicating our conclusions, in intelligible form, in recognition of the right of other free minds to utilize them in making their own decisions.*

—Ronald A. Fisher

A aplicação de técnicas computacionais em problemas de predição de valores futuros de séries temporais tem despertado o interesse da comunidade científica, bem como de entidades governamentais e de empresas do setor privado (MORETTIN; TOLOI, 2006). Entre as situações em que a predição de valores futuros de uma série temporal é tida como informação estratégica ao processo de tomada de decisão, pode-se destacar:

- Demanda de uso da banda de transmissão por parte dos assinantes de uma rede de comunicação *wireless*, de forma a estabelecer a alocação eficiente de canais;
- Índice pluviométrico para planejamento do plantio e colheita de grãos;
- Preço da ação de uma determinada empresa negociada em bolsa de valores na construção de um portfólio de investimentos.

Em geral, tem-se à disposição um conjunto de valores ordenados no tempo até determinado instante e a partir dos pontos observados, pode-se extrapolar o horizonte de

alcance/decisão de tal modo a prever valores em uma data futura com uma determinada probabilidade (MORETTIN; TOLOI, 2006), ou seja, com base no modelo proposto para o fenômeno de interesse, supõe-se que existe autocorrelação entre os valores passados e presentes da série temporal sob análise e com isso pode-se estimar seus valores futuros.

Segundo (MORETTIN; TOLOI, 2006), a solução para o problema de previsão discutido acima pode ser obtida de acordo com as seguintes abordagens:

1. Fundamentação na teoria subjacente;
2. Orientação a dados.

No estudo guiado pelo método fundamentado na teoria subjacente, recorre-se ao método científico, sendo possível sumariá-lo nas seguintes etapas sequenciais:

1. Observação dos efeitos do fenômeno de interesse;
2. A partir da teoria existente relacionada ao fenômeno, realiza-se a formulação de um modelo matemático adequado à sua explicação;
3. Submissão do modelo a testes, em que valida-se a resposta do modelo às novas observações, caso o desempenho do modelo não seja satisfatório, retorna-se à primeira etapa, refinando-o com a introdução de novas hipóteses;
4. Realização da previsão de valores futuros do fenômeno estudado com base no modelo formulado.

Esta abordagem apresenta como ponto negativo, o fato de requerer do projetista responsável pela modelagem, conhecimentos específicos acerca do fenômeno de interesse, sendo inconveniente pois: (i) para cada aplicação, será necessário que o projetista obtenha conhecimentos específicos acerca do fenômeno a ser modelado e (ii) a fundamentação teórica necessária pode ainda não ter sido desenvolvida no momento ou ser proibitivamente complexa para aplicações práticas.

De modo a superar as limitações expostas acima, o estudo orientado a dados desconsidera aspectos fenomenológicos inerentes da série temporal em estudo, sendo necessário somente o acesso aos dados de entrada e saída extraídos do fenômeno (SHAN et al., 2015). Com base nos dados obtidos, aplica-se procedimentos estatísticos e/ou computacionais,

em que é possível visualizar as correlações existentes entre as variáveis sob análise. Desta forma, formula-se hipóteses acerca da estrutura do modelo explicativo para o fenômeno em análise. Por fim, o modelo selecionado é aplicado à predição de valores futuros da série resultante do fenômeno sob estudo. Esta sistemática de modelagem deu origem ao campo de estudo denominado Ciência de Dados (SHAN et al., 2015).

Visando otimizar os resultados, é possível combinar os dois métodos expostos de forma a construir modelos que agregam técnicas estatísticas e computacionais, aliadas aos conceitos teóricos relacionados ao fenômeno, *e.g.* na seleção de variáveis endógenas, hipótese inicial da relação causa-efeito entre as variáveis do modelo etc. A Econometria surge como exemplo de área do conhecimento em que esta junção de abordagens é utilizada. No campo da Econometria, faz-se uso da Teoria Econômica para concepção de hipóteses acerca da estrutura geral do sistema econômico de interesse, bem como de suas variáveis relevantes. Após a etapa de concepção, aplica-se procedimentos estatísticos e computacionais, cujo objetivo é alinhar o resultado esperado pela teoria com aquele observado na prática (MORETTIN; TOLOI, 2006).

Neste trabalho optou-se pelo método orientado a dados para análise e predição de séries temporais, em que fez-se uso da formulação estatística de séries temporais aliada à aplicação de algoritmos computacionais oriundos dos seguintes campos: (i) Processamento Digital de Sinais, (ii) Inteligência Artificial e (iii) Teoria de Informação, no levantamento de heurísticas que otimizem a definição da arquitetura do modelo. Com base neste modelo, efetua-se a predição de valores futuros da série temporal de interesse.

O presente capítulo é estruturado da seguinte maneira. Na Seção 2.1 tem-se um resumo dos principais conceitos associados ao mercado financeiro necessários para a compreensão deste trabalho. A Seção 2.2 contém a descrição dos postulados que compõem a Teoria de Dow, utilizada na análise do mercado financeiro. Na Seção 2.3 elicitam-se alguns aspectos fundamentais da abordagem clássica para análise de séries temporais. Na Seção 2.4 efetua-se uma introdução às técnicas bio-inspiradas oriundas da Inteligência Artificial. Por fim, na Seção 2.5 descreve-se algumas métricas de desempenho aplicadas aos modelos de predição relevantes no escopo deste trabalho.

## 2.1 Mercado Financeiro

Nesta seção é feita uma breve introdução aos conceitos básicos relacionados ao mercado financeiro. Inicialmente, tem-se a definição do mercado financeiro, sua composição, bem como dos principais mecanismos de fluxo financeiro entre seus agentes constituintes. Em seguida faz-se uma descrição de algumas variáveis macroeconômicas relevantes para a realização deste trabalho, ressaltando a correlação existente entre estas.

### 2.1.1 Agentes Econômicos e Fluxo Monetário

Uma economia é, fundamentalmente, constituída por dois agentes econômicos (ASSAF, 1999), diferenciados por suas realidades orçamentárias:

1. Poupadores de recursos;
2. Tomadores de recursos.

Denomina-se poupador de recursos, aquele agente em que, ao final de determinado período de tempo, apresenta excesso de recursos financeiros, *i.e.* a quantia recebida é maior do que aquela gasta durante o período em estudo. A esta diferença positiva entre o fluxo financeiro gasto, subtraído daquele arrecadado, dá-se o nome de superávit.

Em complemento ao agente poupador, existe o agente tomador de recursos. Este por sua vez é caracterizado por apresentar falta de recursos financeiros, *i.e.*, a quantia recebida é menor do que aquela gasta durante o período em estudo. A esta diferença negativa entre o fluxo financeiro gasto, subtraído daquele arrecadado, dá-se o nome de déficit.

Em termos simbólicos, sejam  $R_{ag}$  e  $G_{ag}$  a receita obtida por determinado agente e seu gasto, respectivamente, ambos em unidades monetárias (*e.g.* Reais), então a Equação 2.1 define a relação entre  $R_{ag}$  e  $G_{ag}$  e a classificação do agente econômico.

$$\begin{aligned} R_{ag} - G_{ag} > 0 &\Rightarrow \text{Poupador de recursos} \\ R_{ag} - G_{ag} < 0 &\Rightarrow \text{Tomador de recursos} \end{aligned} \tag{2.1}$$

Analisando os perfis orçamentários destes dois agentes, constata-se que, o agente poupador dispõe de recursos excedentes em estado ocioso, enquanto o tomador necessita de recursos para arcar com suas obrigações financeiras.

Este arranjo é considerado ineficiente (PINHO; VASCONCELLOS, 2011), pois os recursos ociosos do poupador têm seu poder de compra reduzido com o decorrer do tempo por efeito da inflação, enquanto os tomadores são impossibilitados de cumprir com suas obrigações financeiras devido à ausência de recursos. Desta forma, para atingir o equilíbrio entre seus orçamentos, estes agentes recorrem ao mercado financeiro, onde agentes poupadores ofertam seus recursos excedentes a serem consumidos pelos agentes tomadores que os demandam (ASSAF, 1999).

Neste contexto, pode-se definir o mercado financeiro como sendo o meio, abstrato ou físico, onde forças de oferta e demanda por recursos equilibram-se (PINHO; VASCONCELLOS, 2011). Dito de outro modo, é o centro em que ocorre a transferência de recursos de poupadores para tomadores (FORTUNA, 2005). A Figura 2.1 ilustra este conceito, onde as setas definem o sentido do fluxo de recursos.



Figura 2.1: Fluxo de Recursos no Mercado Financeiro.

O fluxo de oferta e demanda de recursos no mercado financeiro dá-se pelo seguinte processo (ASSAF, 1999):

1. Poupador oferta a quantia  $P_1$ ;
2. Tomador consome esta quantia no instante  $t_1$ ;
3. Em um instante posterior  $t_2 > t_1$ , o tomador entrega ao poupador a quantia  $P_2$ , formada pela soma das parcelas  $P_1$  e da remuneração exigida pelo tomador por

abdicar de seus recursos durante o período compreendido entre  $t_1$  e  $t_2$ , com isso finaliza-se o negócio.

Esta remuneração do capital de terceiros é denominada juros (ASSAF, 1999), sendo geralmente representada por uma taxa em base percentual, cujo símbolo adotado neste trabalho é  $R_{juros}$ , é obtida mediante a Equação 2.2.

$$R_{juros} = \frac{P_2 - P_1}{P_1} \cdot 100 \quad (\%) \quad (2.2)$$

Pode-se conceituar o juros como sendo um fator de atualização entre duas quantias monetárias separadas por determinado horizonte de tempo. Dito de outra maneira, como sendo o custo do capital de terceiros, cedido mediante contrato de empréstimo estabelecido diretamente, ou via intermediário financeiro, entre o poupador de recursos e o tomador de recursos (PINHO; VASCONCELLOS, 2011).

De maneira a simplificar o entendimento do mercado financeiro, o mesmo é segregado em quatro mercados distintos e complementares com base no perfil dos recursos negociados e de seus prazos de devolução do tomador ao poupador (ASSAF, 1999):

- Mercado Monetário;
- Mercado de Crédito;
- Mercado de Câmbio;
- Mercado de Capitais.

Ressalta-se que nas referências utilizadas, esta separação é feita unicamente por questões didáticas, dado que, na prática, a linha que divide um mercado de outro é tênue e de difícil identificação (ASSAF, 1999).

#### 2.1.1.1 Mercado Monetário

É no mercado monetário onde são executadas as operações de transferência de recursos de curtíssimo prazo, em geral com validade de uma noite (ASSAF, 1999). É neste mercado que o Banco Central (BC) exerce sua política monetária, por meio da qual regula a oferta de moeda na economia e exerce o controle da inflação do país.

Como exemplo de transação típica deste mercado, cita-se a equalização das posições de caixa de instituições financeiras participantes ao final do dia (FORTUNA, 2005). Neste tipo de operação, uma instituição encerra o dia com o valor em caixa superior ao mínimo estipulado por órgãos reguladores e por seu conselho deliberativo. De maneira complementar, tem-se uma instituição que encerrou seu dia com valor em caixa inferior ao mínimo definido. Para equilibrar suas posições de caixa, *i.e.* a instituição com superávit  $I_{sup}$  emprestar seus recursos excedentes para a instituição em déficit  $I_{def}$ , as mesmas recorrem ao mercado monetário, em particular ao mercado interfinanceiro, procedendo da seguinte forma:

1. A instituição  $I_{def}$  emite um título de dívida denominado Certificado de Depósito Interfinanceiro (CDI), mediante o qual confere ao detentor de tal título, o direito de receber o valor que emprestou acrescido da parcela de juros ao final do período contratual;
2. A instituição  $I_{sup}$  recebe o contrato de CDI e em troca empresta o valor estipulado no contrato à taxa de juros previamente pactuada;
3. Ao final do período, a instituição  $I_{sup}$  devolve a posse do CDI para a instituição  $I_{def}$ , retornando o valor previamente tomado por empréstimo, acrescido da parcela referente aos juros acumulados no período de validade do contrato.

Sendo  $D_0$  a data em que a operação teve início, se a mesma tem vencimento em  $D_0 + 1$ , *i.e.* a operação é encerrada um dia útil após ter sido aberta, dá-se o nome de empréstimo *overnight* (FORTUNA, 2005). A média ponderada das taxas dos contratos CDI negociadas neste prazo é de particular importância na formação econômica do Brasil, sendo denominada taxa de Depósito de Interfinanceiro *Overnight*, ou simplesmente DI (CETIP, 2016). Seu valor é diariamente apurado e divulgado pela Central de Custódia e de Liquidação Financeira de Títulos (CETIP) (CETIP, 2016). O DI é uma variável de fundamental relevância nesta obra, pois seu valor afeta diretamente a rotina das empresas e, com efeito, suas ações e, conseqüentemente, o índice da bolsa de valores.

### 2.1.1.2 Mercado de Crédito

No mercado de crédito são feitas transações de curto e médio prazo, tipicamente no horizonte de semanas até meses (ASSAF, 1999).

Em geral, são realizadas operações entre instituições financeiras e pessoas físicas e/ou jurídicas, em que a instituição financeira pode ser bancária ou não. Uma instituição é dita bancária se for apta a captar recursos mediante depósitos à vista (FORTUNA, 2005). Como exemplo de instituições financeiras não-bancárias, pode-se citar as associações financeiras, sociedades de arrendamento mercantil, cooperativas de crédito etc.

### 2.1.1.3 Mercado de Câmbio

Cada país, ou grupo de países, dispõem de uma moeda própria, por exemplo o Real (BRL) no Brasil, o Dólar (USD) nos Estados Unidos da América (EUA), o Euro (EUR) na Zona do Euro etc (FORTUNA, 2005).

Devido a natureza internacional da economia moderna (PINHO; VASCONCELLOS, 2011), surge a necessidade de transacionar-se a moeda de um país pela moeda de outro país. Por exemplo, um investidor localizado no EUA e dispõe de Dólares, mas que deseja investir no Brasil necessita de Reais para realizar seu investimento.

A moeda pode ser considerada uma mercadoria (FORTUNA, 2005) e, portanto, é possível estabelecer seu preço, denominado taxa de câmbio. A taxa de câmbio expressa a razão entre duas moedas (ASSAF, 1999), *e.g.*, supondo que a taxa de câmbio  $BRL/USD$  seja 3,00, isto significa que, para adquirir 1 unidade de Dólar (USD) é necessário dispor de 3 unidades de Real (BRL).

Para este trabalho, a taxa de câmbio entre Dólar e Real, representa uma das variáveis relevantes de análise, sendo melhor elucidada em seção posterior.

### 2.1.1.4 Mercado de Capitais

Conforme (FORTUNA, 2005), o objetivo do mercado de capitais reside nas operações de longo prazo, atuando no segmento em que instituições bancárias não apresentam interesse em estarem presentes.

Mediante o mercado de capitais, empresas podem angariar fundos para expansão e manutenção de suas atividades produtivas com a emissão de títulos diretamente ao investidor. Desta forma, dispensa-se a necessidade do intermediário bancário, reduzindo o custo da captação (FORTUNA, 2005), pois, dado o longo prazo, a reumenaração bancária que seria exigida, definida com base na diferença entre a taxa de captação de recursos e a de empréstimo de recursos, a operação pode tornar-se economicamente inviável (ASSAF,

1999). Além disso, o mercado de capitais possibilita que o investidor do tipo pessoa física faça parte de negócios produtivos, tornando-se sócio de empresas e tendo seu rendimento atrelado à lucratividade das mesmas ao longo do tempo.

A principal forma de captação de recursos no mercado de capitais dá-se pela emissão de títulos denominados ações (ASSAF, 1999). Uma ação representa a menor parcela do capital social de uma empresa e confere ao detentor da mesma, direitos societários sobre a empresa emissora (FORTUNA, 2005). Entre estes direitos, tem-se a participação nos lucros mediante a distribuição de dividendos, participação do processo de tomada de decisão de negócio em assembleias etc.

É no mercado de capitais que o IBOVESPA é construído, sendo este o objeto de estudo do presente trabalho a ser detalhado em seção posterior.

### 2.1.2 Variáveis Macroeconômicas

Ao modelar um sistema econômico, faz-se a escolha de um conjunto de variáveis a serem explicadas pelo modelo, a estas variáveis dá-se o nome de variáveis explicadas ou endógenas (PINHO; VASCONCELLOS, 2011). O valor assumido por estas variáveis é obtido a partir da aplicação de uma função  $f(\cdot)$ , que considera em sua entrada outro conjunto de variáveis, denominadas variáveis explicativas ou exógenas, que possuem determinada capacidade de explicar o comportamento das variáveis endógenas de interesse.

Por exemplo, supondo que, com base na taxa de juros  $R_{juros}$ , índice de inflação  $R_{inflacao}$  e taxa de câmbio  $R_{cambio}$ , deseja-se estimar o valor de determinado índice de ações  $P_{acoes}$ . Esta relação pode ser expressa a partir da Equação 2.3.

$$P_{acoes} = f(R_{juros}, R_{inflacao}, R_{cambio}) \quad (2.3)$$

Onde  $f(\cdot)$  representa alguma relação funcional entre os valores das variáveis exógenas e a estimativa do valor da variável endógena.

#### 2.1.2.1 Taxa de Juros

Um dos conceitos fundamentais da Economia Monetária (PINHO; VASCONCELLOS, 2011), diz que não é possível comparar a mesma quantia monetária em dois instantes distintos de tempo. Sendo assim, é necessário atualizar uma destas quantias por determinado fator

de compensação temporal de modo a serem comparáveis pontualmente no eixo do tempo. Por exemplo, uma unidade de Real  $R\$1,00$  na data presente  $D_0$  será diferente de uma unidade de Real em uma data futura  $D_0 + n$ , em que  $n > 0$  representa o número de unidades de tempo transcorridas desde  $D_0$ .

A taxa de juros  $R_{juros}$ , também conhecida como custo do capital, representa este fator de atualização. Define-se a taxa de juros como sendo a remuneração mínima que o agente poupador aceita ao ceder o direito de utilização de determinada quantia para o agente tomador de recursos por um determinado período de tempo (PINHO; VASCONCELLOS, 2011).

Em um regime de capitalização composta (WALMSLEY, 1998), a Equação 2.4 define a relação entre o valor futuro  $F_v$  de certa quantia na data atual  $P_v$ , transcorridos  $n$  períodos de tempo, incorrendo à taxa de juros  $R_{juros}$  durante o período.

$$F_v = P_v(1 + R_{juros})^n \quad (2.4)$$

A taxa de juros é uma das variáveis que exercem influência no desempenho da economia produtiva e por consequência nas ações das empresas listadas em bolsa de valores (PINHO; VASCONCELLOS, 2011).

Por exemplo, ao elevar a taxa de juros, torna-se mais rentável o investimento em títulos de dívida e, com efeito, reduz-se a atratividade do investimento no meio produtivo. Portanto impacta-se no preço das ações, reduzindo-os. Dito de outro modo, a taxa de juros apresenta relação, aproximadamente, inversa com o preço das ações (KEYNES, 2016).

### 2.1.2.2 Inflação

A inflação  $R_{inflacao}$  representa o aumento persistente e generalizado do nível de preços em uma economia (PINHO; VASCONCELLOS, 2011). Em um processo inflacionário, todos os bens comercializados no país sofrem aumento de preços que mantém-se em elevação sustentada após determinado período de tempo.

Para mensurar a inflação de uma economia de forma quantitativa, foram propostos os índices de inflação (PINHO; VASCONCELLOS, 2011). Estes índices são calculados pela variação da soma ponderada dos preços de um conjunto de bens entre sucessivos instantes de tempo.

Através dos índices de inflação, o setor empresarial pode planejar sua produção e o governo pode exercer sua política monetária de modo a manter o processo inflacionário no patamar desejado.

No Brasil, o índice de inflação oficial é o Índice Nacional de Preços ao Consumidor Amplo (IPCA), cuja mensuração e divulgação é de responsabilidade do Instituto Brasileiro de Geografia e Estatística (IBGE) desde 1980 (IBGE, 2016). O IPCA abrange o custo de famílias com rendimentos mensais compreendidos entre um e quarenta salários-mínimos, qualquer que seja a fonte de rendimentos e residentes em áreas urbanas (IBGE, 2016).

### 2.1.2.3 Taxa de Câmbio

A taxa de câmbio  $R_{cambio}$  expressa a relação entre duas moedas (PINHO; VASCONCELLOS, 2011), *i.e.* o preço de uma moeda (*e.g.* Real) expresso através de unidades de outra moeda (*e.g.* Dólar) (ASSAF, 1999).

Assim como a taxa de juros, a taxa de câmbio representa uma variável econômica relevante ao setor produtivo. Pois, se uma empresa importa bens do exterior, esta empresa incorre em débitos expressos em moeda estrangeira, por exemplo o Dólar e se a taxa de câmbio aumenta no período, a dívida desta empresa aumenta na base de sua moeda local, por exemplo o Real. Desta maneira, o preço da ação desta empresa tende a diminuir e com isso impactar no índice de ações que engloba esta ação específica.

### 2.1.2.4 Índice de Ações

Com o objetivo de avaliar o desempenho global das ações de uma economia, foram desenvolvidos os índices de ações (ASSAF, 1999). Um índice de ação compreende a soma ponderada dos preços de um conjunto de ações selecionadas através de determinado critério, fornecendo uma estimativa para a tendência e a variação do comportamento da economia nacional ao longo do tempo (FORTUNA, 2005).

No caso específico deste estudo, objetiva-se predizer o valor futuro do preço de fechamento do IBOVESPA. Sendo que, este índice é calculado e divulgado todos os dias úteis pela bolsa de valores oficial de São Paulo (BMFBOVESPA, 2016). Sua formulação teve início no ano de 1968, a partir da concepção de uma carteira teórica de ações (ASSAF, 1999), atribuiu-se o valor de 100 pontos na data de sua criação, desde então não foram feitas alterações em sua estrutura geral. A leitura do IBOVESPA, dá-se a partir da variação de

seus pontos entre dois instantes de tempo, esta leitura indica a direção do grupo de ações que fazem parte do índice no período de tempo considerado, oferecendo uma visão geral do comportamento da economia nacional de forma agregada.

A participação de cada ação no IBOVESPA, tem relação direta com a representatividade deste título no mercado a vista em termos de número de negócios e do volume financeiro gerado, sendo atualizada a cada doze meses (ROQUE, 2009). Essa representatividade é obtida pelo índice de negociabilidade de ação  $k$ ,  $I_k$ , calculado pela Equação 2.5 (ROQUE, 2009).

$$I_k = \sqrt{\frac{N_k \cdot V_k}{N_T \cdot V_T}} \quad (2.5)$$

Onde:

$N_k$  Representa o número de negócios realizados com a ação  $k$  no mercado a vista nos últimos doze meses;

$N_T$  Representa o número total de negócios no mercado a vista nos últimos doze meses;

$V_k$  Representa o valor, em moeda corrente, movimentado com a ação  $k$  no mercado a vista nos últimos doze meses;

$V_T$  Representa o valor total, em moeda corrente, movimentado no mercado nos últimos doze meses.

O cálculo do valor do IBOVESPA no instante  $n$ ,  $I_{ibov}(n)$ , é feito através do somatório dos produtos entre preço da ação e seu peso no índice. Esta relação é expressa na Equação 2.6 (ROQUE, 2009).

$$I_{ibov}(n) = \sum_{k=1}^{N_{acoes}} P_k(n) \cdot Q_k(n) \quad (2.6)$$

Onde:

$N_{acoes}$  Representa o número total de ações componentes do IBOVESPA;

$P_k(n)$  Representa o preço da ação  $k$  no instante  $n$ ;

$Q_k(n)$  Representa o fator de ponderação da ação  $k$  no instante  $n$ .

## 2.2 Teoria de Dow

Ao observar o comportamento do mercado de ações no século XX, Charles Dow verificou um conjunto de propriedades gerais nos movimentos realizados pelos ativos estudados (FORTUNA, 2005). Tais propriedades foram sumarizadas em nove postulados, conjuntamente denominados Teora de Dow (PINHEIRO, 2014).

### 2.2.1 Movimentação do Mercado

Em relação ao tempo de duração, o mercado possui três tendências:

1. Primária - É a mais longa entre as três, com duração de meses até anos;
2. Secundária - Possui duração intermediária, variando de semanas até meses;
3. Terciária - Tendência de menor duração, compreendendo o período de dias até semanas.

### 2.2.2 Tendência do Volume

A quantidade de ativos negociados no mercado, *i.e.* o volume negociado, deve acompanhar os preços. Por exemplo, em uma tendência de alta, o volume negociado deve aumentar conforme os preços aumentam e decair conforme os preços decaem.

### 2.2.3 Fases da Tendência Primária de Alta

A tendência primária de alta, é composta por três fases sequências, são elas:

1. Acumulação - Agentes dotados de maior quantidade de informações detectam a subvalorização dos preços e iniciam o processo de compra, fazendo com que os preços subam. Entretanto, tal subida não é percebida pelos demais agentes do mercado;
2. Subida - A informação é difundida amplamente e os preços iniciam o processo de alta acentuada. Este movimento é ocasionado pelas compras dos demais agentes que até então, estavam fora do mercado;

3. Euforia - Os agentes dotados de maior quantidade de informações, que iniciaram o processo de alta, detectam que os preços encontram-se sobrevalorizados e iniciam o processo de venda. Com certo atraso, os demais agentes percebem o movimento e dão início ao fluxo exacerbado de vendas, com isso os preços decaem intensamente.

### **2.2.4 Fases da Tendência Primária de Baixa**

A tendência primária de baixa é composta por três fases sequências, são elas:

1. Distribuição - Após os preços atingirem níveis de máxima. Com a demanda por ativos ainda em alta, as ofertas de venda são finalizadas gradualmente por agentes com maior disposição de informações e percepção da sobrevalorização dos preços;
2. Pânico - Os demais agentes do mercado iniciam o fluxo de venda de maneira intensa levando os preços a baixas significativas. Enquanto isso, os agentes que iniciaram o processo de baixa diminuem suas vendas;
3. Desaceleração - Devido a inércia das posições, os preços estabilizam-se e tem-se início a uma nova fase de acumulação.

### **2.2.5 Relação entre Médias e Informações**

Os preços dos ativos em uma data presente, são reflexos de sua história, bem como da informação presente no momento. Exceção a esta regra deve-se a perturbações anormais, em geral com baixa probabilidade de ocorrência, por exemplo ataques terroristas, catástrofes naturais etc.

### **2.2.6 Confirmação Conjunta de Médias**

Os preços de ativos distintos, mas pertencentes ao mesmo setor econômico devem possuir a mesma direção de alta ou baixa, se for suposto que existe tendência de alta ou de baixa.

Quando Charles Dow elaborou este princípio, a economia dos EUA, era composta pela atividade industrial e de transportes. Neste contexto, para comprovar que existe uma tendência de alta no mercado, tem-se a necessidade de que, se o preço das ações referentes às indústrias subirem, as ações referentes ao transporte devem subir conjuntamente.

Caso uma das médias falhe em ocorrer, isto pode ser entendido como um sinal de alerta para a potencial reversão da tendência vigente em um momento posterior.

### **2.2.7 Indefinição da Tendência**

Além do processo de alta e/ou baixa, existe uma terceira possibilidade, em que os preços não apresentam tendência definida, mantendo-se constantes. Esta situação decorre do balanceamento de forças de compra e venda atuando no ativo em questão, promovendo o estado de equilíbrio.

### **2.2.8 Aproximação da Média pelo Preço de Fechamento**

Os preços de fechamento caracterizam o consenso estabelecido pelos agentes econômicos acerca da tendência desenvolvida ao longo das negociações realizadas durante o dia. Logo, pode-se aproximar o preço de fechamento como sendo o representante do valor do índice de ações.

### **2.2.9 Continuidade da Tendência**

De forma a evitar potenciais ruídos nos preços dos ativos, deve-se considerar somente sinais explícitos para caracterizar a reversão da tendência vigente. Segundo pesquisas realizadas, somente o preço de fechamento acima de um topo ou fundo anterior, é capaz de caracterizar uma reversão (PINHEIRO, 2014).

## **2.3 Séries Temporais**

Inicialmente, a solução do problema de predição era posta mediante aplicação de procedimentos puramente estatísticos que consideram os dados como uma função amostra representativa de um processo estocástico subjacente (MORETTIN; TOLOI, 2006). A partir deste modelo, encontra-se as correlações entre valores passados, presente e futuros nos dados. Com isso, efetua-se a predição dos valores nos instantes futuros desejados.

Neste contexto, a presente seção descreve aspectos fundamentais de séries temporais e sua utilização na abordagem clássica para o problema de predição de valores.

### 2.3.1 Processos Estocásticos e Séries Temporais

Enquanto um processo determinístico representa a evolução do sistema ao longo do tempo quando não há incerteza acerca das transições entre estados, um processo estocástico caracteriza a evolução do sistema quando transições entre estados é carregada de incerteza (KAY, 1993). Com base nesta concepção, pode-se conceituar um processo estocástico como sendo uma sequência de variáveis aleatórias indexadas no tempo (MONTGOMERY; JENNINGS, 2008).

Matematicamente, um processo estocástico de tempo discreto é representado pela função  $S_X(X = x; n)$ , com  $n \in \mathbb{Z}$ , em que:

$x$  Representa o valor assumido pela variável aleatória  $X$ ;

$n$  Representa o indexador de determinado conjunto de inteiros.

A caracterização completa de um processo estocástico de tempo discreto é feita mediante sua função de distribuição de probabilidade  $N$ -dimensional  $F(x_1, \dots, x_N; n_1, \dots, n_N)$ , expressa pela Equação 2.7 (KAY, 1993).

$$F(x_1, \dots, x_N; n_1, \dots, n_N) = P_r(X(n_1) \leq x_1, \dots, X(n_N) \leq x_N) \quad (2.7)$$

Entretanto, ressalta-se que são poucos os fenômenos em que a função de distribuição é conhecida (KAY, 1993). Desta forma, utiliza-se comumente sua representação parcial, mediante os momentos da função de distribuição, em particular: (i) média  $\mu_X(n)$ , (ii) variância  $\sigma_X^2(n)$  e (iii) função de autocorrelação  $\rho_X(n_1, n_2)$ . Estas estatísticas são descritas pelas equações 2.8, 2.9 e 2.10 (KAY, 1993), respectivamente.

$$\mu_X(n) = E_X[x(n)] = \sum_{i=0}^N x_i(n) \cdot P_r[x_i(n)] \quad (2.8)$$

$$\sigma_X^2(n) = VAR_X[x(n)] = E[x^2(n)] - E[x(n)]^2 \quad (2.9)$$

$$\rho_X(n_1, n_2) = Corr_X[x(n_1), x(n_2)] = \frac{E[x_1(n) \cdot x_2(n)] - E[x_1(n)] \cdot E[x_2(n)]}{\sqrt{\sigma_X^2(n_1)\sigma_X^2(n_2)}} \quad (2.10)$$

Diversos fenômenos são modelados mediante o arcabouço teórico de processos estocásticos, cita-se como exemplo (MORETTIN; TOLOI, 2006):

- Quantidade de ligações em uma central telefônica por minuto;
- Potência sonora em um sistema de compressão de voz;
- Índice de ações na bolsa de valores.

Ao considerar o processo estocástico  $S_X(X = x; n)$ , fixando-se o indexador  $n = n_0$ , tem-se  $S_X(X = x; n = n_0)$ , que representa uma variável aleatória  $X$ . Enquanto que, fixando-se  $X = x_0$ , obtem-se uma possível realização do processo estocástico  $S_X(X = x_0; n)$ , denominada função amostra ou série temporal, representada pelo símbolo  $x(n)$  (MORETTIN; TOLOI, 2006).

### 2.3.1.1 Estacionariedade e Ergodicidade

Diz-se que o processo estocástico  $S_X(X = x, n)$  é estacionário em sentido estrito, se sua distribuição de probabilidades é invariante ao deslocamento de tempo (KAY, 1993). Esta condição é expressa matematicamente pela Equação 2.11

$$F(x_1, \dots, x_N; n_1, \dots, n_N) = F(x_1, \dots, x_N; n_1 + \tau, \dots, n_N + \tau) \quad (2.11)$$

Entretanto, a condição de estacionariedade em sentido estrito é demasiadamente forte e são poucos os fenômenos que obedecem esta restrição (KAY, 1993). Com isso, foi proposta uma condição mais flexível, denominada estacionariedade no sentido amplo, expressa mediante três condições:

- $\mu_X(n) = \mu_X$
- $\sigma_X^2(n) = \sigma_X^2$
- $\rho_X(n_1, n_2) = \rho_X(l), \quad l = n_2 - n_1$

Já a hipótese de ergodicidade diz que as estatísticas de um processo estocástico (*e.g.* média, autocorrelação etc) podem ser determinadas a partir de uma única realização do processo, *i.e.* a partir da observação de uma série temporal (KAY, 1993).

Quando o processo apresenta as propriedades de estacionariedade no sentido amplo, bem como ergodicidade, pode-se substituir funções estatísticas por suas versões temporais.

Destaca-se a função de autocorreção do processo  $x(n)$ , que, para processos de tempo discreto, passa a ser definida pela Equação 2.12, em que  $l$  é o atraso desejado.

$$\rho_X(l) = \sum_{n=0}^{N-1} x(n)x(n-l) \quad (2.12)$$

### 2.3.2 Decomposição de Séries Temporais

Na abordagem clássica para modelagem e predição de valores, faz-se a suposição de que os dados sob observação são oriundos de alguma distribuição de probabilidades e que existe correlação entre eles (MONTGOMERY; JENNINGS, 2008). Com isso, lança-se o processo de ajuste dos dados a determinada função determinística acrescida de uma parcela aleatória, representando o erro do modelo ou resíduo (MONTGOMERY; JENNINGS, 2008).

Segundo (MORETTIN; TOLOI, 2006), uma série temporal pode ser modelada considerando um modelo aditivo de três elementos, conforme a Equação 2.13.

$$x(n) = T(n) + C(n) + e(n) \quad (2.13)$$

Em que  $T(n)$  é a tendência da série, representando sua direção;  $C(n)$  é a componente cíclica e  $e(n)$  é uma parcela aleatória decorrente do erro do modelo.

Classicamente, modela-se a série temporal considerando cada parcela individualmente (MORETTIN; TOLOI, 2006). Existem métodos destinados a modelar cada parcela, cada qual apresenta pontos positivos bem como negativos, tais como: precisão, complexidade etc. Entre os modelos utilizados, cita-se (MONTGOMERY; JENNINGS, 2008): (i) alisamento exponencial, (ii) média móvel, (iii) métodos em frequência e (iv) ARIMA. Entre os quais, o método ARIMA é analisado neste trabalho devido a sua aplicabilidade no problema de predição de séries temporais financeiras (MORETTIN; TOLOI, 2006);

### 2.3.3 Modelo ARIMA para Séries Temporais

O modelo ARIMA é composto por três termos a saber *Autoregressive* (AR), *Moving Average* (MA) e *Integrated* (I) que dão nome ao modelo. Cada componente de um modelo ARIMA tem como responsabilidade modelar parcialmente características presentes na série temporal em análise (KHASHEI; BIJARI, 2010).

Na apresentação a seguir,  $n = 1, \dots, N_{obs}$  representa o instante de observação do conjunto de  $N_{obs}$  observações da série temporal de tempo discreto, amostrado a partir de  $t = n \cdot T_a$ , em que  $T_a$  é a taxa de amostragem,  $x(n)$  é o valor da série temporal,  $e(n)$  é o erro de predição do modelo, caracterizado pelo ruído com distribuição de probabilidades gaussiana padrão  $\mathcal{N}(0, 1)$  com função de densidade de probabilidade  $f_X(X = x)$  exibida na Equação 2.14, com média  $\mu_X = 0$  e variância  $\sigma_X^2 = 1$ . O resíduo  $e(n)$  é exibido na Figura 2.2, em que o eixo das abscissas representa o instante de observação e o eixo das ordenadas compreende a amplitude de  $e(n)$ .

$$f_X(X = x) = \frac{1}{\sigma_X \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x - \mu_X}{\sigma_X} \right)^2} \quad (2.14)$$

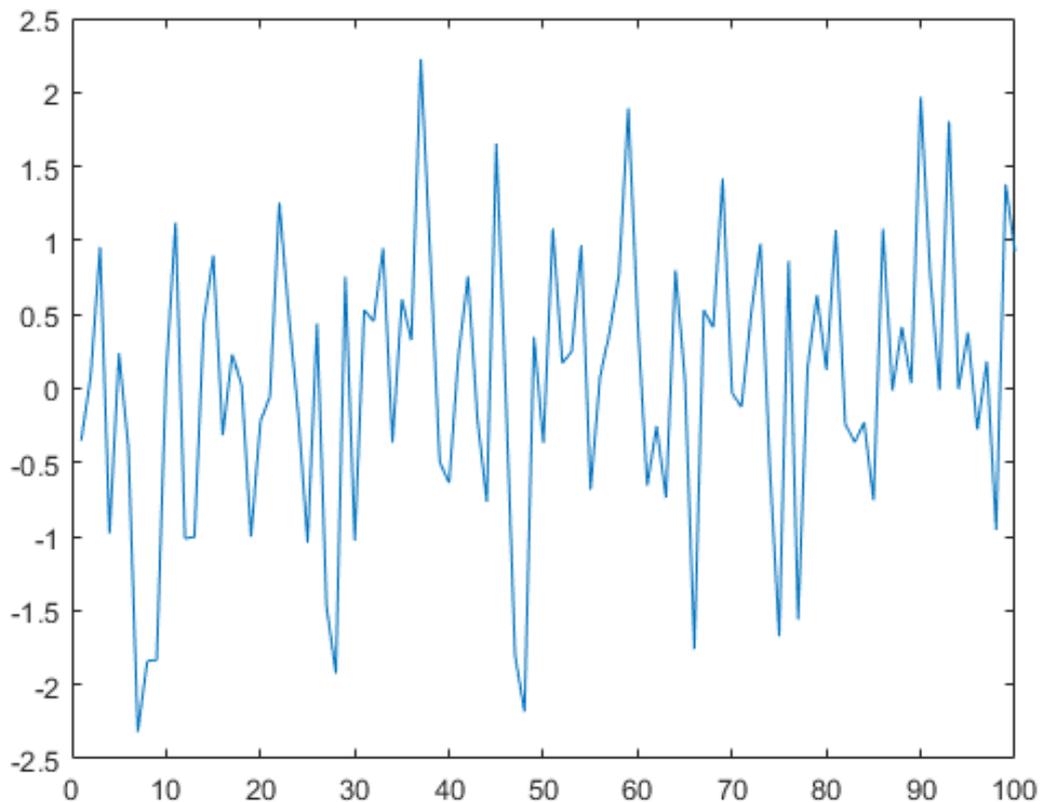


Figura 2.2: Comportamento do Erro de Predição  $e(n)$ .

### 2.3.3.1 Modelo AR

A parcela  $AR(p)$  de ordem  $p$ , tem como objetivo estimar o valor futuro da série temporal  $x(n)$  mediante a combinação linear de seus termos passados. A estrutura da componente AR é exibida na Equação 2.15, sendo  $\phi_i$ ,  $i = 1, \dots, p$ , os coeficientes de auto-regressão.

$$x(n) = - \sum_{i=1}^p \phi_i \cdot x(n-i) + e(n) \quad (2.15)$$

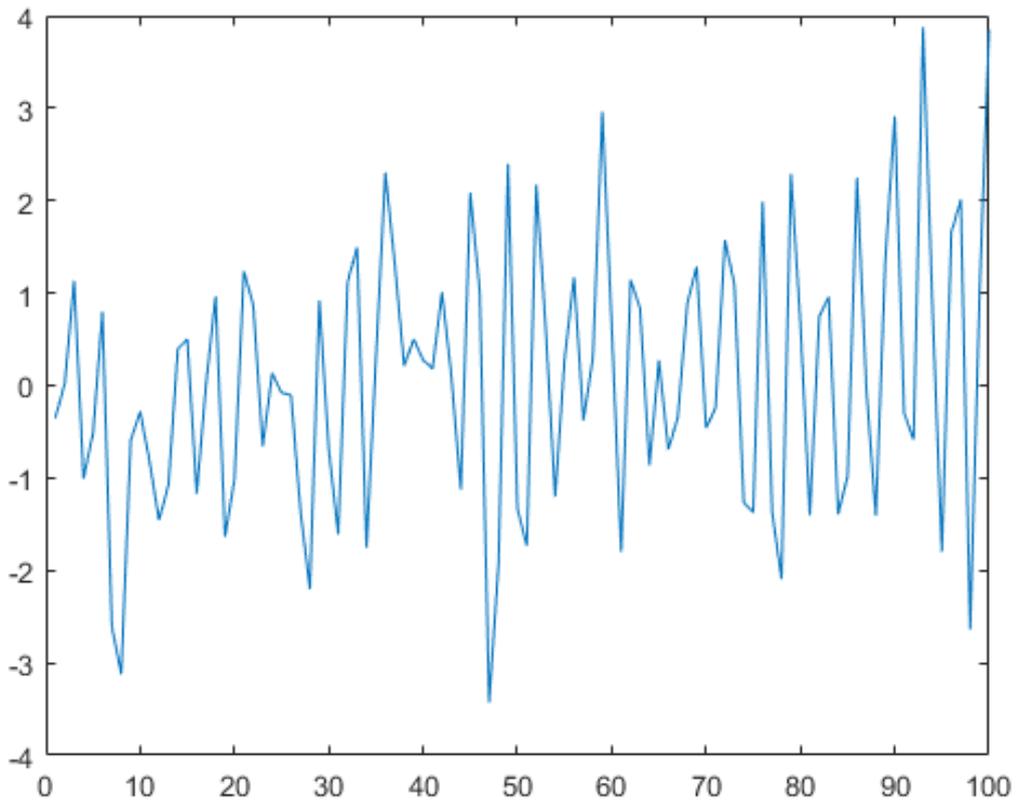


Figura 2.3: Série Temporal Obtida pelo Modelo AR(3).

A Figura 2.3 exibe um exemplo de comportamento para o modelo AR de terceira ordem AR(3) baseado na Equação 2.15, em que o eixo das abscissas representa o instante de observação e o eixo das ordenadas compreende a amplitude do sinal obtido mediante o modelo AR(3).

### 2.3.3.2 Modelo MA

Enquanto a componente  $AR(p)$  estima o valor da série temporal  $x(n)$  com base em seus valores passados, a componente  $MA(q)$  de ordem  $q$ , estima o valor futuro da série com base na ponderação das parcelas de erro de predição defasadas. A estrutura da componente MA é exibida na Equação 2.16, em que  $\theta_i$ ,  $i = 1, \dots, q$ , são os coeficientes de média móvel.

$$x(n) = \sum_{i=1}^q \theta_i \cdot e(n-i) + e(n) \quad (2.16)$$

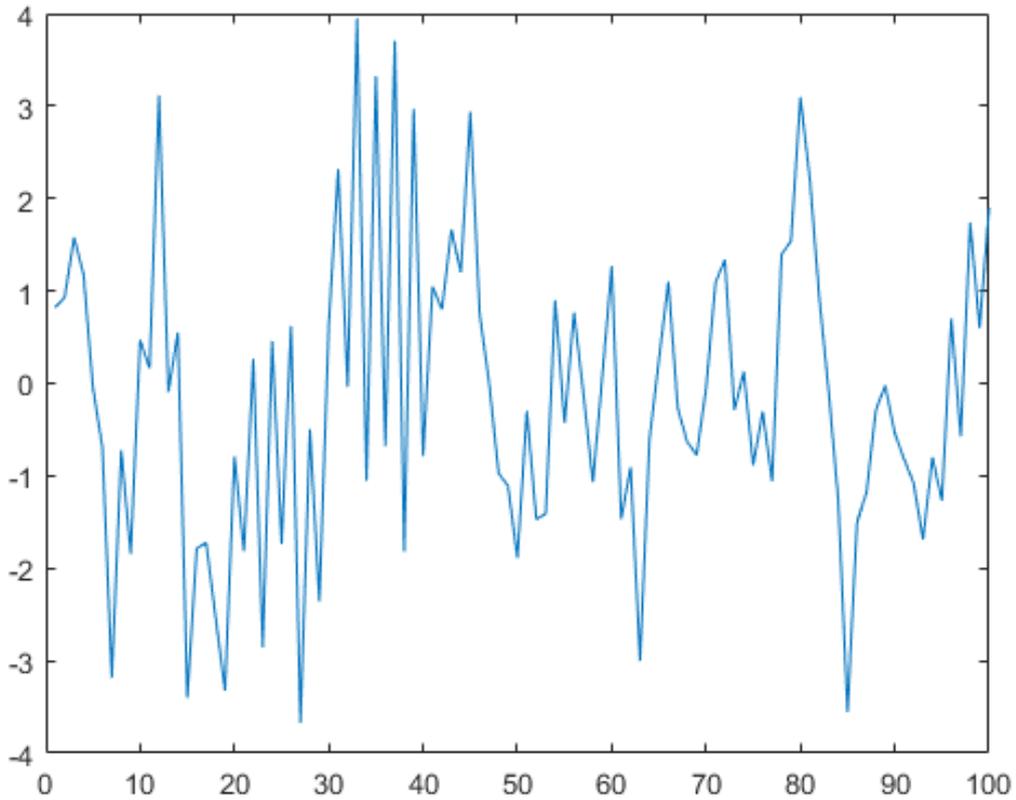


Figura 2.4: Série Temporal Obtida pelo Modelo MA(3).

A Figura 2.4 exibe um exemplo de comportamento para o modelo MA de terceira ordem MA(3) baseado na Equação 2.16, em que o eixo das abscissas representa o instante de observação e o eixo das ordenadas compreende a amplitude do sinal obtido mediante o modelo MA(3).

### 2.3.3.3 Modelo ARMA

É possível combinar os modelos AR( $p$ ) e MA( $q$ ) e assim obter o modelo ARMA( $p, q$ ) (ZHANG, 2003), em que tanto valores passados da série temporal  $x(n)$ , quanto seus erros de predição  $e(n)$ , são considerados na estimação de valores futuros da série. A Equação 2.17 exhibe a estrutura geral deste modelo.

$$x(n) = - \sum_{i=1}^p \phi_i \cdot x(n-i) + \sum_{i=1}^q \theta_i \cdot e(n-i) + e(n) \quad (2.17)$$

De forma a simplificar a notação, em (MONTGOMERY; JENNINGS, 2008), introduz-se o operador de atraso  $L^i[\cdot]$ . A aplicação deste operador em uma função  $x(n)$ , resulta em  $L^i[x(n)] = x(n-i)$ . Com isso, a Equação 2.17 é reescrita pela Equação 2.18.

$$\begin{aligned} x(n) &= - \sum_{i=1}^p \phi_i L^i x(n) + \sum_{i=1}^q \theta_i L^i e(n) + e(n) \\ \Leftrightarrow (1 + \sum_{i=1}^p \phi_i L^i) x(n) &= (1 + \sum_{i=1}^q \theta_i L^i) e(n) \\ \Leftrightarrow \phi(L) x(n) &= \theta(L) e(n) \end{aligned} \quad (2.18)$$

A Figura 2.5 exhibe um exemplo de comportamento para o modelo ARMA(3, 3) baseado na Equação 2.18, em que o eixo das abscissas representa o instante de observação e o eixo das ordenadas compreende a amplitude do sinal obtido mediante o modelo ARMA(3, 3).

### 2.3.3.4 Modelo I

A componente  $I(d)$  com ordem de diferenciação  $d$ , é necessária para remoção de tendências polinomiais eventualmente presentes na série temporal e, desta forma, promover sua estabilidade. Esta classe de tendências é retirada da série mediante aplicação de diferenciações sucessivas, definida na Equação 2.19 para o caso de primeira ordem. Tem-se como exemplo, o caso da série temporal  $x(n)$  composta pela parcela de média  $\mu_X$  acrescida do resíduo  $x(n)$ ,  $x(n) = \mu_X + e(n)$ , em que, aplicando a diferenciação de primeira ordem, *i.e.* com  $i = 1$ , tem-se  $(1 - L)x(n) = [\mu_X - x(n)] - [\mu_X - e(n-1)] = e(n-1) - e(n)$ .

$$(1 - L)^d x(n) \Rightarrow x(n) - x(n-d) \quad (2.19)$$

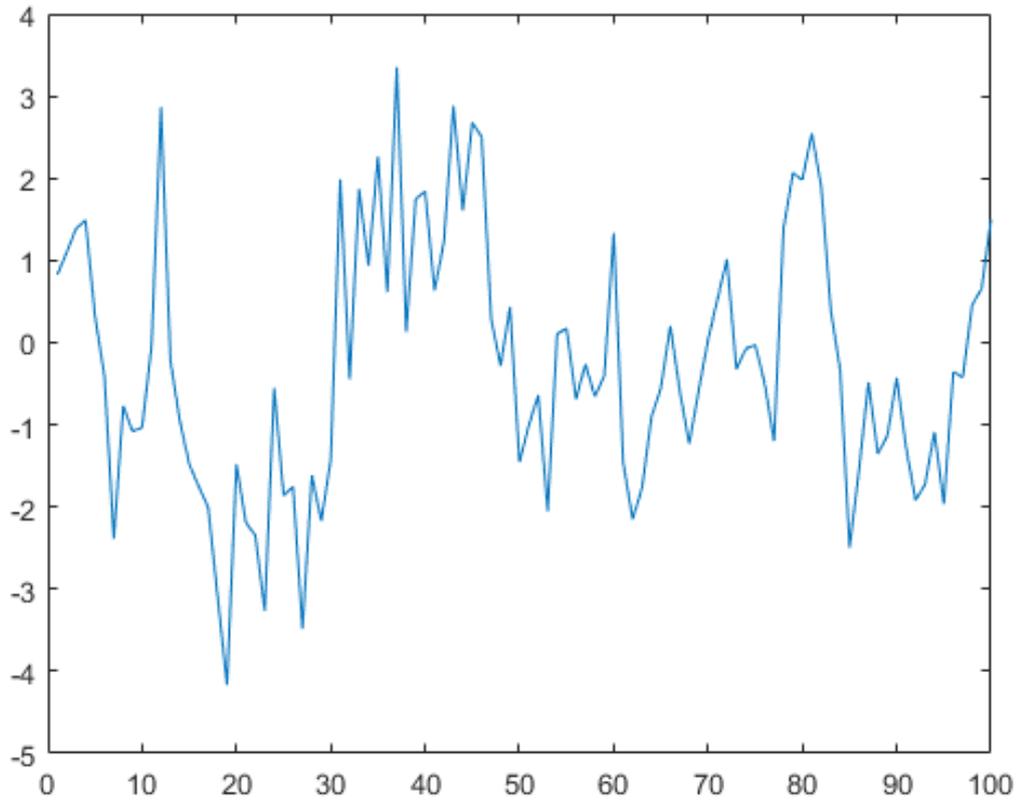


Figura 2.5: Série Temporal Obtida pelo Modelo ARMA(3, 3).

A Figura 2.6 exibe o exemplo de comportamento obtido pelo modelo I(1). Neste exemplo, é feita a remoção da média  $\mu_X = 10$  presente na série  $x(n) = 10 + e(n)$ , resultando na nova série temporal  $x_d(n) = e(n - 1) - e(n)$ . Na Figura 2.6, o eixo das abscissas representa o instante de observação e o eixo das ordenadas compreende a amplitude do sinal obtido mediante o modelo I(1).

Na Figura 2.6, são apresentadas duas curvas: uma azul com linha contínua que representa a série temporal original  $x(n)$ , com média  $\mu_x = 10$  e uma vermelha com linha pontilhada, que representa a série  $x_d(n)$ , com média  $\mu_{x_d} = 0$ , correspondente a série  $x(n)$  após ser processado pelo modelo I(1).

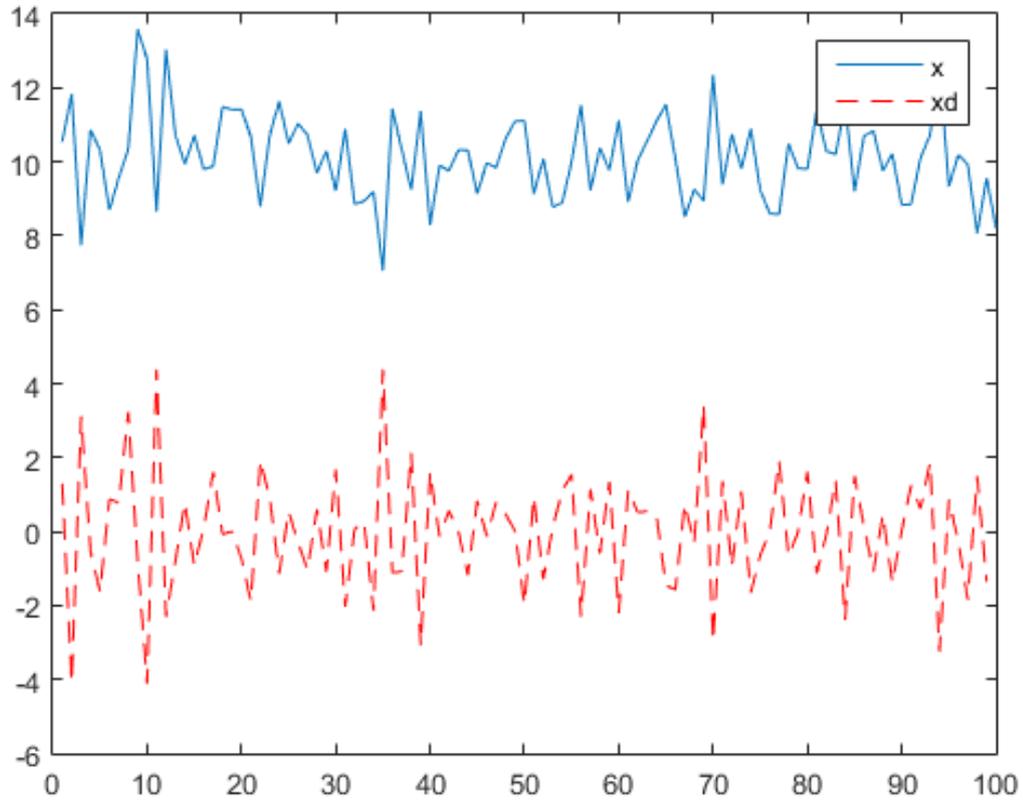


Figura 2.6: Séries Temporais Obtidas pelo Modelo I(1).

### 2.3.3.5 Modelo ARIMA

Por fim, é possível combinar os componentes  $ARMA(p, q)$  e  $I(d)$  e desta forma conceber o modelo  $ARIMA(p, d, q)$ . Neste modelo, os valores passados da série temporal são integrados e combinados com os termos de erro. A Equação 2.20 exhibe a estrutura deste modelo.

$$\phi(L)(1 - L)^d x(n) = \theta(L)e(n) \quad (2.20)$$

A Figura 2.7 exhibe o exemplo de comportamento da série temporal  $x(n)$  obtida pelo modelo  $ARIMA(3, 1, 3)$ , em que o eixo das abscissas representa o instante de observação e o eixo das ordenadas compreende a amplitude da série obtido mediante o modelo  $ARIMA(3, 1, 3)$ .

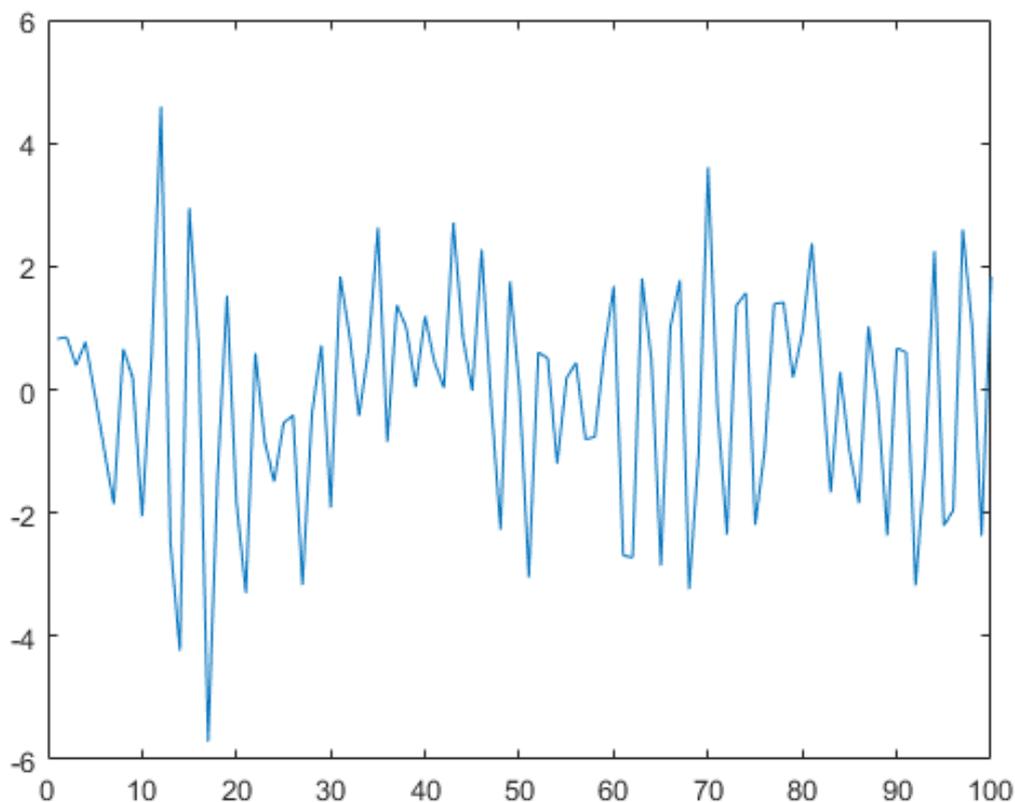


Figura 2.7: Série Temporal Obtida pelo Modelo ARIMA(3, 1, 3).

### 2.3.3.6 Metodologia Box-Jenkins

O projeto de modelos para séries temporais que utiliza a abordagem  $ARIMA(p, d, q)$ , pode ser realizado mediante o emprego da metodologia Box-Jenkins (BOX et al., 2015), composta de três etapas sucessivas:

1. Identificação;
2. Estimação;
3. Verificação.

A identificação consiste em definir a ordem do modelo ARIMA, *i.e.* determinação dos números  $(p, d, q)$ , referentes às parcela AR, MA e I, respectivamente. Nesta etapa, lança-se mão de ferramentas estatísticas, tais como gráficos das funções de autocorrelação

e autocorrelação parcial. Esta tarefa demanda testes (MONTGOMERY; JENNINGS, 2008), em que seleciona-se alguns valores candidatos a ordem do modelo e verifica-se a adequação destes aos dados mediante o processo de busca exaustiva. Este processo é repetido até alcançar o grau de adequação exigido pela aplicação (BOX et al., 2015).

Na estimação, objetiva-se determinar os coeficientes  $\{\phi_i\}$ ,  $\{\theta_i\}$  das parcelas AR e MA respectivamente. O processo de estimação pode ser realizado mediante aplicação de métodos de otimização, tais como Maximum-Likelihood (ML) (BOX et al., 2015), sendo recomendada a utilização de ferramentas computacionais que ofereçam suporte a este método, por exemplo o Matlab.

Por fim, na etapa de verificação, tem-se como objetivo avaliar a proximidade entre os dados de teste com aqueles estimados pelo modelo ARIMA, sendo possível a utilização de métodos objetivos de avaliação, *e.g.* Minimum Mean Square Error (MMSE) e/ou testes estatísticos, *e.g.* teste de Ljung–Box (BOX et al., 2015).

Estas três etapas são realizadas iterativamente até obter-se o resultado desejado para a aplicação em questão.

## 2.4 Inteligência Artificial Conexionista

Segundo visto em (BRAGA; LUDERMIR, 2011), pode-se modelar o comportamento do cérebro humano considerando-o como sendo composto pela interconexão de unidades elementares de processamento denominadas neurônios (LUGER, 2004).

Cada neurônio recebe um conjunto de estímulos em sua entrada, efetua o processamento e produz determinado conjunto de resultados em sua saída. Sendo que, esta saída é enviada às entradas de outros neurônios mediante conexões denominadas sinapses (HAYKIN, 2008). A Figura 2.8 descreve a estrutura de um neurônio fisiológico.

O processamento ocorre mediante o seguinte fluxo (HAYKIN, 2008):

1. Recebimento de impulsos em seus terminais de entrada, denominados dendritos;
2. Processamento dos impulsos no corpo celular;
3. Fornecimento do resultado no terminal de saída, denominado axônio.

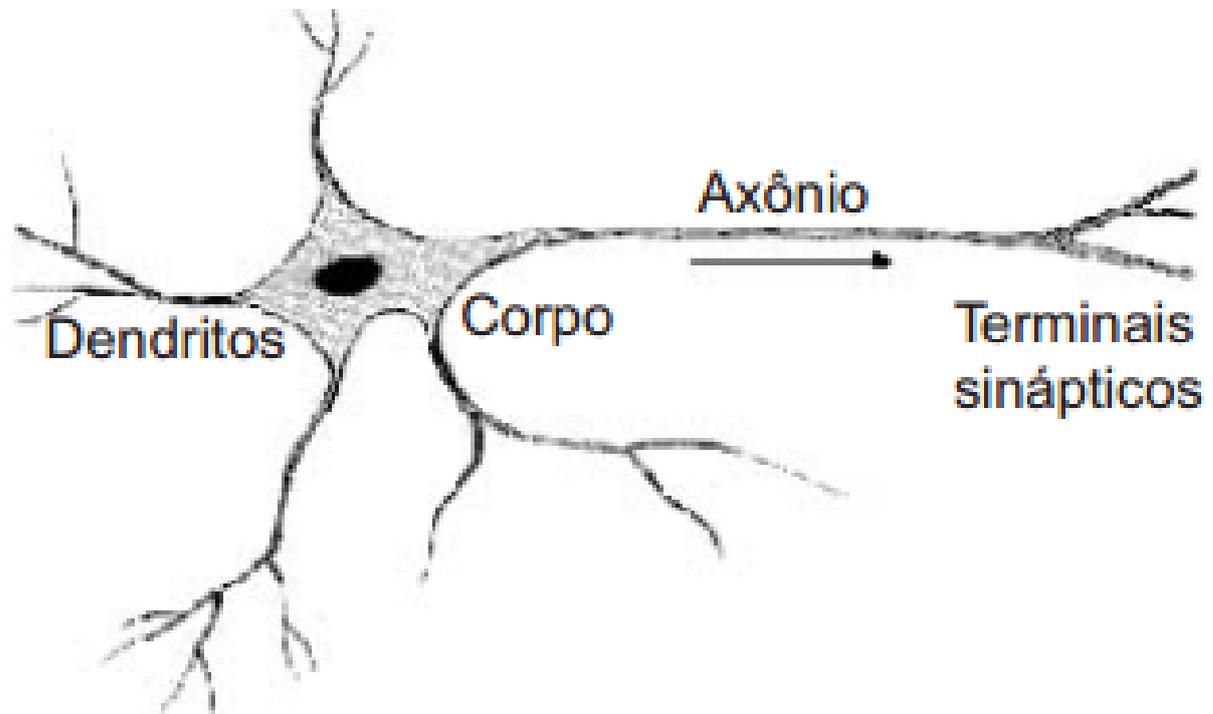


Figura 2.8: Representação do Neurônio Fisiológico. Fonte: (HAYKIN, 2008).

Neste contexto, a ANN consiste em um modelo adaptativo não-linear de processamento implementado em *software* e/ou *hardware*, sendo inspirado no modelo de processamento utilizado no cérebro humano (BRAGA; LUDERMIR, 2011).

Entre as características de uma ANN, pode-se destacar:

- Extração de padrões a partir de um conjunto de dados de treinamento;
- Generalização de resultados, *i.e.*, capacidade de aprender novos padrões, fornecendo resultados satisfatórios para entradas não apresentadas durante a etapa de treinamento.

Um exemplo de arquitetura para uma ANN em que, na sua entrada aplica-se a série temporal  $x(n)$  e na sua saída observa-se a série  $y(n)$ , composta por 1 camada de entrada com 2 neurônios, 2 camadas ocultas de 3 neurônios cada e 1 camada de saída com 1 neurônio é exibido na Figura 2.9.

Segundo (KHASHEI; BIJARI, 2010), as atribuições de cada camada são:

**Camada de entrada** - Obtenção de dados a partir do ambiente externo e, eventualmente, aplicação de pré-processamento para adequá-los ao propósito do modelo;

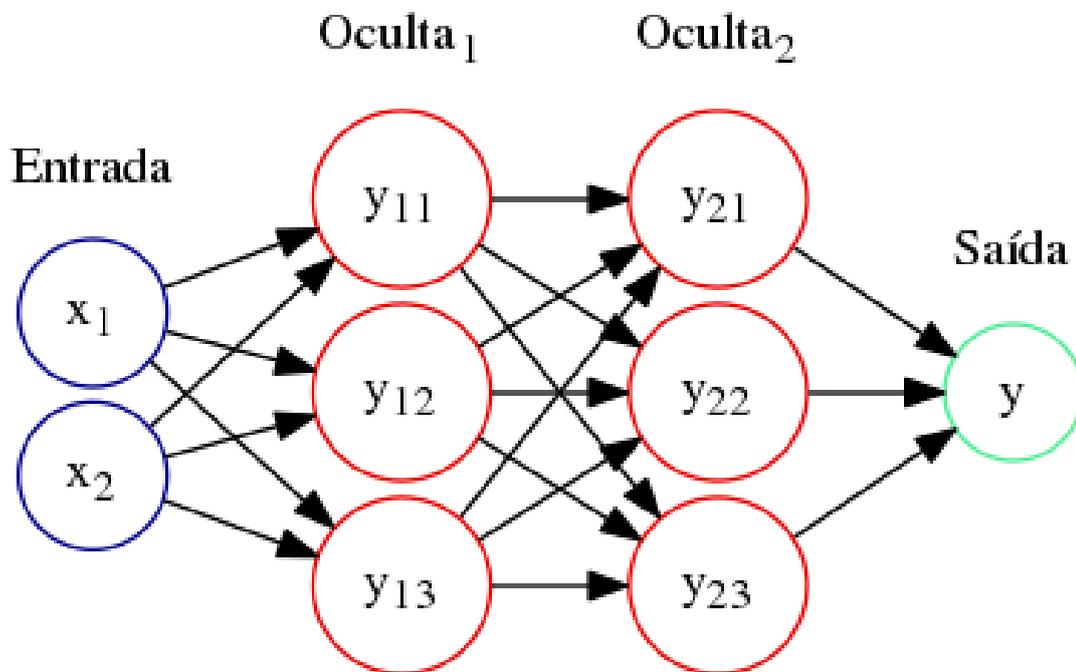


Figura 2.9: Exemplo de Arquitetura de ANN em Camadas.

**Camada oculta** - Realização do processamento a partir dos dados de entrada, determinação de correlações e, posteriormente, aplicação de entradas ainda não observadas;

**Camada de saída** - Disponibilização dos resultados obtidos a partir do processamento realizado nas camadas ocultas.

### 2.4.1 Neurônio Artificial

Inspirado no modelo do neurônio fisiológico, em 1943, os pesquisadores Warren McCulloch e Walter Pitts propuseram o primeiro modelo matemático do neurônio artificial, denominado MCP (MCCULLOCH; PITTS, 1943). O neurônio MCP é composto por  $j = 1, \dots, N_{term}$  terminais de entradas binárias  $x_j(n)$ , caracterizadas pelo conjunto de pesos sinápticos  $\{w_j\}$  que ponderam cada entrada individualmente, resultando no potencial de ativação  $v(n)$ , uma função de ativação do tipo degrau unitário de Heaviside  $g(\cdot)$  com limiar  $\lambda$  e um terminal de saída binária  $y(n)$  (HAYKIN, 2008).

A relação entre estes componentes neuronais é expressa pela Equação 2.21, em que a função  $u(\cdot)$  é exibida na Equação 2.22.

$$v(n) = \sum_{j=1}^{N_{term}} x_j(n) \cdot w_j \quad (2.21)$$

$$y(n) = g(v) = \begin{cases} 0, & v(n) < \lambda \\ 1, & v(n) \geq \lambda \end{cases} \quad (2.22)$$

Com a estrutura do neurônio MCP, é possível aproximar as funções lógicas: OR, AND e NOT (HAYKIN, 2008). Entretanto, conforme dito em (HAYKIN, 2008), o neurônio MCP apresenta deficiências que impossibilitaram sua aplicabilidade em problemas mais complexos, a saber (BRAGA; LUDERMIR, 2011):

- Aplicável somente às entradas e saídas binárias;
- Limitação quanto à função de ativação;
- Não dispunha de metodologia para otimização analítica dos pesos sinápticos.

Com base nas limitações elicitadas, em 1958, Frank Rosenblatt propôs seu modelo do neurônio artificial denominado Perceptron (ROSENBLATT, 1958). O neurônio perceptron corresponde a uma extensão do modelo do neurônio MCP, que soluciona suas limitações (HAYKIN, 2008), incluindo sua aplicação a problemas envolvendo variáveis contínuas, em contraposição ao modelo MCP que admitia apenas variáveis binárias (KOVACS, 2006).

Semelhante ao modelo do neurônio MCP, o potencial de ativação do perceptron é definido pela Equação 2.23, em que a diferença em relação ao modelo MCP definido pela Equação 2.21, é a adição da parcela denominada *bias*  $\lambda$ , responsável por definir uma transformação afim da entrada para a saída, deslocando a região de decisão a partir da origem (KHASHEI; BIJARI, 2010).

$$v(n) = \sum_{j=1}^{N_{term}} x_j(n) \cdot w_j + \lambda \quad (2.23)$$

Já com relação a saída do perceptron, tem-se à disposição outras funções de ativação, além da função limiar utilizada no modelo MCP, responsáveis pela compressão dos valores do potencial de ativação, possivelmente no intervalo  $(-\infty, \infty)$  para uma faixa estreita de

valores (COPPIN, 2010). A única restrição, que deve-se ao algoritmo de aprendizagem do perceptron, a ser descrito posteriormente, é que a função de ativação seja suave, *i.e.*, admita derivada. Entre as funções de ativação comumente adotadas, destaca-se as família de funções *sigmoid* (HAYKIN, 2008), entre as quais cita-se as funções logística (*logsig*) e tangente hiperbólica (*tansig*), definidas nas equações 2.24 e 2.25, respectivamente.

$$g(v) = \text{logsig}(v) = \frac{1}{1 + e^{-v}}, \quad g : (-\infty, \infty) \rightarrow [0, 1] \quad (2.24)$$

$$g(v) = \text{tansig}(v) = \frac{1 - e^{-v}}{1 + e^{-v}}, \quad g : (-\infty, \infty) \rightarrow [-1, 1] \quad (2.25)$$

O conhecimento adquirido pelo neurônio artificial é codificado em seus pesos sinápticos. Portanto, é necessário que, inicialmente, estes valores sejam estimados para solucionar o problema em questão, a este processo de ajuste dos pesos sinápticos, dá-se o nome de treinamento (HAYKIN, 2008). Entre os avanços obtidos com o modelo perceptron, destaca-se a formulação analítica de sua regra de aprendizagem, denominada Lei de Aprendizagem do Perceptron (COPPIN, 2010). Este algoritmo incrementa iterativamente o peso sináptico da entrada  $j$  no instante  $i$ ,  $w_j(i)$ , com a parcela  $\Delta w_j(i) = w_j(i) - w_j(i - 1)$ , a partir da regra descrita na Equação 2.26 (KOVACS, 2006).

$$\Delta w_j(i) = \eta [1 - \hat{y}_j(i) y_j(i)] y_j(i) x_j(i) \quad (2.26)$$

Em que  $\hat{y}_j(i)$  é a saída obtida pelo perceptron,  $y_j(i)$  é a saída desejada,  $x_j(i)$  é a entrada  $j$  e  $\eta$  é uma constante denominada taxa de aprendizagem, utilizada para acelerar o processo de treinamento.

Apesar de suas limitações teóricas e práticas, o modelo do neurônio perceptron, forneceu a base do algoritmo de aprendizagem *backpropagation* (LUGER, 2004), comumente empregado no treinamento de diversas arquiteturas de ANN, inclusive aquelas dispostas em múltiplas camadas a serem descritas nas seções seguintes.

#### 2.4.1.1 Algoritmo de Treinamento

O correto funcionamento de uma ANN é condicionado pelo processo de treinamento, de modo que os parâmetros livres do modelo, a saber os pesos sinápticos  $\{w_{jk}\}$ , *i.e.*, conexão entre os neurônios  $j$  e  $k$ , sejam ajustados aos dados sob análise (COPPIN, 2010). Conforme

visto em (HAYKIN, 2008), existem três paradigmas de aprendizagem, ambos inspirados na Teoria da Aprendizagem oriunda do campo da Psicologia:

- Aprendizagem supervisionada;
- Aprendizagem por reforço;
- Aprendizagem não-supervisionada.

No escopo deste trabalho de pesquisa, destaca-se a aprendizagem supervisionada, composta por uma malha de controle fechada, em que, iterativamente, efetua-se a comparação entre a resposta obtida pela ANN com a resposta desejada, informada pelo componente denominado professor (HAYKIN, 2008).

O emprego da aprendizagem supervisionada necessita do professor, que armazena o conhecimento do problema a ser transmitido à ANN (LUGER, 2004). Para cada sinal de estímulo  $x(i)$ ,  $i = 1, \dots, N_{est}$ , a ANN fornece a saída  $\hat{y}(i)$ , enquanto o professor informa a saída desejada  $y(i)$  (KOVACS, 2006). A partir de cada par ordenada  $[x(i), y(i)]$  apresentado à ANN, calcula-se o sinal de erro,  $e(i) = \|\hat{y}(i) - y(i)\|_N$ , em que  $\|\cdot\|_N$  representa alguma métrica de distância entre vetores no espaço vetorial  $N$ -dimensional  $\mathbb{R}^N$ , por exemplo a distância Euclidiana definida na Equação 2.27, em que  $\vec{A} = (a_1, \dots, a_N)$  e  $\vec{B} = (b_1, \dots, b_N)$ ,  $m = 1, \dots, N$  (KOVACS, 2006). Para cada iteração, procura-se minimizar o sinal de erro, aproximando a resposta obtida pela ANN com a resposta informada pelo professor.

$$\|\vec{A} - \vec{B}\|_N = \sqrt{\sum_{m=1}^N (a_m - b_m)^2} \quad (2.27)$$

A implementação do aprendizado supervisionado é comumente realizada mediante o algoritmo *backpropagation* (HAYKIN, 2008), que tem por objetivo encontrar o ponto de mínimo na superfície de erro em função do valor dos pesos sinápticos. Para orientar a localização do ponto desejado, utiliza-se o gradiente da função de erro, desta forma, o algoritmo também é conhecido como método do gradiente descendente (COPPIN, 2010). O algoritmo *backpropagation* representa uma generalização da Lei de Aprendizagem do Perceptron e é implementado mediante as seguintes etapas sucessivas:

**Alimentação Direta** - Apresenta-se os dados na camada de entrada, para cada camada oculta, aplica-se a função de transferência que rege o comportamento do neurônio

artificial e assim sucessivamente até que a informação seja disponibilizada à camada saída;

**Realimentação** - Ao obter o valor de saída da ANN, compara-se a mesma com o valor desejado, calculando-se o sinal de erro que é fornecido para cada camada anterior retroativamente. Para cada camada, calcula-se a contribuição do peso sináptico associado a cada neurônio na composição do sinal de erro, ajustando-o a partir desta participação.

Seja  $i = 1, \dots, N_{obs}$  o índice de uma determinada saída desejada,  $y(i)$  e sua correspondente saída obtida pela ANN  $\hat{y}(i)$ , o erro total de saída cometido pelo modelo ANN para cada dupla  $i$ ,  $[y(i), \hat{y}(i)]$  é  $E(i)$ , sendo definido pela Equação 2.28.

$$E(i) = \frac{1}{2}[y(i) - \hat{y}(i)]^2 \quad (2.28)$$

Este erro de saída, conforme definido pela Equação 2.28, é retropropagado da saída no sentido da entrada e cada neurônio entre estes dois pontos tem seus pesos sinápticos ajustados de acordo com sua contribuição na composição do erro total.

Esta regra de ajuste dos pesos é denominada Regra Delta Generalizada (KOVACS, 2006), sendo definida pela Equação 2.29.

$$\Delta w_{jk}(i) = -\eta \frac{\partial E(i)}{\partial w_{jk}(i)} \quad (2.29)$$

Em que:

$\Delta w_{jk}(i)$  - é o incremento do peso sináptico entre determinado neurônio  $j$  com outro neurônio  $k$  de uma camada subsequente no instante  $i$ ;

$\eta$  - é a taxa de aprendizagem, sendo responsável por determinar a celeridade do processo de treinamento.

### 2.4.2 Arquitetura *Multilayer Perceptron*

Em uma ANN do tipo *Multilayer Perceptron* (MLP), efetua-se a interconexão de neurônios do tipo perceptron, capazes de segmentar o espaço de estados em duas regiões linearmente separáveis (HAYKIN, 2008).

A MLP é caracterizada pela agregação em camadas de neurônios do tipo perceptron, cada qual, para o caso de entrada vetorial e saída escalar, apresenta relação entre cada componente da série temporal de entrada  $\vec{x}(n)$  e saída  $\hat{y}(n)$  definida pela Equação 2.30, onde o vetor de pesos sinápticos  $\vec{w}$  é responsável por codificar o aprendizado do neurônio e a parcela denominada *bias*  $\lambda$ , é necessária para deslocar a reta de separação do espaço de estados a partir da origem (HAYKIN, 2008). A seguir, efetua-se o produto interno entre a série temporal de entrada  $\vec{x}(n)$  e  $\vec{w}$ , gerando o resultado intermediário  $v(n)$  que, por sua vez, é aplicado à função de ativação  $g(\cdot)$ , obtendo-se assim na saída, a série temporal  $\hat{y}(n)$  que é propagada aos neurônios das camadas à frente e assim sucessivamente (BRAGA; LUDERMIR, 2011).

$$y(n) = g[v(n)] = g[\vec{x}(n) \cdot \vec{w} + \lambda] \quad (2.30)$$

Desta análise, pode-se notar a importância do treinamento, em que os vetores de ponderação  $\vec{w}$  e *bias*  $\lambda$  são ajustados a partir do algoritmo de aprendizagem selecionado (KHASHEI; BIJARI, 2010). Durante o treinamento, a MLP identifica as correlações presentes nos dados, possibilitando uma generalização do modelo para entradas ainda não observadas. Devido sua simplicidade e resultados satisfatórios obtidos, a arquitetura MLP representa uma das mais utilizadas no projeto de sistemas baseados em ANN (KOVACS, 2006).

### 2.4.3 Arquitetura *Radial Basis Function*

A arquitetura *Radial Basis Function* (RBF) tem sua principal finalidade na aproximação de funções (HAYKIN, 2008). Isto é realizado mediante a inclusão de funções de ativação do tipo base radial que representam sintonizadores em torno de determinada região de interesse, denominada protótipo (KOVACS, 2006).

Na RBF, para cada entrada, calcula-se a distância Euclidiana entre a entrada e o centro da RBF. Para cada distância obtida, aplica-se a função de ativação, responsável por determinar a similaridade entre a entrada e o protótipo, quanto mais semelhantes, maior será o valor da função de ativação, enquanto que, conforme diminui-se a similaridade, tem-se que a função de ativação tende assintoticamente a zero. Por fim, na camada de saída soma-se todos os valores obtidos ponderados pelos respectivos pesos sinápticos (HAYKIN, 2008).

Entre as funções de ativação utilizadas na RBF, pode-se destacar a gaussiana modificada, definida pela Equação 2.31, em que para cada instante  $n$ , o parâmetro  $\mu_x$  representa o centro da função de ativação e  $\beta_x$  determina seu espalhamento em torno do centro.

$$y = e^{\beta_x(x-\mu_x)^2} \quad (2.31)$$

#### 2.4.4 Arquitetura *Generalized Regression Neural Network*

A ANN do tipo *Generalized Regression Neural Network* (GRNN) (SPECHT, 1991), é caracterizada por quatro camadas: (i) entrada, (ii) detecção de padrões, (iii) somatória e (iv) saída.

Na camada de entrada é feito o mapeamento entre a entrada  $x_j(n)$ ,  $j = 1, \dots, N_{term}$  com seu respectivo neurônio.

Os neurônios da camada de entrada são conectados àqueles da camada de detecção de padrões, onde cada neurônio representa um determinado padrão de treinamento e sua saída corresponde à distância entre a entrada e cada padrão de treinamento.

Já na camada de somatória é formada por dois neurônios, denominado  $S_{GRNN}$  e  $D_{GRNN}$ ,  $S_{GRNN}$  efetua a somatória ponderada das saídas da camada de detecção de padrões para cada neurônio  $j$  por um determinado peso sináptico  $w_j$ , considerando 1 camada oculta, já  $D_{GRNN}$  efetua a somatória, mas não a pondera pelos pesos sinápticos.

Por fim, a camada de saída efetua a divisão entre as somas  $S_{GRNN}$  e  $D_{GRNN}$ , obtidas na camada de somatória.

Pode-se adotar a função gaussiana como responsável por mensurar a similaridade entre a entrada e cada padrão de treinamento, neste caso, a saída  $y$  para cada instante  $n$  da GRNN é definida pela Equação 2.32, em que  $\zeta_x$  determina o grau de espalhamento da função gaussiana.

$$y = \frac{\sum_{j=1}^{N_{term}} w_j \cdot e^{\zeta_x(x-x_j)^2}}{\sum_{j=1}^{N_{term}} e^{\zeta_x(x-x_j)^2}} \quad (2.32)$$

#### 2.4.5 Arquitetura *Focused Time Lagged Feedforward*

As arquiteturas vistas até o presente momento não incorporam a dimensão temporal ao modelo, sendo assim, a mesma entrada aplicada em instantes distintos de tempo, produz

o mesmo resultado (HAYKIN, 2008). Esta propriedade da resposta ser independente do instante de tempo, possibilita classificar estas arquiteturas como sendo estáticas (KOVACS, 2006).

Entretanto, algumas aplicações possuem dados dispostos de forma sequencial (*e.g.* séries temporais) e com isso, demandam a incorporação desta dependência temporal na arquitetura da ANN, neste caso classifica-se a ANN como sendo dinâmica (HAYKIN, 2008).

Pode-se acoplar o efeito temporal na ANN mediante a utilização de uma memória de curto prazo em série com uma ANN estática (*e.g.* MLP, RBF, GRNN etc).

O modelo elementar para memórias de curto prazo consiste na utilização do filtro *Linear Time Invariant* (LTI) do tipo *Tapped Delay Line* (TDL). A estrutura TDL é composta por uma série de elementos de atrasos unitários, cada qual efetuando o deslocamento do sinal de entrada  $x(n)$  em uma amostra através de sua convolução com a função  $\delta(n-1)$ .

Conforme a maneira em que acopla-se a memória à ANN, tem-se diferentes arquiteturas, dentre as quais, a mais simples consiste na *Focused Time Lagged Feedforward* (FTDNN), em que a memória e, conseqüentemente, a estrutura temporal, encontram-se somente na camada de entrada da ANN (HAYKIN, 2008).

### 2.4.6 Arquitetura *Recurrent Neural Network*

As arquiteturas vistas até o presente momento pertencem a classe da ANN do tipo *feedforward*, em que não há laços de realimentação entre camadas (HAYKIN, 2008). Complementar a esta categoria, existe a classe *Recurrent Neural Network* (RNN), em que os sinais resultantes das camadas ocultas e/ou de saída, são realimentados às camadas ocultas anteriores e/ou à camada de entrada. A existência de realimentação na RNN faz com que a mesma adquira a capacidade de armazenar o estado do sistema dinâmico sob análise (HAYKIN, 2008), esta propriedade é útil na análise e predição de séries temporais, em que os valores passados e presente da série exercem influência nos seus valores futuros (BRAGA; LUDERMIR, 2011).

Entre as arquiteturas de RNN existentes, destaca-se, no contexto da predição de séries temporais, duas arquiteturas em que a realimentação é realizada entre a camada de saída e a camada de entrada: *Nonlinear Autoregressive* (NAR) e *Nonlinear Autoregressive with External Input* (NARX).

A arquitetura NAR é utilizada quando a predição da série temporal de interesse  $x(n)$

é feita somente com base nos seus valores passados, já na arquitetura NARX, além de considerar-se os valores passados da série temporal  $x(n)$ , inclui-se a influência exercida pela série temporal exôgena  $u(n)$ .

## 2.5 Critérios de Desempenho em Predições

Ao realizar predições de séries temporais, surge a necessidade de mensurar a proximidade entre os valores reais e aqueles preditos pelo modelo (MONTGOMERY; JENNINGS, 2008). A esta diferença entre valor real e predito, dá-se o nome de erro de predição. Neste contexto, o critério de desempenho deve avaliar os diferentes modelos e determinar aquele que minimiza o erro da predição segundo alguma métrica.

O erro da predição no instante  $n = 1, \dots, N_{obs}$ , em que  $N_{obs}$  é o total de observações realizadas, entre o valor real  $y(n)$  e sua predição  $\hat{y}(n)$  é definido pela Equação 2.33.

$$e(n) = y(n) - \hat{y}(n) \quad (2.33)$$

Para quantificar o erro total da predição  $e(n)$ , comumente utiliza-se o MSE definido na Equação 2.34 (MONTGOMERY; JENNINGS, 2008).

$$MSE = \frac{1}{N_{obs}} \sum_{n=1}^{N_{obs}} e^2(n) \quad (2.34)$$

A desvantagem de utilizar as equações 2.33 e 2.34, reside no fato de que seus valores dependem da magnitude do valor assumido pela variável a ser predita. Para corrigir este aspecto, normaliza-se o erro de predição  $e(n)$  pelo valor da variável a ser predita  $y(n)$ , resultando no erro de predição normalizado  $e_{norm}(n)$ , esta relação é exibida na Equação 2.35 (MONTGOMERY; JENNINGS, 2008).

$$e_{norm}(n) = \frac{e(n)}{y(n)} \cdot 100 \quad (2.35)$$

Seu correspondente indicador de erro total é denominado *Mean Percent Error* (MPE), explicitado na Equação 2.36 (MONTGOMERY; JENNINGS, 2008).

$$MPE = \frac{1}{N_{obs}} \sum_{n=1}^{N_{obs}} e_{norm}^2(n) \quad (2.36)$$

Outro critério relacionado ao MSE, é o MSE normalizado pela variância da série temporal  $VAR[y(n)]$ ,  $MSE_{norm}$ , definido na Equação 2.37, que quantifica a capacidade do modelo em explicar os dados observados e com isso, reduzir o grau de variabilidade do erro de predição.

$$MSE_{norm} = \frac{MSE}{VAR[y(n)]} \quad (2.37)$$

Além disso, ao experimentar diferentes modelos, é apreciável comparar o desempenho de cada um, para ao final, selecionar aquele que melhor descreve o processo estocástico subjacente.

Existem diferentes critérios de seleção com origem da Teoria de Informação e cada qual propõe quantificar a informação perdida na consideração do modelo em relação à realidade (MONTGOMERY; JENNINGS, 2008). Nestes critérios, penaliza-se os modelos que contenham elevada quantidade de parâmetros, tornando estes critérios parcimoniosos. Entre os principais critérios de seleção, pode-se ressaltar as seguintes propriedades desejadas:

**Parcimônia** - Entre dois modelos de mesmo desempenho, seleciona aquele com menor quantidade de parâmetros;

**Consistência** - A probabilidade de selecionar o modelo verdadeiro entre os demais tende à unidade à medida que a amostra populacional aumenta;

**Eficiência Assintótica** - Seleciona uma sequência de modelos enquanto a quantidade de dados disponíveis para treinamento aumenta, para os quais a variância do erro de predição um passo à frente aproxima-se daquela obtida pelo modelo verdadeiro, no mínimo tão rápido quanto qualquer outro critério.

Entre os critério de seleção, pode-se destacar (MONTGOMERY; JENNINGS, 2008): (i) *Akaike Information Criterion* (AIC), (ii) *Schwarz Information Criterion* (SIC) e (iii) *Akaike Information Criterion Corrected* (AICC).

A Equação 2.38 exhibe o critério AIC, em que  $N_p$  representa o número de parâmetros do modelo, tais como coeficientes do modelo ARIMA, número de conexões entre neurônios do modelo ANN etc. Este critério é assintoticamente estável, entretanto, não é consistente.

$$AIC = \ln\left[\frac{1}{N_{obs}} \sum_{n=1}^{N_{obs}} e^2(n)\right] + \frac{2N_p}{N_{obs}} \quad (2.38)$$

Já o SIC apresenta-se consistente, porém não é assintoticamente estável, a Equação 2.39 exhibe a relação para este critério.

$$SIC = \ln\left[\frac{1}{N_{obs}} \sum_{n=1}^{N_{obs}} e^2(n)\right] + \frac{N_p \cdot \ln(N_{obs})}{N_{obs}} \quad (2.39)$$

Cada critério de desempenho apresenta determinado conjunto de propriedades e desta forma, é necessário avaliar quais propriedades são desejadas para a aplicação em questão.

Neste contexto, para avaliar os diferentes modelos propostos neste trabalho, escolheu-se três critérios de avaliação de desempenho: (i) MSE como determinante para seleção do modelo ótimo no sentido de apresentar menor erro de predição e (ii) AIC é utilizado como critério de desempate entre modelos com o mesmo MSE e (iii) MSE normalizado como métrica de desempenho em relação à variabilidade da série temporal original. Além disso, para cada modelo, exhibe-se o comportamento do erro de predição.

# Capítulo 3

## Modelos Propostos

*I just wondered how things were put together.*

—Claude E. Shannon

O mercado financeiro e, em particular o mercado de ações, caracteriza-se pela assimetria de informações acerca dos fatores que governam o preço dos ativos nele existentes, de particular importância o IBOVESPA (BELMONT, 2010).

Com base na informação disponível, cada agente posiciona-se no IBOVESPA, seja comprando e/ou vendendo seus componentes. A agregação de tais posições determina o valor final do índice e, com efeito, o desequilíbrio entre o número de agentes compradores e vendedores implica na movimentação do preço do IBOVESPA ao longo do tempo (PALLADINO, 2008).

Adotando a hipótese *ceteris paribus*, em que, além do fator sob análise, considera-se todos os demais fatores que regem o comportamento do ativo constantes, tem-se que, conforme acredita-se que o valor do IBOVESPA tende a aumentar, cada vez mais investidores tendem a comprá-lo, acelerando o processo de alta. De modo complementar, conforme tem-se a crença de que o valor do IBOVESPA tende a diminuir, investidores tendem a optar por vendê-lo, acelerando o processo de baixa. A sucessão de processos de alta e baixa, manifesta-se na composição da volatilidade do IBOVESPA (BELMONT, 2010).

Este capítulo tem como objetivo definir os modelos propostos neste trabalho, desta forma, tem-se a seguinte organização. Na Seção 3.1 faz-se a descrição formal do problema a ser avaliado. A Seção 3.2 descreve o modelo ARIMA. Já na Seção 3.3, descreve-se os

diferentes modelos baseados na ANN.

### 3.1 Descrição do Problema

O enfoque de estudo do presente trabalho reside na análise e predição da série temporal correspondente ao preço de fechamento do IBOVESPA.

Para avaliar a adequação dos modelos estudados, efetuou-se a divisão dos dados disponíveis em duas categorias: (i) treinamento e (ii) predição.

Os dados de treinamento são utilizados para o processo de aprendizagem, em que os parâmetros livres do modelo são otimizados para melhor se adequarem aos dados.

Já os dados de predição são utilizados para avaliar a acurácia do modelo para dados ainda não apresentados. Neste caso, utilizou-se como critérios de desempenho: MSE, AIC e  $MSE_{norm}$  globais, *i.e.*, considerando ambos os sub-conjuntos de dados de treinamento (adequação aos dados) e o sub-conjunto de dados de predição (capacidade de generalização).

Neste trabalho utilizou-se dados reais do IBOVESPA, tomados em base diária, obtidos a partir de (FINANCE, 2017), correspondentes ao período de 01/01/2005 até 31/12/2015, totalizando  $N_{obs} = 2669$  observações, deste total, 85% correspondem ao sub-conjunto de treinamento e 15% ao sub-conjunto de predição (BRAGA; LUDERMIR, 2011).

Inicialmente, para entender o processo de representação de séries temporais mediante os modelos ARIMA e MLP simples, identificar seus principais parâmetros e avaliar a correteza dos algoritmos implementados, aplicou-se os mesmos ao problema de prever os valores futuros da função estocástica a tempo contínuo  $P_{teste}(t)$  definida na Equação 3.1, que é amostrada com período  $T_a = 10$  ms, resultando na série temporal a tempo discreto  $P_{teste}(n)$ . Nesta equação, objetiva-se aproximar a variabilidade estatística encontrada no IBOVESPA, representada pelo efeito do ruído gaussiano com  $X \sim \mathcal{N}(0, 1)$ , além disso, esta função exibe mudança de tendência para  $t > 1$ .

$$P_{teste}(t) = 0,2 \cdot X + \begin{cases} 2 \cdot \sin(2 \cdot \pi \cdot 2 \cdot t) + \cos(2 \cdot \pi \cdot 10 \cdot t), & t \leq 1 \\ 2 \cdot \sin(2 \cdot \pi \cdot 2 \cdot t) + t^2, & t > 1 \end{cases} \quad (3.1)$$

Para cada modelo empregado, exibe-se o comportamento do valor real do IBOVESPA, sua predição, o erro de predição cometido e o valor dos critérios de desempenho adotados.

## 3.2 Modelo ARIMA

Nesta seção, descreve-se o modelo ARIMA proposto ao problema de predição que, inicialmente, foi validado com a função de teste definida pela Equação 3.1 e, em seguida, aplicado aos dados reais do IBOVESPA.

A otimização do modelo ARIMA envolve a aplicação da metodologia Box-Jenkins (BOX et al., 2015) em suas três etapas: (i) identificação, (ii) estimação e (iii) predição.

Inicialmente utiliza-se as funções de autocorrelação e autocorrelação parcial da série temporal para efetuar a identificação inicial do modelo, *i.e.* a ordem  $p$  do modelo  $AR(p)$  e a ordem  $q$  do modelo  $MA(q)$ . A partir destas estimativas iniciais  $\hat{p}$  e  $\hat{q}$ , aplica-se o método de busca exaustiva em intervalos centrados nas mesmas. Já com relação da ordem da diferenciação  $d$ , correspondente ao modelo  $I(d)$ , pode-se determiná-la mediante a avaliação da hipótese de comportamento estacionário para a série temporal e suas sucessivas diferenciações, ou pode-se empregar o método de busca exaustiva considerando que, comumente,  $d$  encontra-se no intervalo  $0 \leq d \leq 2$  (MORETTIN; TOLOI, 2006).

Já a etapa de estimação corresponde à determinação dos coeficientes dos modelos  $AR(p)$   $\phi_i$ ,  $0 \leq i \leq p$  e  $MA(q)$   $\theta_j$ ,  $0 \leq j \leq q$ . Existem diferentes algoritmos para a realização desta tarefa, neste trabalho, destaca-se o método ML disponível no pacote *Econometrics* do Matlab (MATHWORKS, 2017a).

Por fim, a partir do modelo ARIMA obtido, efetua-se predições da série temporal do IBOVESPA e verifica-se seu desempenho, caso o mesmo não seja aceitável, retorna-se à etapa de identificação, refinando o modelo em um processo iterativo.

## 3.3 Modelo ANN

Nesta seção, descreve modelos propostos baseados na ANN ao problema de predição do IBOVESPA.

Para atingir a otimização do modelo de predição, empregando o método de busca exaustiva, efetua-se o ajuste do parâmetros livres do modelo, são eles:

- Comprimento da janela de predição;
- Número de camadas ocultas;

- Número de neurônios em cada camada oculta;
- Desempenho desejado para o algoritmo de treinamento.

Com relação a janela temporal  $N_W$ , a mesma corresponde ao número de amostras que compõem a janela deslizante utilizada na entrada da ANN para prever o valor da série temporal uma amostra à frente, correspondente à saída da ANN. Esta janela percorre toda a série temporal, em que, cada valor predito é incorporado à série temporal de predição e, conseqüentemente, à janela temporal. Em que, para dados de treinamento, realizou-se a predição um passo à frente, enquanto que, para dados de predição, efetuou-se a predição considerando somente valores preditos pela ANN na composição sucessiva da janela temporal deslizante.

De maneira geral, é possível sintetizar o projeto de modelos de predição baseados em ANN, mediante as seguintes etapas (KAASTRA; BOYD, 1996):

1. Arquitetura - MLP;
2. Topologia - número de camadas e o número de neurônios por camada;
3. Divisão dos dados em treinamento e predição (REZAEI et al., 2015)
  - 85% para treinamento
  - 15% para predição
4. Treinamento da ANN, balanceando a relação entre interpolar exatamente todos os pontos de treinamento e a capacidade de generalizar os resultados;
5. Efetuar predições;
6. Avaliar o desempenho das predições, caso não sejam adequadas, retornar à etapa 1 para refinamento do modelo.

Para implementação dos modelos baseados em ANN propostos, utilizou-se a *toolbox Neural Network*, disponível no Matlab (MATHWORKS, 2017b).

O projeto do modelo ANN, requer a especificação de determinados parâmetros de operação, dentre os mesmos, alguns são gerais a todas as arquiteturas propostas e, neste trabalho, são considerados fixos (MATHWORKS, 2017b):

**Métrica de desempenho** : MSE;

**Algoritmo de aprendizagem** : *Backpropagation* do tipo Levenberg-Marquardt;

**Número máximo de iterações do algoritmo de treinamento** : 1000;

**Função de transferência nas camadas ocultas** : *Tansig*;

**Função de transferência na camada de saída** : *Purelin*;

**Número máximo de verificações para o algoritmo de treinamento** : 6;

**Gradiente mínimo de desempenho para o algoritmo de treinamento** :  $1 \cdot 10^{-7}$ ;

**Taxa de ajuste  $\mu$  do algoritmo Levenberg-Marquardt** : 0,001;

**Fator de decréscimo de  $\mu$**  : 0,1;

**Fator de acréscimo de  $\mu$**  : 10;

**Valor máximo de  $\mu$**  :  $1 \cdot 10^{10}$ .

Conjuntamente, o desempenho desejado, número máximo de iterações, número máximo de verificações, valor máximo de  $\mu$  e o gradiente mínimo de desempenho, constituem o critério de parada do algoritmo de treinamento de Levenberg-Marquardt utilizado no treinamento da ANN implementado pela *toolbox* do Matlab (MATHWORKS, 2017b).

Adicionalmente, seguindo a recomendação de (MATHWORKS, 2017b) para melhorar o desempenho da ANN mediante a redução do efeito da variabilidade atrelada aos valores iniciais dos pesos sinápticos, efetuou-se  $N_{reps}$  repetições do processo de treinamento para o mesmo conjunto de parâmetros, em que, neste trabalho, adotou-se  $N_{reps} = 3$  repetições.

A metodologia utilizada, considera, primeiramente, a MLP simples para a predição da série temporal dos preços de fechamento do IBOVESPA, em que aplica-se somente os valores defasados da mesma na composição da janela temporal. Após, para o mesmo problema, utiliza-se a MLP exógena, em que, com base na Teoria de Dow, compõe-se a janela temporal com os seguintes dados: (i) preço de fechamento, (ii) preço mínimo e (iii) preço máximo.

Desta forma, avalia-se o desempenho das diferentes variações da ANN, para diferentes combinações de configurações, sem e com a presença da variável exógena e, para cada

combinação, aplica-se os critérios do menor MSE, AIC e  $MSE_{norm}$ , ambos considerando o conjunto composto pelos dados de treinamento e predição.

### 3.3.1 Modelo MLP Simples

Para o modelo MLP simples, assim como no modelo ARIMA, inicialmente, aplicou-se o algoritmo ao problema de predição da função de teste definida pela Equação 3.1 e, em seguida, aos dados reais do IBOVESPA.

O modelo da MLP visa aproximar a função  $f(\cdot)$  de causa-efeto entre o valor do preço de fechamento  $P(n)$  e seus valores defasados para  $P(n - 1), \dots, P(n - N_W)$ , em que  $N_W$  representa o comprimento da janela temporal, descrita pelo número de períodos utilizados para predição. Esta relação é descrita pela Equação 3.2.

$$P(n) = f[P(n - 1), \dots, P(n - N_W)] \quad (3.2)$$

Para cada configuração, tem-se a exibição do resultado obtido mediante a variação de quatro parâmetros livres do modelo, são eles:

- $N_W$  - Número de amostras da janela temporal de saída;
- $N_L$  - Número de camadas ocultas;
- $H_l$  - Número de neurônios na camada oculta  $l$ , com  $0 < l \leq N_L$ ;
- $MSE_d$  - MSE desejado para o critério de parada do algoritmo de treinamento.

### 3.3.2 Modelo MLP Exógena

Para o mesmo problema de prever o preço de fechamento do IBOVESPA com base em seus valores passados, conforme a Equação 3.2, vista no caso do MLP simples, nesta seção emprega-se a MLP considerando a variável exógena correspondente às séries temporais dos preços mínimos e máximos do IBOVESPA,  $P_{min}(n)$  e  $P_{max}(n)$ , respectivamente, inseridos na janela temporal com comprimento de  $N_W$  amostras.

O modelo da MLP exógena é expresso pelas equações 3.3 e 3.4.

$$P(n) = f[P(n - 1), \dots, P(n - N_W); u(n - 1), \dots, u(n - N_W)] \quad (3.3)$$

$$u(n) = h[P_{min}(n); P_{max}(n)] \quad (3.4)$$

Assim como no caso da MLP simples, os parâmetros livres a serem otimizados, são:

- $N_W$  - Número de amostras da janela temporal;
- $N_L$  - Número de camadas ocultas;
- $H_l$  - Número de neurônios na camada oculta  $l$ , com  $0 < l \leq N_L$ ;
- $MSE_d$  - MSE desejado para o critério de parada do algoritmo de treinamento.

# Capítulo 4

## Análise dos Resultados

*There is a popular cliché ... which says that you cannot get out of computers any more than you put in. Other versions are that computers only do exactly what you tell them to, and that therefore computers are never creative. The cliché is true only in the crashingly trivial sense, the same sense in which Shakespeare never wrote anything except what his first schoolteacher taught him to write—words.*

—Richard Dawkins

Com base nos conceitos fundamentais vistos no Capítulo 2, neste capítulo exhibe-se os resultados obtidos para os diferentes modelos propostos no Capítulo 3. Para cada modelo implementado, avalia-se os resultados, considerando a evolução da série temporal real, da série predita pelo modelo bem como dos resíduos remanescentes à aplicação do modelo, além disso, exhibe as seguintes métricas de desempenho: (i) MSE, (ii) AIC e (iii)  $MSE_{norm}$ .

O presente capítulo é estruturado da seguinte maneira. Na Seção 4.1 é feita a apresentação e análise dos resultados obtidos com o modelo ARIMA. A Seção 4.2 exhibe os resultados para a ANN do tipo MLP simples, que considera somente valores defasados do preço de fechamento em sua entrada. Já na Seção 4.3 apresenta-se os resultados atingidos com a MLP exógena que inclui em sua entrada, além do preço de fechamento, os valores da série temporal definida pelos preços mínimos e máximos do IBOVESPA. Por fim, na

Seção 4.4 faz-se a comparação do desempenho atingido pelos diferentes modelos utilizados na predição do IBOVESPA.

## 4.1 Resultados do Modelo ARIMA

Nesta seção, descreve-se a aplicação do modelo ARIMA ao problema de predição do IBOVESPA. Inicialmente, considerou-se a predição de uma série temporal de teste obtida a partir da Equação 3.1, criada com o intuito de representar o comportamento aleatório do IBOVESPA nas análises iniciais, possibilitando a verificação da corretude do algoritmo implementado para o modelo ARIMA.

### 4.1.1 Série Temporal de Teste

Primeiramente, avaliou-se a função de autocorrelação e autocorrelação parcial da série temporal de teste, exibidas nas figuras 4.1 e 4.2, respectivamente.

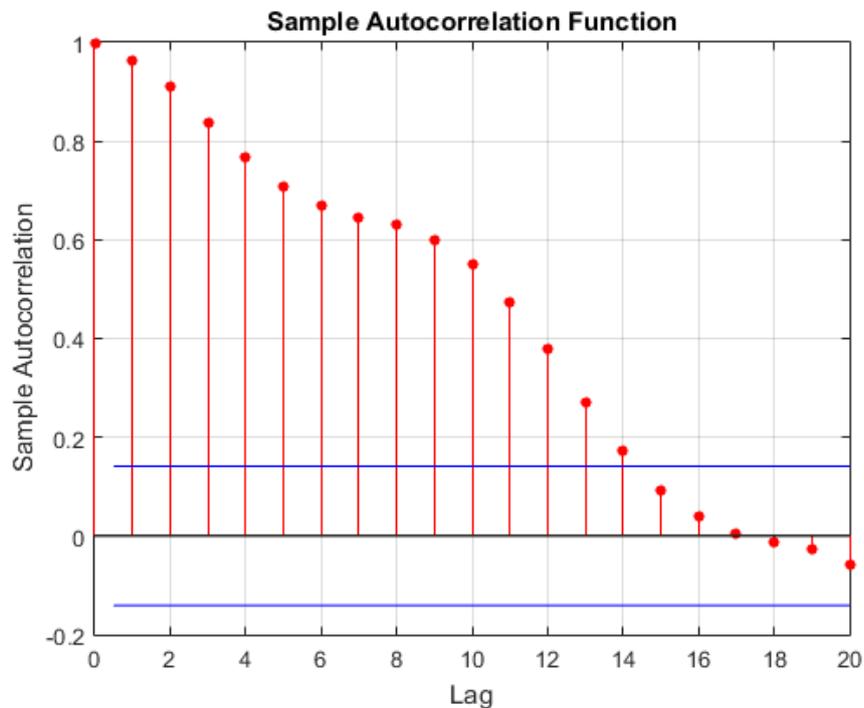


Figura 4.1: Autocorrelação da Série definida pela Equação 3.1.

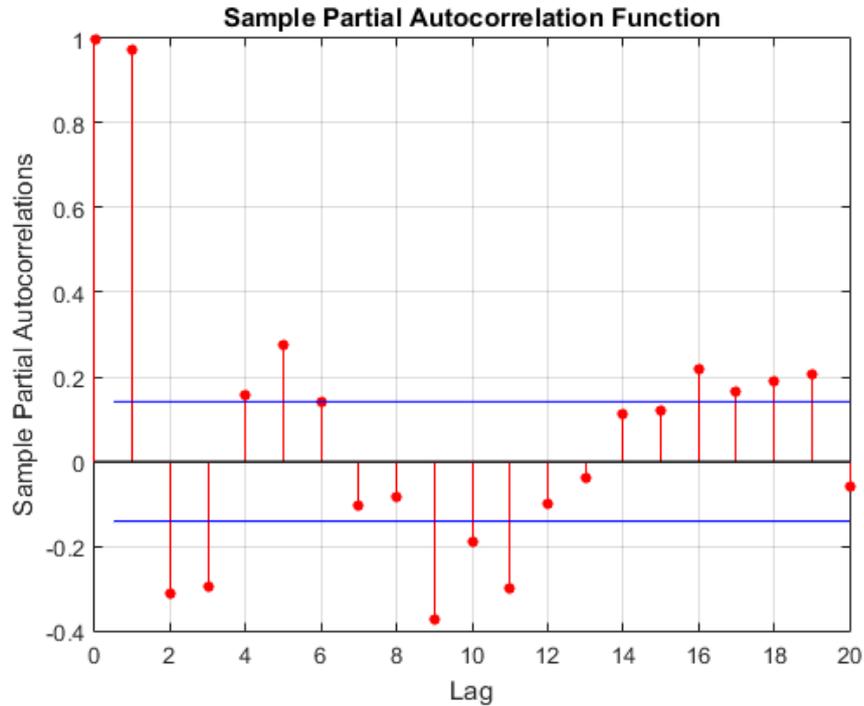


Figura 4.2: Autocorrelação Parcial da Série definida pela Equação 3.1.

Nas figuras 4.1 e 4.2, as duas barras horizontais em destaque na cor azul, delimitam o intervalo de insignificância estatística, dentro do qual os valores da autocorrelação e autocorrelação parcial não apresentam valores estatisticamente confiáveis (MATHWORKS, 2017a).

A partir da análise da função de autocorrelação, nota-se seu lento decaimento, sendo estatisticamente insignificante para atrasos superiores a 18 amostras. Já com relação a função de autocorrelação parcial, nota-se que a mesma cai abruptamente após o segundo atraso, oscilando entre valores que adentram e saem da faixa de relevância estatística.

Com esta análise, supõe-se, a priori, que a ordem de ambos os modelos  $AR(p)$  e  $MA(p)$  encontra-se na faixa  $[0, 18]$ , em que emprega-se o método de busca exaustiva para a determinação da melhor estimativa para  $p$  e  $q$  dentro da faixa de valores pré-estabelecida, considerando como critério de otimalidade o menor MSE global (treinamento e predição). Enquanto que, para a ordem do modelo  $I(d)$ , sua identificação é realizada mediante o método de busca exaustiva considerando à heurística de que seu valor encontra-se na faixa  $[0, 2]$ , isto é, no máximo, a diferenciação de segunda ordem é suficiente para tornar a série temporal estacionária.

Desta forma, o procedimento de busca percorreu os seguintes conjuntos:  $p = \{0, 1, 2\}$ ,  $q = \{0, 1, 2\}$  e  $d = \{0, 1, 2\}$  e alguns dos resultados obtidos são exibidos na Tabela 4.1.

Tabela 4.1: Desempenho do Modelo ARIMA para a Série definida pela Equação 3.1.

AR( $\mathbf{p}$ )	I( $\mathbf{d}$ )	MA( $\mathbf{q}$ )	MSE	AIC
1	0	0	5,58	-3,57
2	0	0	3,93	-3,91
0	0	1	4,85	-3,71
0	0	2	5,56	-3,57
1	1	0	9,60	-3,02
2	1	0	9,08	-3,07
0	1	1	10,70	-2,92
0	1	2	14,50	-2,60
1	0	1	4,89	-3,70
1	1	1	15,90	-2,51
2	1	2	3,55	-3,99

Conforme a Tabela 4.1, o resultado da identificação do modelo otimizado é  $p = 2$ ,  $d = 1$  e  $q = 2$ , ou seja, trata-se do modelo ARIMA(2, 1, 2). Com esta configuração do modelo ARIMA para prever a série temporal de teste, atingiu-se o MSE de 3,55 e AIC de -3,99, em que o valor negativo deve-se ao fato do MSE ser inferior a 1, tornando a função  $\ln(\cdot)$  pertencente à definição do critério AIC, negativa.

Já a estimação dos parâmetros do modelo foi realizada com o algoritmo ML disponibilizado pela *toolbox* do Matlab (MATHWORKS, 2017a), que obteve a parcela constante do modelo ARIMA sendo 0,00978414 e os seguintes coeficientes dos modelos AR(2) e MA(2), respectivamente:  $\phi_1 = 1,46605$ ;  $\phi_2 = -0,849281$  e  $\theta_1 = -1,28538$ ;  $\theta_2 = 0,826373$ .

O comportamento da série temporal do preço de fechamento do IBOVESPA, em conjunto com a série predita pelo modelo ARIMA(2, 1, 2) e do erro de predição entre a série temporal real e sua predição, são exibidos na Figura 4.3.

Considerando que, a variância da série temporal é 3,61 e o MSE ótimo é 3,55, isto corresponde ao MSE normalizado  $MSE_{norm} = 9,82 \cdot 10^{-1}$  ou 98,2%. Desta forma, tem-se que o modelo ARIMA implementado não é capaz de explicar corretamente o comportamento observado da série temporal e, com isso, reduzir a variabilidade do erro de predição.

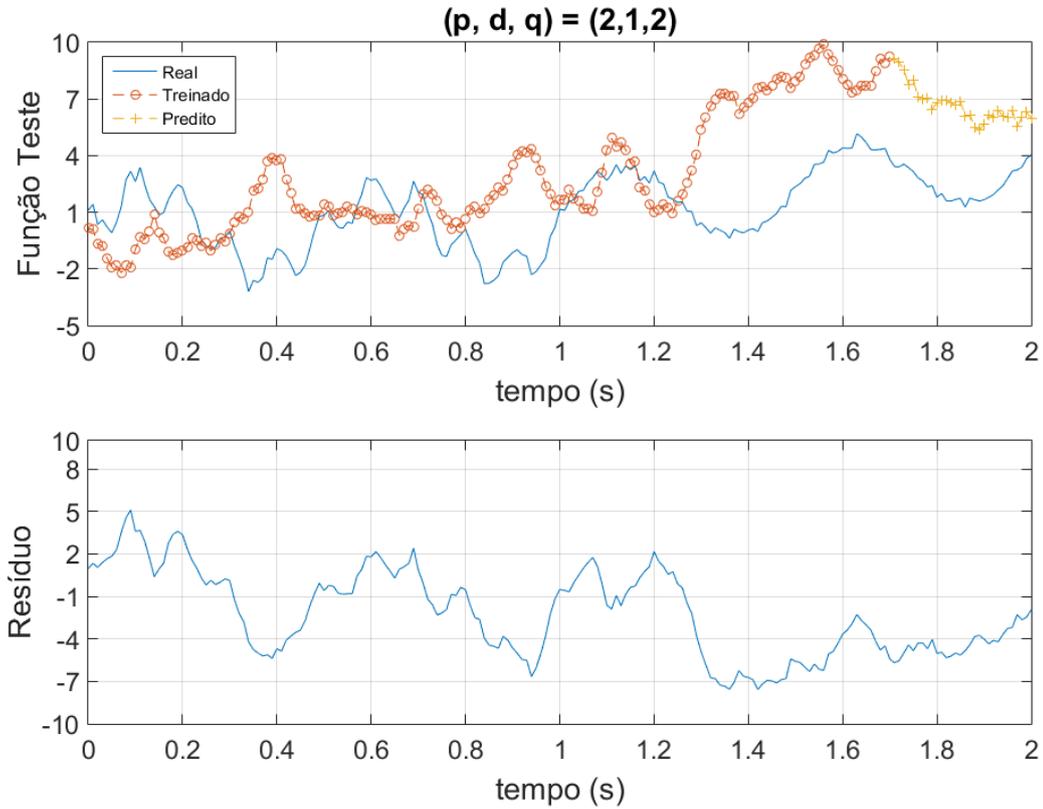


Figura 4.3: Comportamento do Modelo ARIMA para a Série Definida pela Equação 3.1.

A análise da função de teste apresentou-se importante para entender o processo de representação de séries temporais mediante o modelo ARIMA, identificar seus principais parâmetros de desempenho e verificar a corretude do algoritmo implementado.

#### 4.1.2 Dados Reais do IBOVESPA

Após a análise da função de teste, considera-se o mesmo procedimento para o problema de predição da série temporal definida pelo preço de fechamento do IBOVESPA segundo o modelo ARIMA.

Assim como no caso da função de teste, inicialmente exibe-se a função de autocorrelação da série temporal referente ao preço de fechamento do IBOVESPA na Figura 4.4.

A análise da função de autocorrelação revela seu decaimento suave, não exibindo valores relativamente baixos para o atraso a partir do qual a autocorrelação torna-se estatís-

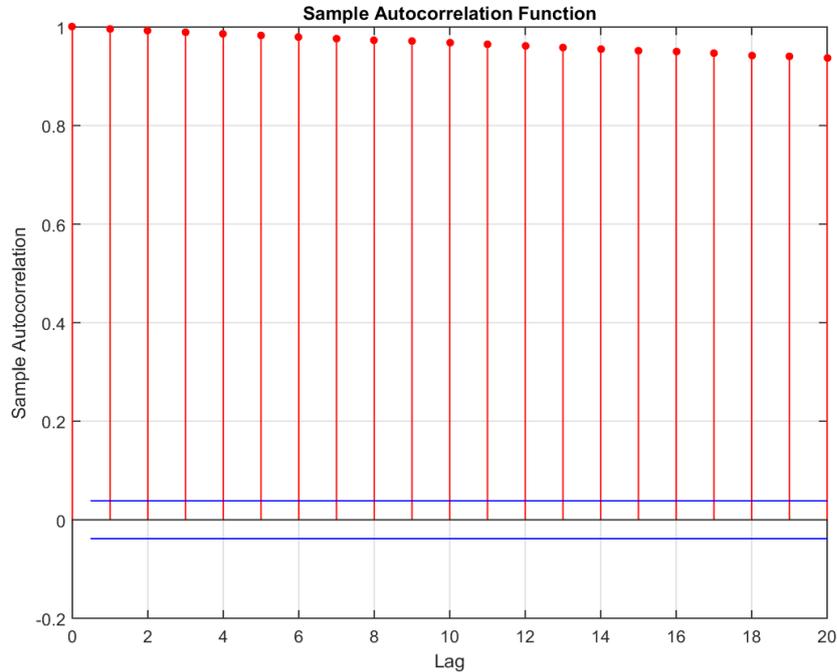


Figura 4.4: Autocorrelação para Dados Reais do IBOVESPA.

ticamente insignificante, isto indica uma forte dependência da função de autocorrelação com o primeiro atraso, que exerce influência estatística nos atrasos de ordem superior. Este fato é corroborado mediante a análise da função de autocorrelação parcial exibida na Figura 4.5, em que, após o primeiro atraso, a mesma decai na faixa de insignificância estatística. Logo, os valores ótimos de  $p$  e  $q$ , encontram-se na faixa  $[0, 2]$ .

Analogamente ao caso da função de teste, nas figuras 4.4 e 4.5, as duas barras horizontais em destaque na cor azul, representam a região de insignificância estatística, dentro da qual os valores da autocorrelação e autocorrelação parcial não apresentam valores estatisticamente confiáveis (MATHWORKS, 2017a).

Desta forma, assim como na análise realizada na série temporal representada pela função de teste, a determinação da ordem do modelo  $I(d)$  é feita mediante à heurística de que seu valor encontra-se na faixa  $[0, 2]$ .

Desta forma, o procedimento de busca percorreu os seguintes conjuntos:  $p = \{0, 1, 2\}$ ,  $q = \{0, 1, 2\}$  e  $d = \{0, 1, 2\}$  e alguns dos resultados obtidos são exibidos na Tabela 4.2.

Segundo a análise da Tabela 4.2, o resultado otimizado da identificação com o método da busca exaustiva é  $p = 2$ ,  $d = 0$  e  $q = 0$ , ou seja, trata-se portanto do modelo ARIMA(2,

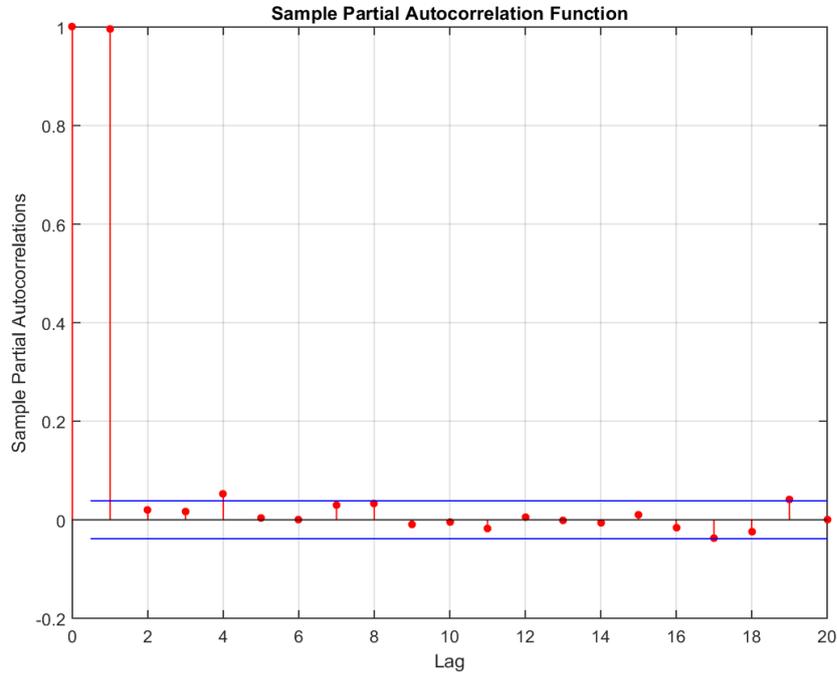


Figura 4.5: Autocorrelação Parcial para Dados Reais do IBOVESPA.

0, 0) que recaiu no modelo AR(2), resultando em um MSE de  $1,21 \cdot 10^8$  e um AIC de 10,70.

Já a estimação dos parâmetros do modelo foi realizada com o algoritmo ML disponibilizado pelo Matlab, que obteve a parcela constante do modelo AR sendo 181,863 e os seguintes coeficientes dos modelos AR(2):  $\phi_1 = 0,97588$ ;  $\phi_2 = 0,0207412$ .

A série temporal real do preço de fechamento do IBOVESPA, em conjunto com sua predição realizada pelo modelo AR(2) e o erro de predição obtido são exibidos na Figura 4.6.

Ao considerar o valor obtido para o MSE de  $1,21 \cdot 10^8$  e a variância da série temporal real de  $1,5 \cdot 10^8$ , tem-se que o  $MSE_{norm}$  é de  $8,40 \cdot 10^{-1}$ , ou seja, 84,0%, indicando que o modelo proposto não é capaz de explicar satisfatoriamente os dados e com isso reduzir a variabilidade do erro de predição.

Tabela 4.2: Desempenho do Modelo ARIMA para Dados do IBOVESPA.

AR( <b>p</b> )	I( <b>d</b> )	MA( <b>q</b> )	MSE ( $10^8$ )	AIC
1	0	0	1,54	10,90
2	0	0	1,21	10,70
0	0	1	4,08	11,90
0	0	2	7,20	12,50
1	1	0	1,34	10,80
2	1	0	4,66	12,10
0	1	1	2,24	11,30
0	1	2	2,22	11,30
1	0	1	1,35	10,80
1	1	1	2,00	11,20
2	1	2	2,55	11,50

## 4.2 Resultados do Modelo MLP Simples

Nesta seção descreve-se a aplicação da arquitetura MLP simples ao problema de predição do IBOVESPA. Primeiramente, considerou-se a predição de uma série temporal de teste obtida a partir da Equação 3.1, possibilitando a verificação do algoritmo implementado.

Com base no procedimento de quatro etapas, realizou-se a subdivisão dos dados em 85% para treinamento e 15% para predição e empregou-se o método de busca exaustiva para otimização da MLP simples considerando as métricas dos menores MSE e AIC globais com a variação dos seguintes parâmetros: comprimento da janela temporal ( $N_W$ ), número de camadas ocultas ( $N_L$ ), o número de neurônios por camada ( $H_l, 1 \leq l \leq N_L$ ) e critério de parada ( $MSE_d$ ).

### 4.2.1 Série Temporal de Teste

Para cada configuração, determinou-se o MSE e o AIC do conjunto de dados completos, *i.e.*, treinamento e predição. Inicialmente, o  $MSE_d$  utilizado correspondeu ao valor padrão definido pela *toolbox* do Matlab sendo  $MSE_d = 0$ . Desta forma, alguns dos resultados obtidos para a variação dos parâmetros  $N_W \{2, 7, 14, 18\}$ ,  $N_L \{1\}$  e  $H_1 \{10, 30, 60, 100\}$ , são listados na Tabela 4.3.

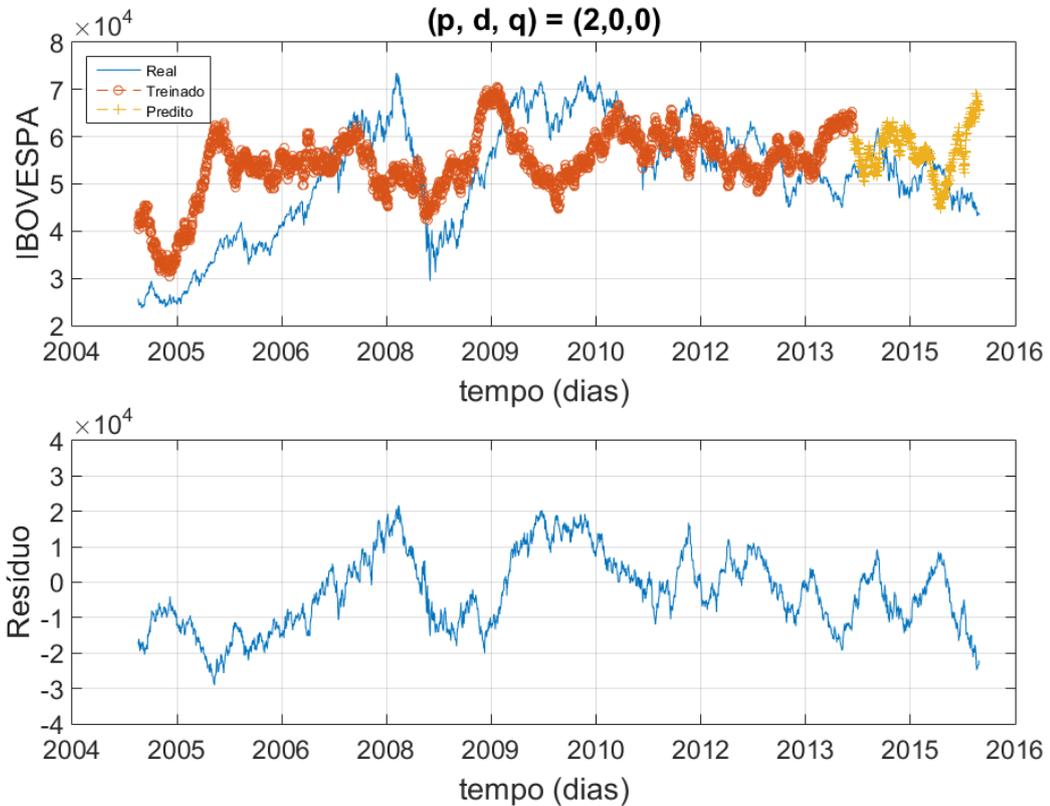


Figura 4.6: Comportamento do Modelo ARIMA para Dados do IBOVESPA.

A inspeção dos resultados exibidos na Tabela 4.3 mostra que, o AIC eleva-se conforme eleva-se o número de neurônios, pois, conforme descrito pela Equação 2.38, o AIC é diretamente proporcional ao parâmetro de penalização  $N_p$ . Enquanto que, o MSE não exibe relação direta com o número de neurônios, demonstrando que a curva de erro  $MSE(H_1)$  exibe múltiplos pontos de mínimo. Entretanto, com relação ao comprimento da janela temporal  $N_W$ , nota-se que, enquanto o AIC exibe consistência, com pequenas variações conforme aumenta-se o comprimento da janela temporal, o MSE exibe comportamento oscilatório. Estas oscilações podem ser explicadas verificando que, para janelas temporais com poucas amostras, o modelo exibe baixa capacidade de memória, não adequando-se satisfatoriamente às correlações presentes nos dados, ao passo que, para janelas compostas por muitas amostras, o modelo torna-se demasiadamente sensível ao ruído inerente aos dados.

Segundo a Tabela 4.3, para o critério de parada  $MSE_d = 0$ , a configuração otimizada

Tabela 4.3: Desempenho da MLP Simples para a Série definida pela Equação 3.1.

$N_W$	$H_1$	<b>MSE</b> ( $10^{-1}$ )	<b>AIC</b>
2	10	3,20	-0,73
2	30	6,60	0,79
2	60	34,70	3,64
2	100	17,10	4,53
7	10	2,40	-0,52
7	30	7,22	2,37
7	60	15,30	5,81
7	100	8,44	11,10
14	10	2,86	0,35
14	30	0,91	2,39
14	60	2,80	8,29
14	100	3,25	14,80
18	10	3,03	0,80
18	30	7,41	5,68
18	60	1,22	12,10
18	100	18,50	20,50

resultante no menor MSE global igual a  $9,13 \cdot 10^{-2}$  e seu AIC associado de 2,39 é composta pela janela temporal de  $N_W = 14$  pontos,  $N_L = 1$  camada formada por  $H_1 = 30$  neurônios.

Após a determinação da configuração otimizada  $(N_W, H_1) = (14, 30)$ , variou-se o critério de parada  $MSE_d$ , de 0 para 1,81; 0,36 e 0,04, que correspondem a 50%, 10% e 1% da variância da série temporal ( $VAR_{teste} = 3,61$ ). Os resultados obtidos para estes critérios são exibidos na Tabela 4.4.

Tabela 4.4: Critérios de Parada da MLP Simples para a Série definida pela Equação 3.1.

<b>MSE<sub>d</sub></b>	<b>MSE</b> ( $10^{-1}$ )	<b>AIC</b>
1,81	0,15	5,20
0,36	1,32	2,76
0,04	1,32	2,76

Segundo a Tabela 4.4, o menor MSE é  $1,51 \cdot 10^{-1}$ , que corresponde ao  $MSE_d = 1,81$ , entretanto, o mesmo é, aproximadamente, 39,5% superior ao MSE de  $9,13 \cdot 10^{-2}$  atingido para o caso  $MSE_d = 0$ .

Logo, mediante a análise dos parâmetros considerados, o MSE otimizado é  $9,13 \cdot 10^{-2}$  e seu AIC associado é 2,39, atingido para  $MSE_d = 0$  e  $(N_W, H_1) = (14, 30)$ .

O comportamento da série temporal de teste, em conjunto com a predição efetuada pela MLP simples, bem como o erro de predição são exibidos na Figura 4.7.

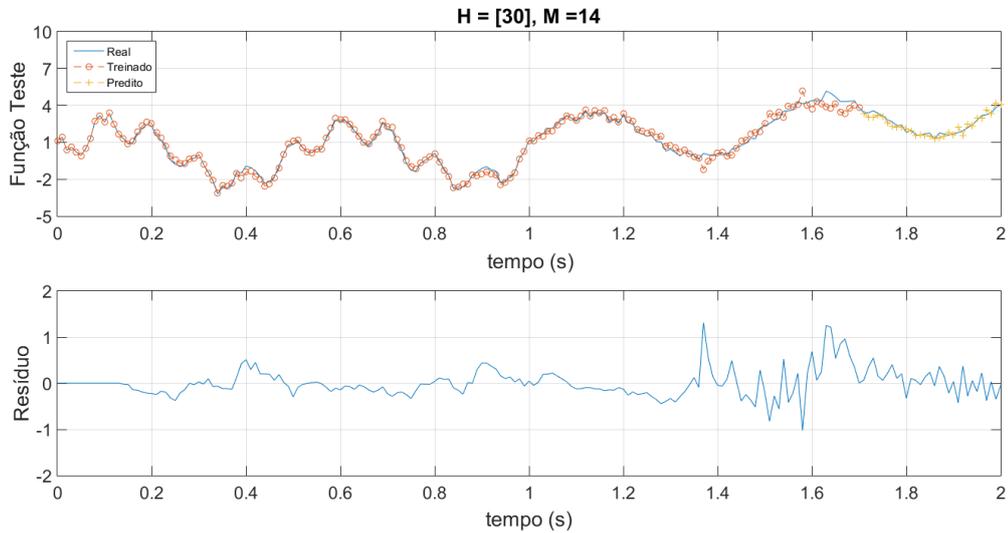


Figura 4.7: Comportamento da MLP Otimizada para a Série definida pela Equação 3.1.

Considerando que a variância da série temporal é 3,61 e o MSE ótimo é  $9,13 \cdot 10^{-2}$ , isto corresponde ao  $MSE_{norm}$  de  $2,53 \cdot 10^{-2}$  ou 2,53%. Demonstrando a capacidade do modelo em explicar a série temporal, reduzindo a variabilidade do erro de predição.

Por fim, nota-se que o resíduo é aproximadamente constante para dados de treinamento e apresenta oscilações quando o modelo é aplicado aos dados de predição, isto é condizente com o fato de que a tendência de um modelo de predição é de desviar-se do valor real da série temporal conforme aumenta-se o horizonte de predição (MONTGOMERY; JENNINGS, 2008).

A análise da função de teste apresentou-se importante para entender o processo de representação de séries temporais mediante o modelo MLP, identificar seus principais parâmetros de desempenho e verificar a corretude do algoritmo implementado.

### 4.2.2 Dados Reais do IBOVESPA

Após a análise dos resultados obtidos com a série temporal de teste definida pela Equação 3.1, aplicou-se a MLP simples ao problema de predição da série temporal do preço de fechamento do IBOVESPA com base em seus valores passados, considerando o período de 01/01/2005 até 31/12/2015.

Assim como no caso da função de teste, na análise do modelo de ANN aplicado ao problema de predição do IBOVESPA, adotou-se o método de busca exaustiva para realizar a otimização da configuração da ANN segundo o critério do menor MSE global.

Inicialmente, assim como no caso da função de teste, o  $MSE_d$  utilizado correspondeu ao valor padrão definido pela *toolbox* do Matlab sendo  $MSE_d = 0$  e com este critério de parada, o método de busca exaustiva compreendeu, a priori, na consideração de  $N_L = 1$  camada oculta (o caso  $N_L = 2$  camadas ocultas é analisado posteriormente) e no aninhamento de 2 laços de repetição, cada qual responsável pela variação de um parâmetro livre da MLP, a saber: comprimento da janela temporal  $N_W \{2, 7, 10, 14\}$  e o número de neurônios na camada oculta  $H_1 \{20, 40, 50, 80\}$ , os resultados obtidos são listados na Tabela 4.5.

Ao avaliar os resultados exibidos na Tabela 4.5, nota-se que, de maneira análoga ao caso da função de teste, ao aumentar o número de neurônios, o AIC aumenta, isto pois aumenta-se o parâmetro de penalização  $N_p$  da Equação 2.38 que é diretamente proporcional ao AIC. Já com relação ao MSE, este não apresenta relação direta com o número de neurônios, em que a curva de erro  $MSE(H_1)$  exibe múltiplos pontos de mínimo. Entretanto, com relação ao comprimento da janela temporal  $N_W$ , nota-se que, enquanto o AIC exibe consistência, com pequenas variações conforme aumenta-se o comprimento da janela temporal, já o MSE exibe comportamento oscilatório, que pode ser explicado verificando que, para janelas temporais pequenas, o modelo exibe baixa capacidade de memorização das correlações presentes nos dados, ao passo que, o aumento da janela torna o modelo mais sensível a ruídos existentes nos dados.

A partir da Tabela 4.5, para o critério de parada  $MSE_d = 0$  e  $N_L = 1$ , a configuração otimizada resultante no menor MSE global igual a  $6,74 \cdot 10^5$  e AIC associado de 13,60 é composta pela janela temporal de  $N_W = 10$  pontos e  $H_1 = 40$  neurônios.

Após a determinação da configuração otimizada  $(N_W, H_1) = (10, 40)$ , variou-se o critério de parada  $MSE_d$ , de 0 para 7,23; 1,45 e 0,15, que correspondem a 50%, 10% e 1%

Tabela 4.5: Desempenho da MLP Simples para Dados do IBOVESPA e  $N_L = 1$ .

$N_W$	$H_1$	MSE ( $10^5$ )	AIC
2	20	7,71	13,60
2	40	7,99	13,70
2	50	8,86	13,80
2	80	7,84	13,80
7	20	7,25	13,60
7	40	6,96	13,70
7	50	7,04	13,80
7	80	6,98	14,00
10	20	7,55	13,70
10	40	6,74	13,80
10	50	6,93	13,90
10	80	7,79	14,30
14	20	7,23	13,70
14	40	7,82	14,00
14	50	6,88	14,00
14	80	6,77	14,40

da variância da série temporal ( $VAR_{teste} = 1,45 \cdot 10^8$ ). Os resultados obtidos para estes critérios são exibidos na Tabela 4.6.

Tabela 4.6: Critérios de Parada da MLP Simples para Dados do IBOVESPA e  $N_L = 1$ .

$MSE_d$ ( $10^5$ )	MSE ( $10^6$ )	AIC
7,23	25,80	17,40
1,45	2,95	15,30
0,15	7,50	13,90

Segundo a Tabela 4.6, ao alterar o critério de parada  $MSE_d$  entre os valores propostos, variou-se o MSE, sendo que o MSE ótimo é  $2,95 \cdot 10^6$ , aproximadamente, além disso, este valor é, aproximadamente, 77,2% superior ao MSE de  $6,74 \cdot 10^5$  atingido para o caso  $MSE_d = 0$ . Logo, mediante a análise dos parâmetros considerados, o MSE otimizado é

$6,74 \cdot 10^5$  e seu AIC associado é 13,60, atingido para  $MSE_d = 0$  e  $(N_W, H_1) = (10, 40)$ .

Após a avaliação da MLP simples com  $N_L = 1$ , procedeu-se para a mesma análise, mas desta vez com a MLP constituída por duas camadas ocultas, *i.e.*  $N_L = 2$ , conforme (HAYKIN, 2008), duas camadas ocultas são suficientes para aproximar satisfatoriamente funções contínuas e assim não avaliou-se os casos em que  $N_L > 2$ . Inicialmente, considerando o critério de parada padrão da *toolbox* do Matlab  $MSE_d = 0$ , variou-se os seguintes parâmetros:  $N_W \{2, 4, 7, 10, 14\}$ ,  $H_1 \{20, 40, 80\}$  e  $H_2 \{10, 40, 50\}$ . Os resultados obtidos para as configurações citadas são exibidos na Tabela 4.7.

Tabela 4.7: Desempenho da MLP Simples para Dados do IBOVESPA e  $N_L = 2$ .

$N_W$	$H_1$	$H_2$	MSE ( $10^5$ )	AIC
2	20	10	7,77	13,80
2	20	40	7,90	14,30
2	40	10	8,16	14,00
2	80	50	8,63	16,90
4	20	10	10,00	14,10
4	20	40	8,04	14,30
4	40	10	7,34	14,60
4	80	50	9,17	17,00
7	20	10	7,24	13,80
7	20	40	7,98	14,40
7	40	10	14,00	14,70
7	80	50	8,31	17,10
10	20	10	8,12	13,90
10	20	40	6,56	14,20
10	40	10	6,93	14,10
10	80	50	8,47	17,30
14	20	10	7,71	13,90
14	20	40	7,21	14,40
14	40	10	8,44	21,13
14	80	50	9,38	17,70

Semelhante as análises anteriores, ao avaliar os resultados exibidos na Tabela 4.7, tem-

se que, ao aumentar a soma do número de neurônios nas duas camadas, o AIC aumenta, isto deve-se ao fato de que, conforme a Equação 2.38, o parâmetro de penalização  $N_p$  é diretamente proporcional ao AIC. De outro modo, o MSE não apresenta relação direta com o número de neurônios, em que a superfície de erro  $MSE(H_1, H_2)$  indica conter múltiplos pontos de mínimo. Entretanto, com relação ao comprimento da janela temporal  $N_W$ , nota-se que, enquanto o AIC exibe consistência, variando marginalmente conforme aumenta-se o comprimento da janela temporal, o MSE exibe comportamento oscilatório, que pode ser explicada mediante à constatação de que, para janelas temporais pequenas, o modelo exibe baixa capacidade de memória, não adequando-se bem às correlações presentes nos dados, ao passo que, para janelas grandes o modelo torna-se demasiadamente sensível ao ruído contido nos dados.

A partir a Tabela 4.7, para o critério de parada  $MSE_d = 0$  e  $N_L = 2$ , a configuração otimizada resultante no menor MSE global igual a  $6,56 \cdot 10^5$  e AIC associado de 14,20 é composta pela janela temporal de  $N_W = 10$  pontos,  $H_1 = 20$  e  $H_2 = 40$  neurônios na primeira e segunda camadas ocultas, respectivamente.

Após a determinação da configuração otimizada  $(N_W, H_1, H_2) = (10, 20, 40)$ , variou-se o critério de parada  $MSE_d$ , de 0 para 7,23; 1,45 e 0,15, que correspondem a 50%, 10% e 1% da variância da série temporal ( $VAR_{teste} = 1,45 \cdot 10^8$ ). Os resultados obtidos para estes critérios são exibidos na Tabela 4.8.

Tabela 4.8: Critérios de Parada da MLP Simples para Dados do IBOVESPA e  $N_L = 2$ .

$MSE_d$ ( $10^5$ )	MSE ( $10^6$ )	AIC
7,23	23,10	17,80
1,45	7,77	16,70
0,15	0,87	14,50

Segundo a Tabela 4.8, ao alterar o critério de parada  $MSE_d$  entre os valores propostos, implicou-se na variação do MSE, sendo que, o valor ótimo é igual a  $8,70 \cdot 10^5$ , sendo, aproximadamente, 24,6% superior ao MSE de  $6,56 \cdot 10^5$  atingido para o caso  $MSE_d = 0$ .

Logo, mediante a análise dos parâmetros considerados, o MSE otimizado é  $6,56 \cdot 10^5$  e seu AIC associado é 14,20, atingido para  $MSE_d = 0$  e  $(N_W, H_1, H_2) = (10, 20, 40)$ , esta configuração é exibida na Figura 4.8.

O comportamento da série temporal do IBOVESPA, conjuntamente com a série tem-

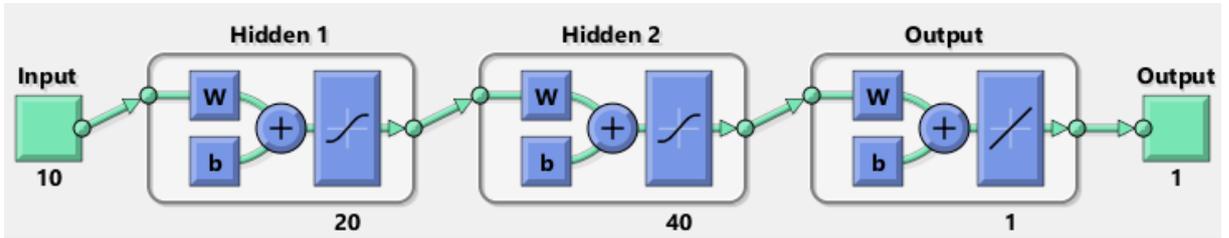


Figura 4.8: Configuração da MLP Simples Otimizada para Dados do IBOVESPA.

poral predita pelo modelo MLP simples otimizado e o erro de predição, são exibidos na Figura 4.9.

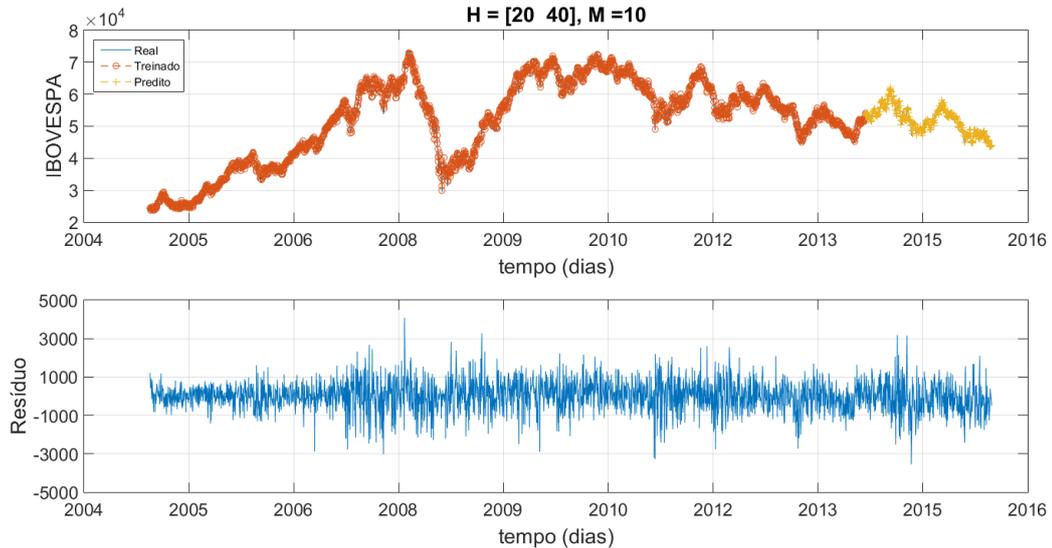


Figura 4.9: Comportamento da MLP Simples Otimizada para Dados do IBOVESPA.

O valor  $6,56 \cdot 10^5$ , obtido para o MSE pode parecer elevado, entretanto, a variância da série temporal real é  $1,45 \cdot 10^8$  e com isso, o  $MSE_{norm}$  é  $4,53 \cdot 10^{-3}$ , ou seja, 0,45% da variância da série temporal real, indicando que o modelo da MLP simples é capaz de explicar satisfatoriamente os dados observados e com isso, reduzir significativamente a variabilidade do erro de predição.

### 4.3 Resultados do Modelo MLP Exógeno

Nesta seção, descreve-se a aplicação da arquitetura MLP exógena ao problema de predição do IBOVESPA. Diferentemente da MLP simples, na MLP exógena considera-se a presença de uma variável explicativa à entrada da ANN, no caso, tem-se a série temporal composta pela junção dos preços mínimos e máximos do IBOVESPA, bem como de seus valores defasados.

De forma semelhante as análises realizadas anteriormente, o critério de parada adotado, inicialmente, correspondeu ao valor padrão definido pela *toolbox* do Matlab, em que  $MSE_d$  é igual a 0 e com este critério de parada, o método de busca exaustiva compreendeu, a priori, na consideração de 1 camada oculta ( $N_L = 1$ ) (o valor  $N_L = 2$  camadas ocultas é analisado posteriormente) e no aninhamento de 2 laços de repetição, cada qual responsável pela variação de um parâmetro livre da MLP, a saber: comprimento da janela temporal  $N_W \{2, 7, 10, 14\}$  e o número de neurônios na camada oculta  $H_1 \{20, 40, 50, 80\}$ . Alguns destes resultados são listados na Tabela 4.9.

Ao avaliar os resultados exibidos na Tabela 4.9, nota-se que, de forma análoga às demais análises realizadas, ao aumentar o número de neurônios, o AIC aumenta, isto pois aumenta-se o parâmetro de penalização  $N_p$  conforme descrito pela Equação 2.38. Enquanto que, com relação ao MSE, este não apresenta relação direta com o número de neurônios, em que a curva de erro  $MSE(H_1)$  exibe múltiplos pontos de mínimo. Entretanto, com relação ao comprimento da janela temporal  $N_W$ , nota-se que, enquanto o AIC é consistente, apresentando pequenas variações conforme aumenta-se o comprimento da janela temporal, o MSE exibe comportamento demasiadamente oscilatório, que pode ser explicado verificando que, para janelas temporais pequenas, o modelo apresenta baixa capacidade de memorização das correlações presentes nos dados, ao passo que, o aumento da janela torna o modelo mais sensível a ruídos existentes nos dados.

A partir a Tabela 4.9, para o critério de parada  $MSE_d = 0$  e  $N_L = 1$ , a configuração otimizada resultante no menor MSE global igual a  $6,23 \cdot 10^5$  e AIC associado de  $5,75$  é composta pela janela temporal de  $N_W = 10$  pontos e  $H_1 = 80$  neurônios.

Após a determinação da configuração otimizada  $(N_W, H_1) = (10, 80)$ , variou-se o critério de parada  $MSE_d$ , de 0 para  $7,23$ ;  $1,45$  e  $0,15$ , que correspondem a 50%, 10% e 1% da variância da série temporal ( $VAR_{teste} = 1,45 \cdot 10^8$ ). Os resultados obtidos para estes critérios são exibidos na Tabela 4.10.

Tabela 4.9: Desempenho da MLP Exógena para Dados do IBOVESPA e  $N_L = 1$ .

$N_W$	$H_1$	MSE ( $10^5$ )	AIC
2	20	7,56	5,75
2	40	7,34	5,83
2	50	7,32	5,89
2	80	7,51	6,09
7	20	6,42	5,80
7	40	6,49	6,15
7	50	6,46	6,32
7	80	6,37	6,81
10	20	6,41	5,930
10	40	6,56	6,43
10	50	6,35	6,63
10	80	6,23	7,32
14	20	6,56	6,13
14	40	6,34	6,75
14	50	6,58	7,11
14	80	6,78	8,11

Segundo a Tabela 4.10, a variação do critério de parada  $MSE_d$  entre os valores propostos, acarretou no menor MSE de  $7,66 \cdot 10^5$  que corresponde a  $MSE_d = 0,15 \cdot 10^5$ , além disso, este valor é, aproximadamente, 18,7% superior ao MSE de  $6,23 \cdot 10^5$  atingido para o caso  $MSE_d = 0$ . Logo, mediante a análise dos parâmetros considerados, o MSE otimizado é  $6,23 \cdot 10^5$  e seu AIC associado é 7,32, atingido para  $MSE_d = 0$  e  $(N_W, H_1) = (10, 80)$ .

Após a avaliação da MLP exógena com  $N_L = 1$ , procedeu-se para a mesma análise, mas desta vez, com a MLP formada por duas camadas ocultas, *i.e.*  $N_L = 2$ . Inicialmente, considerando o critério de parada padrão da *toolbox* do Matlab  $MSE_d = 0$ , variou-se os seguintes parâmetros:  $N_W \{2, 4, 7, 10, 14\}$ ,  $H_1 \{20, 40, 80\}$  e  $H_2 \{10, 40, 50\}$ . Os resultados obtidos para as configurações citadas são exibidos na Tabela 4.11.

Semelhante ao caso da MLP simples, ao avaliar os resultados exibidos na Tabela 4.11, nota-se que ao aumentar o número total de neurônios, o AIC aumenta, isto pois aumenta-se o parâmetro de penalização  $N_p$ , que, conforme descrito pela Equação 2.38 é

Tabela 4.10: Critérios de Parada da MLP Exógena para Dados do IBOVESPA e  $N_L = 1$ .

$MSE_d (10^5)$	$MSE (10^5)$	AIC
7,23	39,30	9,16
1,45	27,00	8,78
0,15	7,66	7,52

diretamente proporcional ao AIC. Já o MSE não apresenta relação direta com o número de neurônios, indicando que a superfície de erro  $MSE(H_1, H_2)$  exibe múltiplos pontos de mínimo. Entretanto, ao avaliar o comprimento da janela temporal  $N_W$ , tem-se que, enquanto o AIC exibe consistência, tendendo a subir conforme aumenta-se o comprimento da janela temporal, o MSE exibe comportamento oscilatório, que pode ser explicado verificando que, para janelas temporais compostas por poucas amostras, o modelo não adquire a capacidade adequada de memória e com isso, não ajusta-se satisfatoriamente às correlações existente nos dados, ao passo que, para janelas maiores, o modelo torna-se sensível ao ruído presente nos dados.

A partir a Tabela 4.11, para o critério de parada  $MSE_d = 0$  e  $N_L = 2$ , a configuração otimizada resultante no menor MSE global igual a  $6,23 \cdot 10^5$  e seu AIC associado é  $10,30$  é composta pela janela temporal de  $N_W = 10$  pontos,  $H_1 = 80$  e  $H_2 = 50$  neurônios na primeira e segunda camadas ocultas, respectivamente.

Assim como no caso  $N_L = 1$ , após a determinação da configuração otimizada para  $N_L = 2$ ,  $(N_W, H_1, H_2) = (10, 80, 50)$ , variou-se o critério de parada  $MSE_d$ , de 0 para 7,23; 1,45 e 0,15, que correspondem a 50%, 10% e 1% da variância da série temporal ( $VAR_{teste} = 1,45 \cdot 10^8$ ). Os resultados obtidos para estes critérios são exibidos na Tabela 4.12.

Segundo a Tabela 4.12, a variação do critério de parada  $MSE_d$  entre os valores propostos, acarretou no menor MSE de  $28,70 \cdot 10^5$  que corresponde a  $MSE_d = 0,15 \cdot 10^5$ , mas este valor é, aproximadamente, 78,2% superior ao MSE  $6,23 \cdot 10^5$  atingido para o caso  $MSE_d = 0$ . Logo, mediante a análise dos parâmetros considerados, o MSE otimizado é  $6,23 \cdot 10^5$  e seu AIC associado é  $10,30$ , atingido para  $MSE_d = 0$  e  $(N_W, H_1, H_2) = (10, 80, 50)$ .

Ao comparar os casos  $N_L = 1$  e  $N_L = 2$ , tem-se que para ambos, o MSE obtido é  $6,23 \cdot 10^5$ , entretanto, o caso  $N_L = 1$  apresenta o AIC igual a 7,32, enquanto que, para

Tabela 4.11: Desempenho da MLP Exógena para Diferentes Configurações.

$N_W$	$H_1$	$H_2$	MSE ( $10^5$ )	AIC
2	20	10	7,26	5,85
2	20	40	7,45	6,36
2	40	10	7,36	6,12
2	80	50	7,16	9,00
4	20	10	7,10	5,92
4	20	40	6,83	6,37
4	40	10	6,87	6,22
4	80	50	6,95	9,33
7	20	10	6,95	6,03
7	20	40	6,79	6,49
7	40	10	6,68	6,46
7	80	50	6,61	9,80
10	20	10	7,10	6,18
10	20	40	6,34	6,56
10	40	10	6,73	6,73
10	80	50	6,23	10,30
14	20	10	6,99	6,64
14	20	40	6,63	6,78
14	40	10	6,82	7,10
14	80	50	15,50	11,90

$N_L = 2$  o mesmo é 10,30, ou seja, aproximadamente 40,7% maior.

Logo, sendo o AIC o critério de desempate para seleção do modelo, tem-se que o processo de otimização considerado neste trabalho resultou na MLP exógena com a seguinte configuração  $(N_L, N_W, H_1) = (1, 10, 80)$ , esta configuração é exibida na Figura 4.10, em que o valor 30 exibido na entrada da MLP exógena corresponde ao produto do número de pontos da janela temporal (10) pelo número de coordenadas no vetor de entrada (3), ou seja  $3 \cdot 10$  ou de forma equivalente, 30 pontos.

O comportamento da série temporal do IBOVESPA, em conjunto com a série predita pelo modelo MLP exógeno e seu respectivo erro de predição, são exibidos na Figura 4.11.

Tabela 4.12: Critérios de Parada da MLP Exógena para Dados do IBOVESPA e  $N_L = 2$ .

$MSE_d (10^5)$	$MSE (10^5)$	AIC
7,23	28,70	14,10
1,45	33,20	11,90
0,15	14,16	11,10

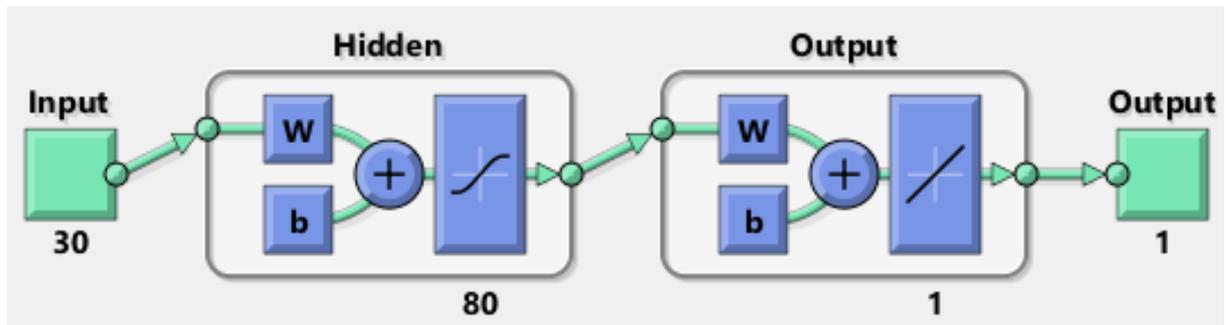


Figura 4.10: Configuração da MLP Exógena Otimizada para Dados do IBOVESPA.

Considerando a variância da série temporal de  $1,45 \cdot 10^8$  e o MSE otimizado de  $6,23 \cdot 10^5$ , tem-se que  $MSE_{norm}$  é igual a  $4,31 \cdot 10^{-3}$ . Logo, o MSE é 0,43% da variância da série temporal real, o que ilustra a capacidade do modelo MLP exógeno em explicar adequadamente os dados observados e desta forma reduzir a variabilidade do erro de predição.

## 4.4 Comparação entre os Modelos

Para o problema de predição do preço de fechamento do IBOVESPA considerado neste trabalho, a Tabela 4.13 sintetiza os resultados obtidos por cada modelo otimizado segundo os critérios do menores MSE e AIC.

Tabela 4.13: Desempenho dos Diferentes Modelos para Dados do IBOVESPA.

Modelo	$MSE (10^5)$	AIC	$MSE_{norm} (10^{-3})$
ARIMA	1210,00	10,70	840,00
MLP Simples	6,56	14,20	4,53
MLP Exógeno	6,23	7,32	4,31

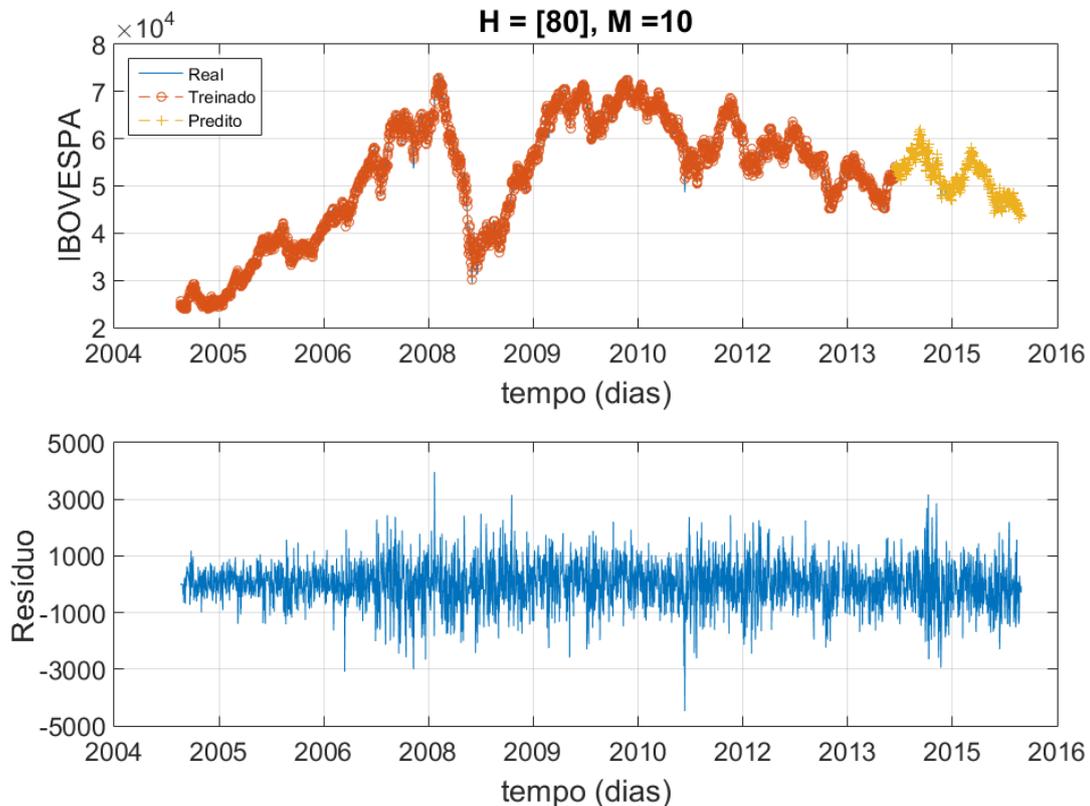


Figura 4.11: Comportamento da MLP Exógena Otimizada para Dados do IBOVESPA.

Mediante a análise da tabela, tem-se que o modelo ARIMA apresentou desempenho inferior aos modelos bio-inspirados que utilizam a técnica da ANN. Por exemplo, o MSE otimizado para o modelo ARIMA é de  $1,21 \cdot 10^8$ , enquanto que, para o modelo da MLP simples apresenta MSE igual a  $6,56 \cdot 10^5$ , indicando uma diferença relativa de 99,46%.

Enquanto que, com relação aos diferentes modelos baseados na ANN, partindo-se do MLP simples para sua versão exógena, com MSE iguais a  $6,56 \cdot 10^5$  e  $6,23 \cdot 10^5$ , respectivamente, ou seja, uma diferença relativa de, aproximadamente, 5,03%, mas tal ganho torna-se mais significativo ao considerar o AIC de cada modelo, sendo 14,0 para a MLP simples e 7,32 para a MLP exógena, que traduz-se na diferença relativa de, aproximadamente, 48,45%.

Portanto, a redução na variabilidade observada entre os modelos indica que existe correlação entre o preço de fechamento (variável alvo) e os preços mínimo e máximo (variável exógena) e a captura desta correlação é capaz de reduzir o erro de predição e

---

com isso, otimizar a adequabilidade do modelo em explicar os dados observados. Assim, segundo os critérios do MSE, AIC e  $MSE_{norm}$ , tem-se que a MLP exógena apresentou o melhor desempenho entre os modelos avaliados.

# Capítulo 5

## Conclusão

*If I have seen further it is by standing on the shoulders  
of giants.*

—Isaac Newton

Neste trabalho destacou-se a importância da análise e predição de séries temporais, em particular de séries financeiras (BRAGA; LUDERMIR, 2011). Verificou-se diferentes variações do modelo estatístico denominado ARIMA, bem como diferentes modelos baseados na ANN: MLP simples e MLP exógena. Para cada modelo, estipulou-se um sub-conjunto de seus parâmetros de operação como sendo fixos, enquanto o restante foi considerado livre e portanto, parte integrante do procedimento de otimização. Assim, a análise de cada modelo correspondeu a variação de cada parâmetro de otimização e a verificação de sua influência nos critérios de desempenho adotados neste trabalho: MSE, AIC e  $MSE_{norm}$ .

Conforme a análise dos resultados obtidos, determinou-se que o modelo MLP exógeno apresentou o melhor resultado segundo os critérios dos menores MSE, AIC globais e  $MSE_{norm}$ , em que MSE do modelo foi igual a  $6,23 \cdot 10^5$ , seu AIC correspondeu a 7,32 e seu  $MSE_{norm}$  de  $4,31 \cdot 10^{-3}$ . Desta forma, chega-se à conclusão de que, entre os modelos avaliados neste trabalho, a MLP exógena apresentou-se como modelo ótimo no sentido de resultar nos menores MSE, AIC e  $MSE_{norm}$  globais.

Por fim, ressalta-se a importância de entender o processo de representação de séries temporais segundo os modelos estudados e validar os algoritmos implementados com dados gerados mediante funções matemáticas simplificadas e adequadas ao problema em análise,

antes do modelo ser posto em prática frente aos dados reais de interesse. Assim, é possível antecipar a identificação de erros decorrentes de implementação errônea, bem como avaliar os parâmetros de maior relevância ao problema sendo analisado.

## 5.1 Trabalhos Futuros

Ao longo da análise dos modelos propostos neste trabalho, avaliou-se o problema de predição dos valores da série temporal representada pelo preço de fechamento do IBOVESPA, em que, inicialmente, considerou-se somente valores defasados da série temporal e em seguida, adicionou-se o efeito das séries temporais exógenas compostas pelos preços mínimo e máximo do IBOVESPA.

Com relação aos modelos baseados na ANN, neste trabalho focou-se na arquitetura MLP com e sem variável exógena, desta forma, pode-se sugerir como estudos futuros, a aplicação das demais arquiteturas elicitadas: RBF, GRNN, FTDNN, NAR e NARX.

Para o estudo dos modelos analisados, alguns parâmetros foram considerados como sendo fixos, enquanto outros foram selecionados como parte do procedimento de otimização. Desta forma, ressalta-se a oportunidade de considerar estes parâmetros tidos como fixos, no processo otimização, por exemplo, número máximo de validações, algoritmo de treinamento etc.

Além das variáveis acopladas nos modelos exógenos, sugere-se a inclusão de outros fatores que podem exercer influência explicativa na formação do preço de fechamento do IBOVESPA, por exemplo, as variáveis macroeconômicas: taxa juros básica da economia, índice de inflação, taxa de câmbio etc. A partir destas modificações, pode-se capturar novas correlações presentes nos dados e, desta forma, é plausível que se obtenha resultados superiores aos obtidos neste trabalho e assim reduzir o erro de predição do modelo (MONTGOMERY; JENNINGS, 2008).

Pode-se destacar também o potencial de utilizar os resultados preditos com os modelos avaliados neste trabalho em sistemas de apoio à tomada de decisão, comumente aplicados na negociação automática de ações no mercado financeiro (COPPIN, 2010). Destaca-se por exemplo, os sistemas que fazem uso da Lógica Fuzzy, tipicamente aplicada em tais sistemas devido a sua semelhança com o método de raciocínio aproximado desenvolvido no cérebro humano. À combinação das técnicas ANN e Lógica Fuzzy origina o modelo denominado

Neuro-Fuzzy (LUGER, 2004), em que utiliza-se a ANN para realizar predições, alimentando o sistema fuzzy responsável pelo processo de tomada de decisões, *e.g.* comprando ações em determinado ponto e as vendendo em outro momento oportuno.

# Referências Bibliográficas

ASSAF, A. *Mercado Financeiro*. 2. ed. [S.l.]: Editora Atlas, 1999.

BELMONT, D. F. de S. *Teoria de ondas de Elliott: Uma aplicação ao mercado de ações da bmfbovespa*. UFPB, 2010.

BMFBOVESPA. *Índice Bovespa (Ibovespa)*. sep 2016. Disponível em:  
<[https://www.bmfbovespa.com.br/pt\\_br/produtos/indices/indices\\_amplos/indice\\_bovespa\\_ibovespa](https://www.bmfbovespa.com.br/pt_br/produtos/indices/indices_amplos/indice_bovespa_ibovespa)>.

BOX, G. E. P. et al. *Time Series Analysis: Forecasting and Control*. 5. ed. [S.l.]: Wiley, 2015.

BRAGA, A. de P.; LUDERMIR, A. P. *Redes Neurais Artificiais - Teoria e Prática*. 2. ed. [S.l.]: LTC, 2011.

CETIP. *Calculadora da Taxa DI - Cetip*. sep 2016. Disponível em:  
<<https://www.cetip.com.br/renda-fixa/calculadora-di>>.

COPPIN, B. *Inteligência Artificial*. [S.l.]: Paulus, 2010.

FINANCE, G. *Dados IBOVESPA*. ago 2017. Disponível em:  
<[www.finance.google.com/finance/historicalstartdate=Jan+01%2C+2005&enddate=Oct+22%2C+](http://www.finance.google.com/finance/historicalstartdate=Jan+01%2C+2005&enddate=Oct+22%2C+)>

FORTUNA, E. *Mercado Financeiro Produtos e Serviços*. 4. ed. [S.l.]: Editora Qualitymark, 2005.

HAYKIN, S. O. *Neural Networks and Learning Machines*. 3. ed. [S.l.]: Pearson, 2008.

IBGE. *Sistema Nacional de Índices de Preços ao Consumidor*. sep 2016. Disponível em:  
<[www.ibge.gov.br/home/estatistica/indicadores/precos/inpcipca/defaultinpc](http://www.ibge.gov.br/home/estatistica/indicadores/precos/inpcipca/defaultinpc)>.

KAASTRA, I.; BOYD, M. Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, Elsevier, v. 10, p. 215 – 236, 1996.

KAY, S. M. *Fundamentals of Statistical Signal Processing*. [S.l.]: Prentice Hall, 1993. ISBN 0133457117.

KEYNES, J. *General Theory Of Employment , Interest And Money*. [S.l.]: Atlantic Publishers & Distributors (P) Limited, 2016. ISBN 9788126905911.

KHASHEI, M.; BIJARI, M. An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with applications*, Elsevier, v. 37, n. 1, p. 479–489, 2010.

KOVACS, Z. L. *Redes Neurais Artificiais - Fundamentos e Aplicações*. 4. ed. [S.l.]: Livraria da Física, 2006.

LUGER, G. F. *Inteligência artificial: estruturas e estratégias para a resolução de problemas complexos*. [S.l.]: Bookman, 2004.

MATHWORKS. *Econometrics Toolbox*. feb 2017. Disponível em:  
<<https://www.mathworks.com/products/econometrics.html>>.

MATHWORKS. *Neural Network Toolbox*. feb 2017. Disponível em:  
<<https://www.mathworks.com/products/neural-network.html>>.

MCCULLOCH, W.; PITTS, W. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 1943.

MONTGOMERY, D. C.; JENNINGS, C. L. *Introduction to Time Series Analysis and Forecasting*. 1. ed. [S.l.]: Wiley, 2008.

MORETTIN, P. A.; TOLOI, C. M. C. *Análise de Séries Temporais*. 2. ed. [S.l.]: Editora Edgard Blucher, 2006.

PALLADINO, B. *Na Onda das Ações*. 1. ed. [S.l.]: Saraiva, 2008.

PINHEIRO, J. L. *Mercado de Capitais: fundamentos e técnicas; Atlas*. 7. ed. [S.l.]: Atlas, 2014.

PINHO, D. B.; VASCONCELLOS, M. A. S. *Manual De Economia - Equipe de Professores da USP*. 6. ed. [S.l.]: Editora Edgard Blucher, 2011.

REZAEI, M. et al. River flow forecasting using artificial neural network (shoor ghaen). *Ciência e Natura*, Santa Maria, v. 37, n. 1, p. 207–215, 2015.

ROQUE, R. do C. *Estudo sobre a empregabilidade da previsão do índice BOVESPA usando Redes Neurais Artificiais*. Dissertao (Mestrado) — Universidade Federal do Rio de Janeiro, 2009.

ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Cornell Aeronautical Laboratory*, 1958.

SHAN, C. et al. *The Data Science Handbook: Advice and Insights from 25 Amazing Data Scientists*. [S.l.]: Data Science Bookshelf, 2015.

SPECHT, D. F. A general regression neural network. *IEEE Transactions on Neural Networks*, 1991.

WALMSLEY, J. *New Financial Instruments*. [S.l.]: Wiley, 1998. (Frontiers in Finance Series). ISBN 9780471121367.

ZHANG, G. P. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, Elsevier, v. 50, p. 159 – 175, 2003.