

ANDRÉ ALVES PORTELA SANTOS

**PREVISÃO NÃO-LINEAR DA TAXA DE CÂMBIO REAL/DÓLAR
UTILIZANDO REDES NEURAIIS E SISTEMAS NEBULOSOS**

FLORIANÓPOLIS

2005

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO SÓCIO-ECONÔMICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA
ÁREA DE CONCENTRAÇÃO: ECONOMIA E FINANÇAS

PREVISÃO NÃO-LINEAR DA TAXA DE CÂMBIO REAL/DÓLAR
UTILIZANDO REDES NEURAIS E SISTEMAS NEBULOSOS

Dissertação submetida ao Programa de Pós-Graduação em Economia da Universidade Federal de Santa Catarina como parte dos requisitos para obtenção do grau de Mestre em Economia

ANDRÉ ALVES PORTELA SANTOS

Florianópolis, março de 2005.

**PREVISÃO NÃO-LINEAR DA TAXA DE CÂMBIO REAL/DÓLAR
UTILIZANDO REDES NEURAIIS E SISTEMAS NEBULOSOS**

André Alves Portela Santos

‘Esta dissertação foi julgada adequada para obtenção do Título de Mestre em Economia, Área de Concentração em Economia e Finanças, e aprovada em sua forma final pelo Programa de Pós-Graduação em Economia da Universidade Federal de Santa Catarina.’

Celso Leonardo Weydmann
Coordenador do Programa de Pós-Graduação em Economia

Banca Examinadora:

Newton Carneiro Affonso da Costa Jr.
Orientador

Leandro dos Santos Coelho
Co-orientador

Eraldo Sérgio da Silva
Membro

AGRADECIMENTOS

A meus pais e meus irmãos, pela ajuda durante toda vida;

À família Natale Junkes, em especial à Patrícia, D. Rosa e Sr. Waldemiro, pelo amor e carinho fraternal que por mim sempre tiveram;

Ao meu co-orientador, Prof. Leandro dos Santos Coelho, pela imensurável ajuda em todo as etapas do trabalho;

Ao meu orientador, Prof. Newton Carneiro Affonso, pelo apoio e dedicação ao longo de todos esses anos;

Ao Prof. Eraldo Sérgio da Silva, por ter gentilmente disponibilizado as séries históricas que utilizei neste trabalho;

Aos amigos Marco “Kiko” Picanço, Marcus Emerson Verhine, Jales Dantas da Costa, Luis Gustavo Pinto de Carvalho pela ajuda que me deram na elaboração deste trabalho;

A todos os amigos que fiz no curso de mestrado e na graduação durante esses anos morando nesta linda e inesquecível cidade;

Ao contribuinte brasileiro, que proporcionou o financiamento do meu estudo e da minha formação em uma Universidade pública e gratuita.

RESUMO

SANTOS, A. Previsão não-linear da taxa de câmbio Real/Dólar utilizando redes neurais e sistemas nebulosos. 2005. 98 f. Dissertação (Mestrado em Economia), Universidade Federal de Santa Catarina, 2005.

Este trabalho tem como objetivo investigar a hipótese de que modelos não-lineares de previsão de séries temporais são capazes de fornecer uma previsão fora-da-amostra mais acurada que modelos lineares tradicionais. De uma perspectiva teórica, considera-se que modelos não-lineares são mais precisos dada sua maior capacidade em capturar não-linearidades presentes em séries financeiras. Estudos empíricos anteriores têm corroborado com essa hipótese, enquanto outros a têm questionado, tornando a questão controversa. Os modelos não-lineares analisados foram as redes neurais *perceptron* multicamadas (RN-MLP), as redes neurais função de base radial (RN-RBF) e os sistemas nebulosos (*fuzzy systems*) do tipo Takagi-Sugeno. Esses modelos possuem características adequadas para problemas envolvendo previsão de séries temporais, como a capacidade de realizar aprendizado através de treinamento específico acerca do ambiente ao qual está sendo utilizado, no caso das redes neurais, e a capacidade de lidar com incertezas qualitativas e quantitativas, no caso dos sistemas nebulosos. As estruturas matemáticas NARX e NARMAX (*Nonlinear Autoregressive Moving Average with eXogenous inputs*) foram utilizadas para a representação dinâmica não-linear das séries. Os modelos lineares analisados foram os modelos auto-regressivos (AR), modelos média móvel (MA), modelos auto-regressivos de média móvel (ARMA) e os modelos auto-regressivos de média móvel considerando heteroscedasticidade condicional auto-regressiva dos resíduos (ARMA-GARCH). Os procedimentos de estimação desses modelos, baseados na metodologia Box-Jenkins, estão bem sedimentados dentro da área de séries temporais.

Foram utilizadas séries de retorno da taxa de câmbio Real/Dólar (R\$/US\$) com frequência de 15 min., 60 min. e 120 min. durante o período que vai de 01/01/2002 a 01/01/2003 e séries com frequência diária e semanal durante o período que vai de 01/01/2000 a 01/01/2004. O desempenho preditivo dos modelos foi avaliado segundo o *root mean squared error* (RMSE), o índice de desigualdade U-Theil, o percentual de sinais corretamente previstos e a estatística de falha de previsão Pesaran-Timmermann. Além disso, foi avaliado o retorno e o risco de uma estratégia de negociação estabelecida com base nas previsões geradas pelos modelos. Os resultados indicam que os modelos não-lineares obtiveram um desempenho preditivo superior aos modelos lineares. Em todas as séries analisadas, os modelos não-lineares conseguiram significância estatística em suas previsões, além de um índice U-Theil inferior e um percentual de sinais corretamente previstos superior em relação aos modelos lineares. Constatou-se que a qualidade da previsão está intimamente relacionada com a frequência com que a série é observada. Quanto maior foi a frequência da série, maior foi a sua previsibilidade. Uma possível explicação para esse resultado está na presença de não-linearidade nas séries, encontrada em maior grau nas séries de alta frequência. Os resultados da estratégia de negociação indicaram que as estratégias baseadas nos modelos não-lineares superaram – em termos de risco e retorno – as estratégias baseadas em modelos lineares e a estratégia passiva de investimento. As estratégias não-lineares obtiveram retornos e índices de Sharpe superiores às demais estratégias. Nesse sentido, os resultados desse trabalho corroboram com evidências empíricas anteriores que atestam haver ganho em termos de desempenho preditivo decorrente da utilização de modelos não-lineares em relação aos modelos lineares.

Palavras-Chave: previsão de séries temporais, modelos lineares, modelos não-lineares.

ABSTRACT

SANTOS, A. Previsão não-linear da taxa de câmbio Real/Dólar utilizando redes neurais e sistemas nebulosos. 2005. 98 f. Dissertação (Mestrado em Economia), Universidade Federal de Santa Catarina, 2005.

This study investigates the hypothesis that nonlinear models for time series forecasting are able to provide a more accurate out-of-sample forecast than traditional linear models. From a theoretical perspective, nonlinear models are considered more precise, due to their greater capacity to capture nonlinear features present in financial series. Several previous empirical studies have confirmed this hypothesis while others have questioned such advantages, creating controversy concerning the lack of academic consensus. The nonlinear models of multilayer *perceptron* and radial basis function neural networks, and the Takagi-Sugeno fuzzy system were evaluated. Those models hold appropriate characteristics for time series forecasting problems, like the ability to perform learning through specific training process, in the case of the neural networks, and the ability of modeling qualitative and quantitative uncertainties, in the case of the fuzzy systems. The NARX and NARMAX mathematical structures were used for the nonlinear dynamic representation of the series. The linear models evaluated were the autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), and autoregressive moving average with autoregressive conditional heteroskedasticity (ARMA-GARCH) models. The estimation procedures of these models, based on Box-Jenkins methodology, are well-established in time series literature.

Series of 15min., 60 min. and 120 min. returns from 01/01/2002 to 01/01/2003 and daily and weekly returns from 01/01/2000 to 01/01/2004 of the Brazilian exchange rate (R\$/US\$) were used. The predictive performance was evaluated using the following criteria: root mean squared error (RMSE), U-Theil inequality index, percentage of corrected predicted signals and the Pesaran-Timmermann predictive failure statistic. The return and risk of a trade strategy based on model's forecasts were also evaluated. Results indicated that nonlinear models perform better than their linear counterparts. In all series, nonlinear models achieved statistic significance in its forecasts, smaller U-Theil indexes and higher percentage of corrected predicted signals than linear models. The forecast performance is also closely related to the series' frequency: the higher the series' frequency, the higher the forecastability. A possible explanation for this result it is the presence of nonlinearity in the series, which is in a higher level in high frequency series. The results of the trade strategy indicated that strategies based on nonlinear models surpassed – in terms of return and risk – strategies based on linear models and a passive trade strategy. Nonlinear strategies achieved higher returns and higher Sharpe indexes than other strategies. In this sense, the results of this study corroborate with previous empirical evidence that states a gain, in terms of predictive performance, due to the use of nonlinear models.

Key-words: time series forecasting, linear models, nonlinear models.

SUMÁRIO

LISTA DE TABELAS	IX
LISTA DE QUADROS	XI
LISTA DE FIGURAS	XII
LISTA DE SIGLAS E ACRÔNIMOS	XIII
1 PROBLEMÁTICA DE PESQUISA	1
1.1 INTRODUÇÃO	1
1.2 FORMULAÇÃO DO PROBLEMA DE PESQUISA	6
1.3 OBJETIVOS	10
1.3.1 <i>Objetivo geral</i>	10
1.3.2 <i>Objetivos específicos</i>	11
2 MODELOS ECONOMÉTRICOS DE PREVISÃO DE SÉRIES TEMPORAIS 12	
2.1 MODELOS LINEARES	14
2.1.1 <i>Modelos AR, modelos MA e modelos ARMA</i>	14
2.1.2 <i>Modelos ARMA-GARCH</i>	16
2.2 MODELOS NÃO-LINEARES	17
2.2.1 <i>Redes neurais artificiais</i>	18
2.2.1.1 <i>Redes Neurais perceptron de uma ou mais camadas</i>	22
2.2.1.2 <i>Redes Neurais Função de Base Radial</i>	26
2.2.1.3 <i>Redes neurais em previsão de séries temporais</i>	31
2.2.1.4 <i>Aplicações de redes neurais em Economia e Finanças</i>	34
2.2.2 <i>Sistemas nebulosos (fuzzy systems)</i>	37
2.2.2.1 <i>Estrutura e classificação do sistemas nebulosos</i>	40
2.2.2.2 <i>Sistema nebuloso de TS</i>	42
2.3 SÍNTESE CONCLUSIVA.....	47
3 FUNDAMENTOS PRÁTICOS DA PREVISÃO DE SÉRIES TEMPORAIS..... 49	
3.1 INTRODUÇÃO	49
3.2 DETECÇÃO DE NÃO-LINEARIDADES	50
3.3 ESCOLHA DA ESTRUTURA DOS MODELOS	51
3.5 AVALIAÇÃO DO DESEMPENHO PREDITIVO DOS MODELOS.....	52
3.6 SÍNTESE CONCLUSIVA.....	54
4 ANÁLISE EMPÍRICA	55
4.1 SÉRIES TEMPORAIS UTILIZADAS E TESTES DE ESPECIFICAÇÃO	55

4.2 ESTIMAÇÃO DOS MODELOS	58
4.2.1 Modelos lineares	60
4.2.2 Modelos não-lineares	60
4.3 APRESENTAÇÃO DOS RESULTADOS	61
4.3.1 Estratégia de negociação.....	68
4.4 SÍNTESE CONCLUSIVA.....	73
5 CONCLUSÕES.....	75
6 REFERÊNCIAS.....	78

LISTA DE TABELAS

Tabela 4.1 Estatística descritiva para as séries de retornos da taxa de câmbio Real/Dólar, de acordo com a frequência observada.	57
Tabela 4.2 Teste BDS para presença de não-linearidade nas séries de retorno da taxa de câmbio Real/Dólar.	58
Tabela 4.3 Distribuição do número de observações de cada série histórica de acordo com as etapas de estimação e validação fora-da-amostra.	59
Tabela 4.4 Resultados do teste de previsão de Chow obtidos a partir da estimativa de um modelo AR(1) para as séries de retorno da taxa de câmbio Real/Dólar.	60
Tabela 4.5 Resultados dos modelos lineares e não-lineares de previsão para a série da taxa de câmbio Real/Dólar com frequência de 15 min.	62
Tabela 4.6 Resultados dos modelos lineares e não-lineares de previsão para a série da taxa de câmbio Real/Dólar com frequência de 60 min.	63
Tabela 4.7 Resultados dos modelos lineares e não-lineares de previsão para a série da taxa de câmbio Real/Dólar com frequência de 120 min.	64
Tabela 4.8 Resultados dos modelos lineares e não-lineares de previsão para a série da taxa de câmbio Real/Dólar com frequência diária.	65
Tabela 4.9 Resultados dos modelos lineares e não-lineares de previsão para a série da taxa de câmbio Real/Dólar com frequência semanal.	66
Tabela 4.10 Resumo dos resultados dos melhores modelos lineares de previsão para a série de retornos da taxa de câmbio Real/Dólar de acordo com a frequência observada.	67
Tabela 4.11 Rentabilidade acumulada, desvio padrão e índice de Sharpe das estratégias de negociação aplicadas à série de retornos da taxa de câmbio Real/Dólar com frequência de 15 min.	69
Tabela 4.12 Rentabilidade acumulada, desvio padrão e índice de Sharpe das estratégias de negociação aplicadas à série de retornos da taxa de câmbio Real/Dólar com frequência de 60 min.	70
Tabela 4.13 Rentabilidade acumulada, desvio padrão e índice de Sharpe das estratégias de negociação aplicadas à série de retornos da taxa de câmbio Real/Dólar com frequência de 120 min.	71
Tabela 4.14 Rentabilidade acumulada, desvio padrão e índice de Sharpe das estratégias de negociação aplicadas à série de retornos da taxa de câmbio Real/Dólar com frequência diária.	71

Tabela 4.15 Rentabilidade acumulada, desvio padrão e índice de Sharpe das estratégias de negociação aplicadas à série de retornos da taxa de câmbio Real/Dólar com frequência semanal.....	72
--	----

LISTA DE QUADROS

Quadro 2.1 Tipos principais de funções de ativação	20
Quadro 2.2 Tipos principais de funções de base radial.....	28

LISTA DE FIGURAS

Figura 2.1 Modelo genérico de um neurônio artificial.	19
Figura 2.2 Modelos genéricos de uma RN <i>feedforward</i> (a) e uma RN <i>feedback</i> (b).....	22
Figura 2.3 RN <i>perceptron</i> multicamadas	24
Figura 2.4 Ilustração da distância euclidiana entre dois vetores.	27
Figura 2.5 RN função de base radial.....	29
Figura 2.6 Rede neural sem camada intermediária.....	32
Figura 2.7 Representação clássica (a) e <i>fuzzy</i> (b) da altura de determinada população..	39
Figura 2.8 Modelo nebuloso de Takagi-Sugeno visto como uma aproximação linear local de uma função não-linear (adaptada de Babuska, 2000).....	44
Figura 2.9 <i>Clusters</i> nebulosos em formato elipsoidal (adaptada de Babuska, 2000).....	46
Figura 4.1 Gráficos da série da taxa de câmbio Real/Dólar em nível e dos retornos da taxa de câmbio para a série de 15 min. e para a série semanal.	56

LISTA DE SIGLAS E ACRÔNIMOS

AR	- auto-regressivo.....	10
ARCH	- heteroscedasticidade condicional auto-regressiva.....	6
ARMA	- auto-regressivo de média móvel.....	10
ARMA-GARCH	- modelo ARMA considerando heteroscedasticidade condicional auto-regressiva dos resíduos.....	10
GARCH	-heteroscedasticidade condicional auto-regressiva generalizada..	6
LN	- lógica nebulosa.....	38
MA	- média móvel.....	10
NARX	- <i>Nonlinear autoregressive with exogenous inputs</i>	30
NARMAX	- <i>Nonlinear autoregressive moving average with exogenous inputs</i>	30
RN	- rede neural.....	17
RN-MLP	- rede neural <i>perceptron</i> multicamadas.....	21
RN-MLP-NARX	- rede neural <i>perceptron</i> multicamadas com representação matemática NARX.....	61
RN-MLP-NARMAX	- rede neural <i>perceptron</i> multicamadas com representação matemática NARMAX.....	61
RN-RBF	- rede neural função de base radial.....	25
RN-RBF-NARX	- rede neural função de base radial com representação matemática NARX.....	61
RN-RBF-NARMAX	- rede neural função de base radial com representação matemática NARMAX.....	61
SN	- sistema nebuloso.....	39
TS	- Takagi-Sugeno.....	8
SN-TS	- sistema nebuloso Takagi-Sugeno.....	60
SN-TS-NARX	- sistema nebuloso Takagi-Sugeno com representação matemática NARX.....	61
SN-TS-NARMAX	- sistema nebuloso Takagi-Sugeno com representação matemática NARMAX.....	61

1

Problemática de Pesquisa

Este capítulo apresenta a motivação deste trabalho de pesquisa, analisando quais são as questões-chave relativas ao estudo. A seção 1.1 discute a questão da previsibilidade dos retornos dos ativos e a questão da presença de não-linearidades nas séries financeiras. A seção 1.2 apresenta a formulação do problema de pesquisa, enquanto que a seção 1.3 comenta os objetivos propostos neste trabalho.

1.1 Introdução

O debate acerca da possibilidade de previsão do desempenho do mercado dominou, e ainda domina, a mente de muitos economistas que se engajam na tentativa de provar que o mercado acionário é completamente imprevisível ou então na formulação de modelos que buscam algum indício de previsibilidade, ou seja, provar que é possível, com base naquilo que aconteceu no passado, estimar o que vai ocorrer no futuro.

De acordo com a Hipótese dos Mercados Eficientes (Fama, 1970; 1991), o preço das ações possui uma dinâmica aleatória, ou seja, não segue nenhum tipo de ordem ou padrão. Essa questão é fundamental para o debate relacionado à previsibilidade do mercado, pois caso a evolução dos preços realmente não apresente nenhuma sistemática, ou seja, caso o mercado seja eficiente na sua formação de preço, é impossível prevê-lo. O conceito de mercado eficiente é usualmente associado ao conceito de *martingale* ou jogo justo (Campbell *et al.* 1997), cuja condição é dada por

$$E[P_{t+1} / P_t, P_{t-1}, \dots] = P_t \quad (1.1)$$

ou, equivalentemente,

$$E[P_{t+1} - P_t / P_t, P_{t-1}, \dots] = 0. \quad (1.2)$$

onde E é definido como operador de valor esperado.

Tomando P_t como o preço de um ativo na data t , a hipótese *martingale* assegura que o preço de amanhã é esperado ser igual ao preço de hoje, dada toda a história dos preços do ativo. Alternativamente, a expectativa de variação no preço é zero, quando condicionada à história dos preços; dessa forma o preço pode tanto cair quanto subir. Vista de uma perspectiva de previsão, a hipótese *martingale* implica que a “melhor” previsão do preço de amanhã é simplesmente o preço de hoje. Outro aspecto da hipótese é que as mudanças de preços são não-correlacionadas em todas as defasagens, implicando a ineficácia de todas as regras lineares de previsão das mudanças futuras de preço baseadas apenas no histórico dos preços.

A hipótese *martingale* levou ao desenvolvimento de uma hipótese correlata que se tornou fundamental nos estudos sobre eficiência dos mercados: a hipótese do caminho aleatório (*random walk hypothesis*). A hipótese (ou modelo) do caminho aleatório pode ser representada por

$$P_t = \mu + P_{t-1} + \varepsilon_t. \quad (1.3)$$

Este modelo nos diz que o preço do ativo no período $t+1$ é dado pelo seu preço no momento imediatamente anterior, mais um termo de mudança esperada μ conhecido como *drift*, mais um componente de erro não-previsível. O comportamento do termo ε_t é extremamente importante, e as restrições acerca do comportamento deste termo motivaram Campbell *et al.* (1997) a produzirem três versões do modelo de caminho aleatório:

Caminho aleatório 1: Nesta versão, também chamada de versão “forte” do modelo *random walk*, assume-se que o termo ε_t é identicamente distribuído e independente. A independência implica que não só os incrementos são não-correlacionados, mas também qualquer função não-linear dos incrementos será não-correlacionada.

Caminho aleatório 2: Esta versão considera que os incrementos são independentes mas não necessariamente identicamente distribuídos. O relaxamento dessa hipótese mantém a característica da não-previsibilidade linear além de admitir a presença de fatos estilizados em séries financeiras, como mudanças na volatilidade condicional.

Caminho aleatório 3: A versão mais geral – e mais testada empiricamente – do modelo de caminho aleatório substitui a hipótese de incrementos independentes pela hipótese de incrementos dependentes porém não-correlacionados. Isso significa, por exemplo, satisfazer a condição de que a $Cov[\varepsilon_t, \varepsilon_{t-k}] = 0$, para todo $k \neq 0$, porém a $Cov[\varepsilon_t^2, \varepsilon_{t-k}^2] \neq 0$, para todo $k \neq 0$. Tal processo indica que os incrementos são não-correlacionados, mas são não-independentes, desde que o quadrado dos incrementos é correlacionado.

De fato, a pesquisa empírica sugeriu durante muito tempo a impossibilidade de se refutar o modelo de caminho aleatório nas variações de preço dos ativos financeiros. Entretanto, como notam Campbell *et al.* (1997), o avanço computacional e o surgimento de metodologias econométricas robustas, passaram a sugerir a existência algum nível de previsibilidade nos retornos dos ativos financeiros, indicando que as variações de preço não podiam ser totalmente explicadas pelo modelo do caminho aleatório ou *random walk*.

Assim como nos trabalhos de Lo e Mackinlay (1988) e posteriormente Campbell *et al.* (1997), indícios crescentes de dependência entre as sucessivas mudanças de preço passaram agora a sugerir que pode ser possível identificar algum tipo de tendência nas séries de preço dos ativos, e que estratégias sofisticadas de negociação que visem identificar essas tendências podem superar as estratégias simplistas do tipo *buy-and-hold*. As evidências empíricas utilizando outras abordagens têm sugerido que as variáveis relacionadas aos fundamentos das empresas, tais como a relação preço/lucro, valor contábil/valor de mercado dentre outras, atuam como previsores dos retornos das ações (Fama e French, 1992), dando origem às chamadas estratégias contrárias de investimento (Lakonishok *et al.*, 1994).

Xu (2004) argumenta que algum nível de previsibilidade dos retornos, mesmo que pequeno, pode alterar significativamente o preço dos ativos, afetando os critérios de decisão de alocação do investidor como também sua aversão ao risco. Xu mostra que pequenos níveis de previsibilidade são suficientes para fazer com que uma estratégia de negociação gere ganhos significantes¹. Segundo o autor, a previsibilidade de curto prazo dos retornos (entendida como correlação positiva observada) pode ser atribuída a efeitos presentes na microestrutura do mercado, como negócios não-síncronos e o *spread* entre compra e venda enquanto que a previsibilidade de longo prazo pode ser atribuída a

¹ O autor estima modelos econométricos de previsão que possuem um baixo poder explicativo (coeficiente de correlação múltipla R^2 em torno de 2%) e, utilizando dados reais e dados obtidos por simulação *bootstrap*, conclui que estratégias de negociação baseadas nesses modelos geram retornos ajustados ao risco superiores a uma estratégia *buy-and-hold*.

variações no prêmio pelo risco. Em períodos intermediários, anúncios não-sincronizados de ganhos das empresas podem ter um importante papel na previsibilidade dos retornos.

Uma parcela significativa dos trabalhos relacionados à previsibilidade dos retornos tomou um novo rumo a partir do início da década de 90 graças à introdução de conceitos advindos da Teoria do Caos², tais como *dependência sensível das condições iniciais*, *sistemas dinâmicos não-lineares*, *atratores* e *bifurcações*. Esses conceitos, inicialmente aplicados a diversos outros sistemas da natureza, principalmente na física, tornaram-se novos candidatos à explicação da dinâmica presente nas séries financeiras (Aguirre e Aguirre, 1997) a qual, assim como diversos sistemas da natureza, passou a ser considerada *não-linear*.

A partir do final da década de 80, estudos relatavam que uma determinada série temporal financeira, quando representada utilizando coordenadas apropriadas, o resultado era não mais uma série temporal, mas uma figura geométrica semelhante aos *fractais* e *atratores* (Aguirre e Aguirre, 1997). Esse resultado indicava que uma série temporal financeira, apesar de caracterizada por constantes flutuações, permanece oscilando dentro de coordenadas limitadas no espaço de fase. Os sistemas com dinâmica caótica passaram a ser tratados não como sistemas em que impera uma aleatoriedade pura, e sim como sistemas dinâmicos em que é possível identificar, através de determinada representação, um padrão em sua dinâmica³.

Atualmente, pesquisadores têm relatado a presença de outras importantes propriedades da física como leis de potência, periodicidades e memória longa em dados financeiros, dando origem a um campo específico de conhecimento dentro da economia e da física conhecido como *econofísica* (Da Silva *et al.*, 2005; Gleria *et al.*, 2002). Assim como outros sistemas da natureza, a economia é tratada como um sistema complexo onde suas subunidades interagem entre si de modo não-linear, possibilitando identificar e representar geometricamente as propriedades globais estáveis do sistema.

O advento da Teoria do Caos à Ciência Econômica estimulou diversos pesquisadores a desenvolverem uma fundamentação matemática dos conceitos da teoria. Dentro da macroeconomia, por exemplo, demonstrou-se que modelos dinâmicos podem exibir uma natureza extremamente complexa sem a necessidade de choques exógenos,

² Uma explicação detalhada sobre a Teoria do Caos, sua história e seus principais conceitos, bem como aplicações à Ciência Econômica, é apresentada em Gleisser (2001) e Gleick (1989).

³ É relevante fazer uma distinção entre sistema não-linear e sistema caótico. Este último é um caso particular de um sistema não-linear; nem todo sistema não-linear deve, necessariamente, exibir uma dinâmica considerada caótica. A estimativa do expoente de Lyapunov é um procedimento recorrente para determinar se determinada série é caótica ou não.

com diversos pontos de equilíbrio ou não-convergência a um estado estacionário (Scheinkman, 1990; Boldrin e Deneckere, 1989).

Do ponto de vista da econometria, uma série de testes estatísticos paramétricos e não-paramétricos surgiram no intuito de verificar a presença de não-linearidades nas séries financeiras. Em outras palavras, o objetivo era identificar a presença de estruturas complexas inerentes ao mecanismo gerador de dados da série, responsáveis por produzir de modo endógeno (sem a necessidade de choques) a oscilação e a volatilidade inerentes às séries financeiras (Brock *et al.* 1987).

As evidências empíricas passaram a indicar presença significativa de não-linearidades em séries de diversos tipos de ativos, como câmbio, ações, ouro, contratos futuros etc, dando suporte aos estudos que associavam a Teoria do Caos ao mercado financeiro, como apresentado em Abhyankar *et al.* (1995). Os autores destacam, entretanto, que nos sistemas não-lineares que apresentam dependência sensível das condições iniciais, o conhecimento do mecanismo gerador dos preços pode ser insuficiente para garantir o lucro, já que a acurácia da previsão pode degenerar rapidamente até dar tempo para que investidores executem o negócio, indicando que esses conceitos são de crítica importância para a Hipótese dos Mercados Eficientes.

Essas constatações trouxeram conseqüências importantes ao uso dos tradicionais modelos econométricos lineares de previsão de séries financeiras. Como notam Aguirre e Aguirre (1997), se, de fato, as séries financeiras são caracterizadas pela presença de não-linearidades, pode-se concluir que (i) modelos lineares determinísticos são incapazes de captar as incertezas presentes na série observada e (ii) esses modelos não podem reproduzir, de modo sistemático, as oscilações presentes na série. A utilização desses modelos passa a ser, então, inapropriada. Dessa forma, a atenção voltou-se para formulação de modelos não-lineares de previsão de séries temporais⁴, os quais, de uma perspectiva teórica, seriam mais aptos a capturarem não-linearidades presentes nas séries financeiras, e, portanto, fornecerem uma previsão acurada.

Segundo Clements *et al.* (2004) a racionalidade da utilização de modelos não-lineares deriva do fato de que modelos lineares possuem um desempenho inferior na captura de certos tipos de comportamentos econômicos. Nesse sentido, a existência de um ou mais regimes do tipo recessão e expansão, no caso das variáveis macroeconômicas, ou alta e baixa volatilidade, no caso das variáveis financeiras, justificaria a utilização de modelos não-lineares. Segundo os autores, a medida que as

⁴ Um conceito formal de sistema não-linear e modelo não-linear de previsão é apresentado na seção 2.1 do capítulo 2.

metodologias de estimação, seleção e validação de modelos não-lineares avançarem, é de se esperar uma melhora significativa em seu desempenho preditivo.

Um ponto importante a ser salientado refere-se à utilização de séries temporais financeiras de alta frequência⁵ nos estudos envolvendo previsão dos retornos. Dada sua recente disponibilidade para estudos acadêmicos, séries de alta frequência têm recebido destaque na literatura especializada em previsão de séries temporais. Em termos gerais, ainda são poucos os estudos que utilizam essa modalidade de série temporal na estimação e validação dos modelos de previsão. Deve-se, entretanto, ressaltar que a disponibilidade de séries de alta frequência traz algumas vantagens importantes, que devem ser exploradas. Do ponto de vista teórico, a grande quantidade de observações disponíveis torna os testes de presença de não-linearidades nas séries mais robustos, enquanto que do ponto de vista prático torna-se possível o estabelecimento de estratégias de negociação intra-dia.

1.2 Formulação do problema de pesquisa

A literatura relacionada à previsão de séries temporais financeiras tem registrado, desde a década de 90, importantes avanços relacionados à incorporação de novas metodologias que tentam determinar padrões de relacionamentos presentes nos dados do mercado financeiro. Essas metodologias, em sua maioria computacionalmente intensivas, caracterizam-se pela capacidade de identificação e previsão de sistemas dinâmicos não-lineares, ou seja, sistemas em que as variáveis do ambiente possuem um complexo padrão de inter-relacionamento que se altera ao longo do tempo.

Esse aspecto tem estimulado pesquisadores de diversas formações a aplicarem novas abordagens de identificação de sistemas em vários problemas da área de Finanças, como por exemplo, precificação de opções (Freitas e Souza, 2002), previsão de falência de empresas (Malhotra e Malhotra, 2002) e, principalmente, previsão de séries temporais financeiras (Rubio *et al.*, 2002; Tsaih *et al.*, 1998). Essas abordagens não-lineares têm procurado respaldo teórico no crescente relato de presença de não-linearidades em séries financeiras, o que de alguma forma justificaria a sua utilização.

Nesse sentido, novas abordagens têm sido propostas no intuito de capturarem não-linearidades nas séries temporais financeiras. Importantes avanços em análise não-linear de séries temporais como os processos estocásticos não-lineares ARCH e GARCH

⁵ Série formada por observações coletadas dentro do dia, com uma determinada frequência, por exemplo, 1 min., 5 min, 1 hora.

(Engle, 1982; Bollerslev, 1986), modelos determinísticos não-lineares com dinâmica caótica (Scheinkman, 1990), análise não-paramétrica (Diebold e Nason, 1990), redes neurais (Haykin, 2001; Granger, 1995), modelos auto-regressivos não-lineares (Ocal, 2000) e com transição suave (van Dijk e Terasvirta, 2000; Eitrheim e Terasvirta, 1996), modelos Markov de mudança de regime (Hamilton, 1994) além de modelos de previsão ainda pouco difundidos entre os econométristas, como os sistemas nebulosos e algoritmos genéticos (Coelho e Santos, 2003) têm reacendido o interesse acerca da não-linearidades em dados financeiros, abrindo novas possibilidades para previsão de retorno dos ativos.

De um ponto de vista estritamente teórico, a utilização de modelos não-lineares na previsão de séries financeiras aparenta ser perfeitamente justificável, dada imensa gama de trabalhos empíricos atestando a presença de não-linearidades nesse tipo de série (ver, por exemplo, LeBaron, 1994; Aguirre e Aguirre, 1997; Abhyankar *et al.*, 1995). Entretanto, a afirmação de que esses modelos produzem previsões mais acuradas que modelos lineares tradicionais não pode ser dita com tanta veemência. Dijk e Franses (2003) colocam que, em muitos trabalhos, apesar da substancial evidência de não-linearidade em dados dentro-da-amostra, o uso de modelos não-lineares raramente traduz-se em previsões mais acuradas para dados fora-da-amostra. Em um dos trabalhos mais referenciados na literatura, Stock e Watson (1999) analisam um conjunto de 49 modelos lineares e não-lineares, concluindo que os modelos lineares geralmente “vencem a competição” pela melhor previsão. Resultados semelhantes são encontrados por Swanson e White (1995) e Diebold e Nason (1990).

Várias razões para esse tipo de resultado têm sido sugeridas na literatura. Ramsey (1996) e Dijk e Franses (2003) apontam algumas delas, como por exemplo: i) a não-linearidade detectada em séries temporais pode ser espúria, no sentido de que outras características presentes em séries temporais como heteroscedasticidade, quebras estruturais ou presença de *outliers* podem sugerir a presença de não-linearidade; ii) modelos não-lineares produzem previsões pontuais inferiores aos modelos lineares, entretanto costumam apresentar intervalos de previsão mais precisos que seus concorrentes. Dijk e Franses (2003) contra-argumentam, afirmando que o aparente desempenho inferior dos modelos não-lineares pode estar relacionado a um critério impróprio usado na avaliação das previsões obtidas por tais modelos.

Abhyankar *et al.* (1995), utilizando uma série com 60.000 observações de alta frequência do índice da Bolsa de Valores do Reino Unido (FTSE-100), identificam a presença de dependência não-linear em todas as frequências (principalmente de 1 min. e

5 min.) entretanto parte da não-linearidade encontrada na série pode ser explicada por um fato estilizado na literatura em finanças: a variância condicional dos resíduos (processo ARCH e suas variantes). Em outras palavras, a não-linearidade está em parte “mascarada” por um processo de heteroscedasticidade condicional auto-regressiva. Bollerslev *et al.* (1992) salientam que, uma vez removidos os efeitos GARCH de uma série temporal, testes usuais de não-linearidade tendem a aceitar (ou rejeitar a um baixo intervalo de confiança) a presença de linearidade na série.

O problema do critério de avaliação do desempenho preditivo dos modelos é analisado em Leitch e Tanner (1991). Realizando testes empíricos de previsão das taxas de juros norte-americanas, os autores concluem que critérios de avaliação convencionais baseados na mensuração do tamanho do erro de previsão não apresentam relação sistemática com lucros. Segundo Leitch e Tanner (1991, pg. 580-581)

“É naturalmente melhor analisar diretamente os lucros do que examinar uma variável que está, no máximo, indiretamente relacionada aos lucros. Economistas geralmente assumem que empresas usam previsões porque elas auxiliam os lucros. Dessa forma, um teste de acurácia de previsão mais apropriado é a lucratividade, e não o tamanho do erro de previsão ou o seu valor ao quadrado”.

Dessa forma, nos estudos direcionados à previsão de retornos, torna-se interessante o estabelecimento de uma estratégia de negociação baseada nos resultados previstos pelo(s) modelo(s) utilizado(s) e a avaliação dos ganhos obtidos em comparação a uma estratégia simplista de negociação do tipo *buy-and-hold*.

Ramsey (1996) salienta que os modelos não-lineares são capazes de fornecer um melhor ajustamento à série temporal, entretanto “um bom ajustamento não necessariamente leva a uma boa previsão”. A causa disso reside no fato de que um bom ajustamento significa que o modelo consegue fazer aproximações locais do sistema, enquanto que uma boa previsão requer um conhecimento de propriedades globais. Desta forma, os modelos não-lineares estão sujeitos a grandes erros (pontuais) de previsão, em função da sua sobre-especialização ou sobre-ajustamento em relação à série temporal, ou ainda devido a sua grande sensibilidade em relação a valores (condições) iniciais.

Vale destacar que muitos estudos como, por exemplo, Andrada-Felix *et al.* (2003), Franses e Griensven (1998), Kuan e Liu (1995), Leung *et al.* (2000), Dasgupta *et al.* (1994) e Olson e Mossman (2003) a modelagem não-linear de redes neurais foi superior

aos modelos lineares na previsão das séries econômicas e financeiras. De forma semelhante, os trabalhos de Ocal (2000), utilizando modelos do tipo STAR (*Smooth Transition Auto-Regressive*), Medeiros *et al.* (2002) e Boero e Marrocu (2002) utilizando modelos SETAR (*Self Exciting Threshold Auto-Regressive*), obtiveram erros de previsão inferiores aos modelos auto-regressivos lineares tradicionais. Esses resultados tornam controverso o debate acerca dos benefícios da utilização de modelos não-lineares de previsão de séries temporais.

Duas abordagens relevantes, entre tantas outras, para a previsão do comportamento de sistemas complexos em finanças são os sistemas nebulosos e as redes neurais. Nos últimos anos, pesquisadores (Tseng e Tzeng, 2002; Von Altrock, 2002; León *et al.*, 2002) têm proposto um variado espectro de metodologias de identificação e previsão não-linear baseadas em sistemas nebulosos, a citar os modelos nebulosos do tipo Takagi-Sugeno (TS) para lidarem com sistemas não-lineares, “mal definidos” e que apresentam incertezas.

Derivada inicialmente da teoria clássica dos conjuntos, a qual utiliza como princípio básico a idéia de que um elemento “pertence” ou “não-pertence” a determinado conjunto, a teoria dos conjuntos nebulosos foi desenvolvida a partir de 1965 por Lotfi Zadeh para tratar do aspecto vago da informação (Sandri e Correa, 1999). O *princípio da incompatibilidade*, determinado por Zadeh, afirma que “quando a complexidade de um sistema aumenta, nossa capacidade de fazer afirmações precisas e significativas sobre ele diminui até um limiar no qual a precisão e a relevância tornam-se características mutuamente exclusivas” (Cox, 1994). Posteriormente, surgiram os denominados sistemas nebulosos, que são modelos não-lineares de identificação e previsão de sistemas que utilizam conceitos da teoria dos conjuntos nebulosos.

As redes neurais, por sua vez, têm recebido considerável atenção nos últimos anos em função de suas habilidades em realizar aprendizagem, sendo aplicadas em um grande número de situações. Em previsão de séries temporais, diversas metodologias utilizando redes neurais têm sido propostas na literatura (Narendra e Parthasarathy, 1990; Aguirre, 2004) e resultados interessantes em problemas da área de Finanças têm surgido (Wong e Selvi, 1998).

Dito isso, uma questão ainda controversa colocada aos pesquisadores da área de previsão de séries temporais refere-se ao ganho na acurácia de previsão decorrente da utilização de modelos não-lineares em contraponto aos modelos lineares tradicionais. Até que ponto a rejeição da hipótese nula de presença de linearidade em séries financeiras justifica a utilização de modelos de previsão não-lineares? Mais importante:

será que a utilização de modelos não-lineares, em sua maioria de maior complexidade, gera uma melhor previsão do que os modelos lineares tradicionais a ponto de superar a sua utilização?

Nesse sentido, esse estudo visa, em parte, replicar trabalhos anteriores que compararam a capacidade preditiva de modelos lineares e não-lineares, como, por exemplo, os estudos de Stock e Watson (1999) e Dijk e Franses (2003). Pretende-se fornecer uma contribuição com relação à metodologia e aos critérios de avaliação da previsão, além de discutir em que medida a possível presença de não-linearidade nas séries pode influenciar na sua previsibilidade.

1.3 Objetivos

1.3.1 Objetivo geral

O objetivo geral deste trabalho é verificar se modelos não-lineares de previsão de séries temporais são capazes de gerar uma previsão fora-da-amostra mais precisa que modelos lineares tradicionais. Os modelos não-lineares a serem analisados serão as redes neurais *perceptron* multi-camadas e as redes neurais do tipo função de base radial e os sistemas nebulosos Takagi-Sugeno, enquanto que os modelos lineares serão os modelos auto regressivos (AR), modelos média móvel (MA), modelos auto regressivos de média móvel (ARMA) e modelos auto regressivos de média móvel considerando heteroscedasticidade condicional auto regressiva dos resíduos (ARMA-GARCH).

Serão utilizadas séries de retorno da taxa de câmbio Real/Dólar coletadas nas frequências de 15 min., 60 min., 120 min., diária e semanal. Nas séries de alta frequência (15 min., 60 min., e 120 min.) o período observado vai de 01/01/2002 a 01/01/2003. Nas séries com frequência diária e semanal, o período observado vai de 01/01/2000 a 01/01/2004.

A idéia subjacente ao estabelecimento desse objetivo geral é tentar fornecer mais uma contribuição teórica e empírica à questão relativa ao uso de modelos não-lineares em detrimento de modelos lineares. Enquanto que parcela dos estudos empíricos mostra que modelos não-lineares conseguem uma previsão fora-da-amostra mais acurada que modelos lineares (Kaashoek e van Dijk 2002; Andrada-Félix *et al.*, 2003; Kuan e Liu, 1995), outra parcela dos estudos chega à conclusão oposta (Stock e Watson, 1999; Neely e Weller, 2003; Swanson e White, 1995), tornando a questão controversa. Dessa forma, esse estudo pretende contribuir para elucidação dessa questão.

1.3.2 Objetivos específicos

Os objetivos específicos deste trabalho de pesquisa visam:

- i) Verificar se a previsibilidade está relacionada com a frequência com que a série é observada;
- ii) Verificar através de testes estatísticos a presença de não-linearidades nas séries analisadas;
- iii) Estabelecer critérios de avaliação da capacidade dos modelos em realizar previsões um-passo-a-frente, de modo que se possa escolher, dentre cada categoria, qual modelo obteve melhor desempenho;
- iv) Estabelecer uma estratégia de negociação baseada nas previsões geradas pelos modelos analisados, de modo a verificar se estratégias baseadas em modelos não-lineares são capazes de gerar um retorno ajustado ao risco superior às estratégias baseadas em modelos lineares; e
- v) Estabelecer uma relação entre significância estatística e significância econômica das previsões obtidas.

2

Modelos econométricos de previsão de séries temporais

Neste capítulo serão detalhados os modelos univariados de previsão de séries temporais utilizados no estudo. Inicialmente, uma definição formal de modelo econométrico linear e não-linear será fornecida. Em seguida, serão definidos os modelos lineares estimados no trabalho e, na seqüência, os modelos não-lineares de redes neurais e os sistemas nebulosos do tipo Takagi-Sugeno.

De acordo com Intriligator (1983), os modelos econométricos podem ser lineares ou não-lineares *nas variáveis* e/ou *nos parâmetros*. Por exemplo, um modelo de regressão do tipo $Y = \beta_1 + \beta_2 X_1^2 + \varepsilon$, onde β_1 e β_2 são parâmetros e ε é o termo de erro aleatório, pode ser considerado como um modelo não-linear nas variáveis, porém linear nos parâmetros. Já o modelo $Y = \beta_1 + \sqrt{\beta_2} X_1 + \varepsilon$ é um modelo não-linear nos parâmetros e linear nas variáveis.

Intriligator (1983) nota que os primeiros modelos econométricos formulados são lineares, e que a hipótese de linearidade concentra-se exclusivamente nos parâmetros. Para o autor a premissa da linearidade desempenhou papel importante ao permitir a demonstração de teoremas matemáticos e estatísticos fundamentais para modelos econométricos, incluindo estimação de parâmetros e utilização de modelos de análise estrutural, previsão e avaliação de políticas. Entretanto, o uso de modelos econométricos não-lineares nos parâmetros tem se tornado comum graças a evolução da computação e da análise numérica, que facilitaram a estimação desses modelos.

Os modelos lineares e não-lineares podem ainda ser classificados como paramétricos e não-paramétricos. Nos modelos paramétricos a forma de relacionamento funcional entre as variáveis é conhecida e os parâmetros desses modelos são passíveis de serem estimados por algum critério, como máxima verossimilhança. Nos modelos não-paramétricos, existe a ausência (completa e ou quase completa) de conhecimento a

priori a respeito da forma da função que está sendo estimada. Dessa forma, mesmo que a função seja estimada a partir do ajuste de parâmetros livres, o conjunto de formas que a função pode assumir (classe de funções que o modelo pode prever) é muito amplo.

Segundo Campbell *et al.* (1997), em análise linear de séries temporais, a seqüência de choques ε_t a qual está sujeita a série temporal observada x_t é assumida ser não-correlacionada com ela mesma e com a própria série x_t , mas não necessariamente ser independente e identicamente distribuída (IID). Já em análise não-linear de séries temporais, a seqüência de choques ε_t é tipicamente assumida ser IID, mas busca-se uma possível função não-linear relacionando-a à série x_t . Uma representação geral é dada por⁶

$$x_t = f(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) \quad (2.1)$$

onde os choques são assumidos ter média zero e variância unitária, e $f(\cdot)$ é uma função desconhecida. Campbell *et al.* (1997) notam que a generalidade desta representação a torna difícil de se trabalhar; em geral, a maioria dos modelos costuma recair sobre uma classe mais restrita, que pode ser descrita por

$$x_t = g(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots) + \varepsilon_t h(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots). \quad (2.2)$$

A função $g(\cdot)$ representa a média de x_t condicional à informação passada, ou seja, $E_{t-1}[x_t] = g(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$. A inovação em x_t corresponde ao choque ε_t , e o coeficiente de proporcionalidade é representado pela função $h(\cdot)$. O quadrado dessa função é a variância de x_t condicional à informação passada, ou seja, $E_{t-1}[(x_t - E_{t-1}[x_t])^2] = h(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots)^2$. Modelos com uma função $g(\cdot)$ não-linear são ditos não-lineares *em média*, enquanto que modelos com uma função $h(\cdot)$ ² não-linear são ditos não-lineares *em variância*.

⁶ Essa representação faz uso do Teorema de Representação de Wold, o qual assegura que qualquer série temporal pode ser representada como um processo média móvel dos incrementos, os quais representam a variância incondicional e a autocovariância da série.

2.1 Modelos lineares

2.1.1 Modelos AR, modelos MA e modelos ARMA

Seguindo a definição proposta por Box e Jenkins (1976), Nelson (1973) e Pindyck e Rubinfeld (1991), pode-se definir um processo auto-regressivo com p defasagens AR(p) como

$$y_t = \phi_0 + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t. \quad (2.3)$$

ou

$$\Phi_p(L)y_t, \text{ onde } \Phi_p(L) = (1 - \phi_1 L - \dots - \phi_p L^p) \quad (2.4)$$

onde $\{\varepsilon_t\}$ é um ruído branco normalmente distribuído com média zero e variância constante σ^2 . Os termos $\phi_0, \phi_1, \dots, \phi_p$ são parâmetros e L é o operador de defasagem definido por $L^p y_t = y_{t-p}$. Por sua vez, um processo média móvel com q defasagens MA(q) pode ser definido como

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2.5)$$

ou

$$y_t = \mu + \Theta_q(L)\varepsilon_t, \text{ onde } \Theta_q(L) = (1 + \theta_1 L + \dots + \theta_q L^q) \quad (2.6)$$

onde $\theta_1, \dots, \theta_q$ são parâmetros e L segue a definição anterior. Generalizando, um processo auto-regressivo de média móvel, ARMA(p, q), pode ser escrito como

$$y_t = \phi_0 + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (2.7)$$

ou

$$\Phi_p(L)y_t = \phi_0 + \Theta_q(L)\varepsilon_t \quad (2.8)$$

A estimativa dos parâmetros dos modelos ARMA é realizada da seguinte forma⁷. Seja um vetor β que representa os parâmetros p e q (ϕ e θ) a serem estimados e w o vetor de variáveis y e ε que compõem os termos do modelo ARMA. O procedimento consiste em escolher valores numéricos de β que minimize

$$S(\beta) = \sum_{t=1}^T [\varepsilon_t | w, \beta]^2 = \sum_{t=1}^T [\varepsilon_t]^2, \quad (2.9)$$

⁷ Os procedimentos de estimação dos parâmetros dos modelos ARMA são discutidos em Pindyck e Rubinfeld (1991).

onde $[\varepsilon_t]$ denota os erros condicionais aos valores de w e β . Expandindo $[\varepsilon_t]$ em uma série de Taylor em torno de valores iniciais de β_0 , temos

$$\begin{aligned} [\varepsilon_t] &= [\varepsilon_t | w, \beta_0] + \sum_{i=1}^T (\beta_i - \beta_{i,0}) \frac{\partial [\varepsilon_t]}{\partial \beta_i} \Big|_{\beta=\beta_0} \\ &+ \frac{1}{2} \sum_{i=1}^{p+q} (\beta_i - \beta_0)^2 \frac{\partial^2 [\varepsilon_t]}{\partial \beta_i^2} \Big|_{\beta=\beta_0} + \dots + \end{aligned} \quad (2.10)$$

onde $\beta_{i,0}$ é o valor do primeiro candidato para o parâmetro β , e dessa forma o i -ésimo componente do vetor β_0 . $[\varepsilon_t]$ pode ser aproximado através dos primeiros dois termos da expansão de Taylor. Tem-se então

$$x_{i,t} = - \frac{\partial [\varepsilon_t]}{\partial \beta_i} \Big|_{\beta=\beta_0} \quad (2.11)$$

e

$$[\varepsilon_{t,0}] = [\varepsilon_t | w, \beta_0]. \quad (2.12)$$

Segue-se então que, aproximadamente,

$$[\varepsilon_t] = [\varepsilon_{t,0}] - \sum_{i=1}^{p+q} (\beta_i - \beta_{i,0}) x_{i,t}, \quad (2.13)$$

que pode ser reescrito como

$$[\varepsilon_{t,0}] + \sum_{i=1}^{p+q} \beta_{i,0} x_{i,t} = \sum_{i=1}^{p+q} \beta_i x_{i,t} + [\varepsilon_t]. \quad (2.14)$$

Os parâmetros β_i da equação 2.14 podem ser estimados através de uma regressão linear. Através do método dos mínimos quadrados ordinários (MQO) é feita uma nova estimativa $\hat{\beta}$ de β . Em seguida, usando uma nova expansão de Taylor de $[\varepsilon_t]$ em torno dos valores de $\hat{\beta}_1$, é obtida um nova versão da equação 2.14, que também pode ser estimada através de MQO. Esse processo é repetido sucessivamente até que

$$\hat{\beta}_k - \hat{\beta}_{k-1} \approx 0, \quad (2.14)$$

onde k é o número de iterações requeridas até que a convergência ocorra. $\hat{\beta}_k$ será então a estimativa final dos parâmetros ϕ_1, \dots, ϕ_p e $\theta_1, \dots, \theta_q$.

2.1.2 Modelos ARMA-GARCH

A maioria das séries financeiras não apresenta média constante ao longo do tempo, além de exibir fases de relativa tranqüilidade sendo seguidas por fases de alta volatilidade. Modelos ARCH e GARCH são especialmente concebidos para modelagem e previsão de variância não-constante. De acordo com Fang e Xu (2003), modelos autorregressivos lineares de média móvel considerando heteroscedasticidade condicional (modelos ARMA-GARCH) tornam a variância dos resíduos previsível, capturando fatos estilizados do segundo momento dos retornos.

Um processo ARMA(p,q)-GARCH(p,q) pode ser definido como um processo ARMA(p,q) do tipo

$$\Phi_p(L)y_t = \phi_0 + \Theta_q(L)\varepsilon_t \quad (2.15)$$

onde o processo gerador de $\{\varepsilon_t\}$ é dado por

$$\varepsilon_t = v_t \sqrt{h_t}, \quad v_t \sim \text{IID}(0,1) \quad (2.16)$$

e

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i}. \quad (2.17)$$

Como $\{v_t\}$ é um processo ruído branco independente das realizações passadas de ε_{t-i} e a variância condicional de ε_t é dada por $E_{t-1} \varepsilon_t^2 = h_t$, podemos concluir que a variância condicional de ε_t em (2.17) é dada por h_t .

A estimação dos parâmetros de um modelo GARCH é usualmente feita através da maximização de uma função *log-likelihood* (Enders, 1995). Supondo que a variância condicional de ε_t seja dada por

$$\begin{aligned}
 h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \\
 &= \alpha_0 + \alpha_1 (y_{t-1} - \beta x_{t-1})^2
 \end{aligned}
 \tag{2.18}$$

é possível encontrar os valores dos parâmetros α_0 , α_1 e β através da maximização da função

$$\log \varphi = -(T/2) \ln(2\pi) - (1/2) \sum_{t=1}^T \ln h_t - (1/2) \sum_{t=1}^T h_t (y_t - \beta x_t)^2.
 \tag{2.19}$$

A maioria dos pacotes econométricos tradicionais é capaz de selecionar valores para os parâmetros que maximizem a função $\log \varphi$.

2.2 Modelos não-lineares

Uma representação geral de um modelo não-linear é dada por:

$$y = \varphi(x, \delta) + \varepsilon
 \tag{2.12}$$

onde x e δ são variáveis explicativas, ε é o componente de distúrbio estocástico e φ é uma função que associa variáveis dependentes à variável y . De acordo com Terasvirta *et al.* (1999), a partir dessa definição geral, os modelos não-lineares se subdividem em:

- i. *Modelos auto-regressivos não-lineares*, envolvendo apenas funções das variáveis dependentes (modelos univariados). Tipicamente, apenas funções matemáticas simples têm sido consideradas (ex. funções seno, co-seno, logística etc);
- ii. *Modelos de função de transferência não-lineares*, usando funções da variável dependente defasada e funções de variáveis explicativas defasadas e correntes;
- iii. *Modelos bilineares*, dados pela relação $y_t = \sum_{j,k} \beta_{j,k} y_{t-j} \varepsilon_{t-k}$. Quando acrescidos do produto de uma variável explicativa x_t e um termo residual defasado, tem-se os *modelos bilineares multivariados*;
- iv. *Modelos média móvel não-lineares*, formados pela soma de funções dos resíduos defasados $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-j}$;
- v. *Modelos duplamente estocásticos*, formados pelos produtos cruzados a variável dependente defasada e as variáveis explicativas correntes e defasadas.

Outras classes importantes de modelos não-lineares são os modelos *NARMAX* polinomial e racional, modelos de Wiener, *wavelets*, Hammerstein e Volterra, apresentados em Aguirre (2004) e Coelho (2000).

Neste estudo, a ênfase será dada aos modelos univariados auto-regressivos não-lineares. Uma representação geral desses modelos é dada por

$$y_t = \sum_{j=1}^p \alpha_j \varphi_j(y_{t-j}, \dots, y_{t-p}) + \varepsilon_t. \quad (2.13)$$

onde $\varphi(\cdot)$ é um conjunto de funções que associa as variáveis defasadas à variável corrente. Terasvirta *et al.* (1999) observam que quando a função $\varphi(\cdot)$ é a função logística $\varphi(x) = [1 + \exp(-x)]^{-1}$ tem-se o modelo de redes neurais, que será detalhado adiante.

2.2.1 Redes neurais artificiais

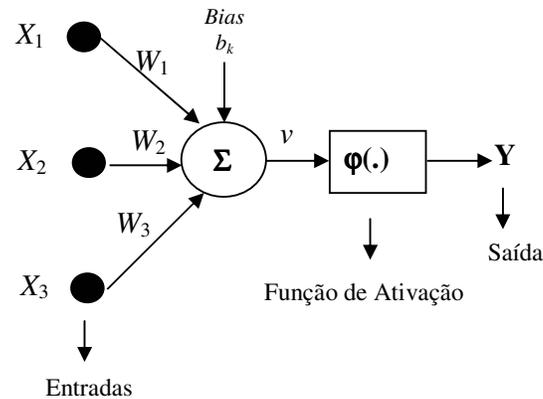
Uma rede neural artificial, ou simplesmente rede neural (RN), pode ser definida como uma estrutura computacional projetada para mimetizar de forma rudimentar a maneira pela qual o cérebro humano desempenha uma particular tarefa do seu interesse (Haykin, 2001). Devido a sua natureza indutiva, as RNs podem sobrepor a formulação teórica tradicional e inferir relações não-lineares complexas entre as observações de uma série temporal. Desta forma, arquiteturas de RNs têm recebido considerável atenção nos últimos anos em função de suas habilidades em realizar aprendizagem, sendo aplicadas em um grande número de situações e demonstrando desempenho superior às abordagens convencionais.

A principal característica de uma RN é a capacidade de aprender, através de um determinado tipo de treinamento, campos específicos de conhecimento. Portugal e Fernandes (1995) notam que as RNs são bem aplicadas em situações onde poucas decisões têm que ser tomadas a partir de uma grande quantidade de dados e situações onde um complexo mapeamento não-linear deve ser aprendido.

Como o cérebro humano, as RNs são constituídas por um conjunto de unidades de processamento conectadas entre si denominadas de neurônios. A figura 2.1 apresenta um modelo genérico de neurônio artificial proposto inicialmente por McCulloch e Pitts (1942), em que X_1 , X_2 e X_3 correspondem aos valores de entrada e Y equivale ao valor de saída do neurônio. Após a multiplicação dos valores de entrada pelos respectivos

pesos W_1 , W_2 e W_3 os produtos obtidos são somados, resultando no potencial de ativação, representado por v . Posteriormente, o valor de v é submetido a uma função de ativação, cujo resultado Y é o valor de saída do neurônio. A entrada da função de ativação pode ser aumentada ou diminuída por um elemento externo chamado de *bias* ou polarização

Figura 2.1 Modelo genérico de um neurônio artificial.



Matematicamente, pode-se escrever:

$$v = \sum_{i=1}^n w_i \cdot x_i \quad \text{e} \quad Y = \varphi(v) + b_k \Rightarrow Y = \varphi\left(\sum_{i=1}^n w_i \cdot x_i\right) + b_k \quad (2.14)$$

A função de ativação, representada na equação acima por $\varphi(\cdot)$, tem como objetivo limitar, dentro de uma escala de valores, a amplitude do sinal de saída. Tipicamente, o intervalo normalizado da amplitude da saída de um neurônio é escrito como o intervalo fechado $[-v, +v]$. O quadro 2.1 a seguir traz os principais tipos de funções de ativação mencionadas na literatura. A função de ativação mais comum na construção de RNs é a *função sigmóide*, que é definida como uma função estritamente crescente que exibe um balanceamento adequado entre comportamento linear e não-linear (Haykin, 2001).

Quadro 2.1 Tipos principais de funções de ativação.

Tipo de Função	Mapeamento Entrada-Saída
Limiar	$\varphi(v) = \begin{cases} 1 & \text{se } v \geq 0 \\ 0 & \text{se } v < 0 \end{cases}$
Linear	$\varphi(v) = v$
Saturação	$\varphi(v) = \begin{cases} +1 & \text{se } v > 1 \\ v & \text{se } v \in [-1, 1] \\ -1 & \text{se } v < -1 \end{cases}$
Sigmóide*	$\varphi(v) = \frac{1}{1 + \exp(-av)}$
Tangente hiperbólica	$\varphi(v) = \frac{e^v - e^{-v}}{e^v + e^{-v}}$

* O parâmetro a indica a declividade da função sigmóide.

A habilidade de uma RN em realizar previsões sobre uma determinada série temporal ou reconhecer padrões em um conjunto de dados provém de sua capacidade de aprendizado sobre o ambiente no qual está inserida. O problema do aprendizado consiste em encontrar, através de um processo iterativo (onde cada entrada provoca uma resposta) e iterativo (reiterado), um conjunto de parâmetros livres que possibilite à rede o desempenho desejado. Em outras palavras, deve-se encontrar o ajuste dos pesos sinápticos W_i e do nível de *bias* que produzam o menor nível de erro entre a resposta desejada e a resposta estimada pela rede (no caso de uma série temporal) ou a melhor fronteira de separação (decisão) entre padrões presentes em um conjunto de dados.

As formas básicas de aprendizagem realizada por uma RN são:

- i. *Aprendizado supervisionado*: para cada padrão de entrada fornecido pela rede deve-se fornecer também a saída esperada até que uma ligação adequada entre os dados de entrada e os dados de saída seja encontrada. A RN é exposta a um vetor de treino extraído do ambiente e um supervisor externo fornece a resposta desejada para este vetor específico, que é o resultado ótimo que a rede deveria apresentar;
- ii. *Aprendizado não-supervisionado*: somente os padrões de entrada são fornecidos, ficando a cargo do sistema descobrir sozinho padrões específicos da população de entrada. Através de identificação de regularidades estatísticas nos dados de entrada, a rede desenvolve a habilidade adaptar os pesos de suas conexões para

representar e agrupar os padrões de entrada de acordo com características similares (de Castro e De Castro, 2001);

- iii. *Aprendizado híbrido*: partes das conexões da RN são ajustadas de modo não-supervisionado enquanto que outra parte é ajustada de modo supervisionado. Coelho (2000) nota que as RNs do tipo função de base radial admitem esse tipo de aprendizagem; e
- iv. *Aprendizado por reforço*: segundo Braga *et al.* (2000) o aprendizado por reforço pode ser visto como um caso particular do aprendizado supervisionado. Enquanto que no aprendizado supervisionado a medida de desempenho da rede baseia-se num nível mínimo de erro pré-estabelecido, no aprendizado por reforço a única informação de treinamento fornecida à rede é se determinada saída está correta ou não, ou seja, não é fornecida à rede a resposta correta para o padrão de entrada.

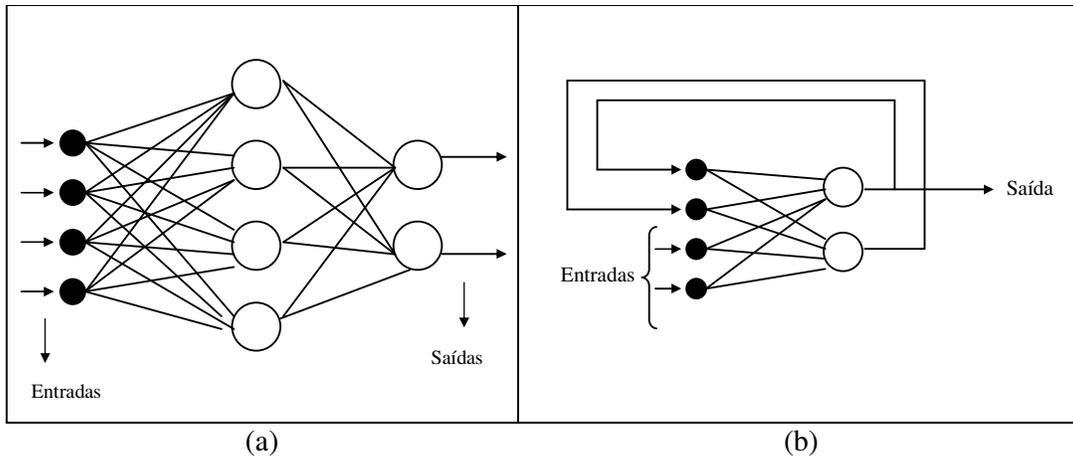
As RNs são, de modo geral, estruturadas em camadas, sendo que cada camada apresenta uma ou mais unidades de processamento (neurônios). Toda rede tem ao menos uma camada de entrada e uma camada de saída. A camada de entrada recebe os sinais de ambiente externo e a camada de saída guarda o resultado final da rede. A RN pode apresentar ainda mais de uma camada intermediária, o que permite que a saída da rede seja gerada por uma função contínua e não-linear das entradas. O número de camadas da rede, o número de neurônios em uma camada, o tipo de conexão entre os neurônios e a topologia da rede definem a arquitetura de uma RN. Com relação ao tipo de conexão entre os neurônios, as RN recebem a seguinte classificação:

- i. *Acíclica (feedforward)*: a saída do neurônio da i -ésima camada da rede não pode ser usada como entrada de neurônios em camadas de índice menor ou igual a i . Podem ser de uma única camada (*single layer*) ou multicamadas (*multilayer*).
- ii. *Cíclica (feedback)*: a saída de um neurônio na i -ésima camada da rede é usada como entrada de neurônios em camadas de índice menor ou a i .

A figura 2.2 apresenta os modelos genéricos de uma rede *feedforward* e uma rede *feedback*. As RNs *feedforward* mais utilizadas são redes *perceptron* de uma camada, *perceptron* multicamadas, redes função de base radial, redes B-splines, *wavelets* e redes CMAC. As principais redes do tipo *feedback* são as redes de Elman e Jordan, as redes

competitivas, as redes de Hopfield e os mapas auto-organizáveis de Kohonen⁸. Neste trabalho, será utilizada uma RN *feedforward* do tipo função de base radial, por ser o modelo de RN mais adequado ao problema que está sendo abordado. Na seção seguinte, será detalhada a estrutura das principais redes *feedforward*, além da justificativa de escolha da RN função base radial em detrimento de outras topologias.

Figura 2.2 Modelos genéricos de uma RN *feedforward* (a) e uma RN *feedback* (b).



2.2.1.1 Redes Neurais *perceptron* de uma ou mais camadas

O *perceptron*, proposto inicialmente por Frank Rosenblatt em 1958, consiste na forma mais simples de uma RN utilizada para classificação de padrões linearmente separáveis, ou seja, que estão em lados opostos de um hiperplano. É representado basicamente por um neurônio com pesos sinápticos ajustáveis e um nível de *bias*.

A saída de uma rede neural *perceptron* multicamadas (RN-MLP) com uma camada oculta e um único neurônio na camada de saída é dada por

$$Y = \varphi_s \left\{ \sum_{i=1}^m w_i \varphi_i \left(\sum_{j=1}^n w_{ij} \varphi_j + b_j \right) + b_s \right\} \quad (2.15)$$

sendo que φ_s é a função de ativação do neurônio da camada de saída. Tal função não precisa ser igual a φ_i , $i = 1, \dots, m$ que, por sua vez, não precisam ser iguais entre si. b_s é o termo de polarização do neurônio da camada de saída, w_i são os pesos da saída de cada neurônio da camada oculta e w_{ij} são os pesos da entrada j , vista pelo i -ésimo neurônio

⁸ Consultar o trabalho de Haykin (2001) para o detalhamento da estrutura das RNs *feedback*.

da camada oculta. Aguirre (2004) observa que em aplicações de modelagem de séries temporais, é comum escolher $b_s=0$.

Uma RN-MLP é uma rede de aprendizado supervisionado constituída por uma camada de entrada (*input layer*), uma ou mais camadas ocultas (*hidden layers*) e uma camada de saída (*output layer*). Com exceção da camada de entrada, todas camadas são constituídas de neurônios. Dessa forma, o projeto de uma RN-MLP envolve basicamente três aspectos: i) a determinação do número de camadas ocultas; ii) a determinação do número de neurônios em cada uma das camadas ocultas e iii) a especificação dos pesos sinápticos que interconectam os neurônios nas diferentes camadas da rede. A função das camadas ocultas é tentar identificar, através de treinamento supervisionado, algum padrão complexo de relacionamento entre os dados de entrada e saída. O seu alto grau de conectividade (sinapses) possibilita à rede aprender tarefas complexas, mesmo aquelas em que um difícil mapeamento não-linear entre dados de entrada e saída devem ser aprendidos. Dessa forma, acrescentando-se um número maior de camadas ocultas, a rede torna-se mais apta a identificar um comportamento complexo entre dados de entrada e saída.

O número de neurônios na(s) camada(s) oculta(s) determinará a capacidade da rede em generalizar, para novos dados apresentados, o seu aprendizado. Guerra e Coelho (2002) sugerem que a medida que eleva-se o número de neurônios presentes na(s) camada(s) oculta(s), a rede torna-se mais especializada, entretanto sua capacidade de generalização diminui. Em outras palavras, a rede torna-se muito especializada nos dados de uma amostra específica, perdendo a capacidade de fazer inferências corretas sobre novos dados ou padrões apresentados.

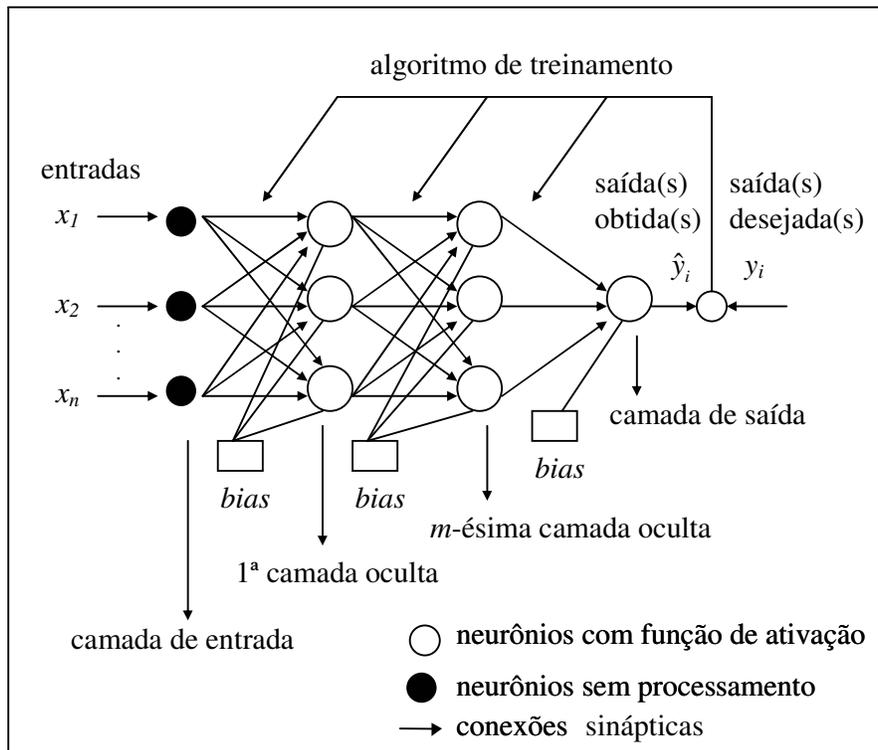
Entretanto, deve-se destacar algumas limitações e desvantagens relacionadas à utilização de RN-MLP. Em primeiro lugar, segundo De Castro e De Castro (2001) não há, infelizmente, regras determinadas para a especificação dos aspectos (i) e (ii). Além disso, a presença de não-linearidade distribuída ao longo da rede e sua alta conectividade tornam difícil a análise teórica das RN-MLP. O aprendizado da rede sobre o ambiente, o qual está representado pelos valores assumidos pelos pesos sinápticos, não pode ser visualizado na estrutura da rede. Em outras palavras, o alto padrão de conectividade impede que se extraia a(s) regra(s), ou o vetor de pesos utilizado, que permitiu(ram) à rede chegar a determinado resultado. Essa restrição faz com que muitas vezes as RNs sejam classificadas como modelos “caixa-preta” (*black box*).

O aspecto (iii) está intimamente relacionado ao algoritmo de treinamento supervisionado utilizado para o ajustamento dos pesos sinápticos da RN. De Castro e De Castro (2001), Coelho (1997), Krose e van der Smagt (1996) afirmam que o algoritmo de treino mais utilizado na literatura que trata de RN-MLP é o algoritmo de retropropagação do erro, ou *backpropagation* como é mais conhecido. A essência desse algoritmo consiste em primeiramente estimar o erro (dado pela diferença entre a saída estimada pela RN e a saída desejada) e em seguida retropropagar da camada de saída para as camadas intermediárias da rede efetuando-se um ajuste nos pesos sinápticos. Esse procedimento é repetido até que seja encontrado um valor de tolerância do erro imposta pelo usuário.

As etapas que regem o algoritmo de retropropagação do erro são sintetizadas por (Coelho, 2000):

- i. Iniciar aleatoriamente os pesos das conexões da RN;
- ii. Apresentar à rede o conjunto de treinamento constituído das entradas e das saídas desejadas;
- iii. Calcular a(s) saída(s) da RN, propagando as saídas dos neurônios de cada camada da RN para os neurônios da próxima camada, passando por suas respectivas funções de ativação;

Figura 2.3 RN *perceptron* multicamadas



- iv. Especificar as saídas desejadas e calcular os erros das camadas;

- v. Ajustar os pesos da RN começando pelas unidades de saída e propagando-se em direção à primeira camada oculta, utilizando as equações⁹:

$$w_{j,i}(n+1) = w_{j,i}(n) + \Delta w_{j,i}(n) \quad (2.16)$$

$$\Delta w_{j,i} = \eta \cdot \delta_j(n) \cdot y_i(n) \quad (2.17)$$

- onde $w_{j,i}$ é o peso a ser atualizado, $\Delta w_{j,i}$ é a mudança a ser realizada no peso da i -ésima camada para a j -ésima camada, η é a taxa ou velocidade de aprendizagem do algoritmo, $y_i(n)$ é a saída apresentada pela rede no instante n e $\delta_j(n)$ é o gradiente local definido por $\delta_j(n) = e_j(n)\phi'_j(v_j(n))$ onde $e_j(n)$ é a diferença (erro) entre a saída desejada e a saída estimada no instante n e $\phi'_j(v_j(n))$ é a derivada primeira da função de saída do j -ésimo neurônio da camada de saída.
- vi. Repetir os passos (iv) e (v) até que o erro obtido apresente um valor de tolerância aceitável (critério de parada de treinamento);
- vii. Utilizar um conjunto de teste para analisar o desempenho do treinamento.

Guerra e Coelho (2002) e Krose e van der Smagt (1996) apontam algumas deficiências inerentes ao algoritmo de retropropagação do erro. A primeira seria a baixa velocidade de aprendizado e, conseqüentemente, o alto custo computacional do algoritmo. A segunda deficiência seria a possibilidade de convergência para mínimos locais, ou seja, o treinamento da rede pode ser interrompido sem que seja atingido o mínimo global da função erro. Isso pode ocorrer quando a função erro apresenta um comportamento muito complexo durante a fase de treinamento da rede.

Uma alternativa ao algoritmo de retropropagação do erro, o qual utiliza somente a primeira derivada (a curvatura) da superfície, seria utilizar a segunda derivada (taxa de mudança na curvatura), reduzindo significativamente o tempo de treinamento. Esse procedimento é utilizado pelo algoritmo Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1943) e constitui-se numa alternativa viável ao tradicional algoritmo de retropropagação do erro (Guerra e Coelho, 2002).

Outro ponto que deve ser destacado acerca da fase de aprendizagem da RN é a possibilidade de ocorrência do problema de *overfitting*, ou seja, a rede especializa-se nos padrões de treinamento e perde sua capacidade de generalização (Freitas e Souza, 2003; Braga *et al.*, 2000). Uma das técnicas mais utilizadas para contornar esse

⁹ A derivação das equações do algoritmo de retropropagação do erro não serão demonstradas nesse trabalho. Consultar o trabalho de Haykin (2000, pg. 188-191) para uma demonstração detalhada dessas equações.

problema é a técnica de *early stopping*, que consiste em treinar a rede com uma determinada amostra de dados, denominada de *grupo de treinamento*, e validar o desempenho da rede utilizando outra amostra de dados, denominado *grupo de validação*. Se os resultados obtidos com a validação são satisfatórios, o treinamento é interrompido. Existe ainda a proposta da criação de um *grupo de treinamento*, *grupo de teste* e outro *grupo de validação* (Ivakhnenko e Ivakhnenko, 1995).

2.2.1.2 Redes Neurais Função de Base Radial

As RNs do tipo função de base radial (*radial basis function*) (RN-RBF) compõem uma classe de RNs cuja camada oculta é definida por um conjunto de funções de base radial, das quais a rede deriva seu nome. As unidades ocultas fornecem um conjunto de funções não-lineares que constituem uma base arbitrária para os padrões (vetores) de entrada, quando eles são expandidos sobre o espaço oculto. Essas funções são chamadas de funções de base radial.

Segundo Orr (1996) a principal característica de uma função de base radial é que sua resposta decresce (ou cresce) monotonicamente a medida que a distância de um ponto central aumenta (ou diminui). Cada função de base radial é centrada em uma particular coordenada do espaço multi-dimensional dos pontos que compõem o espaço dos dados de entrada da RN. Cada uma destas coordenadas particulares caracteriza-se por definir o centro de uma – entre várias possíveis – região de maior aglomeração de pontos do espaço dos dados de entrada.

As RN-RBF foram originalmente concebidas para problemas de interpolação de dados, ou seja, problemas em que dado um conjunto de vetores $\{u_i\}$ e um conjunto de escalares $\{y_i\}$, busca-se uma função $F(\cdot)$ tal que $y_i = F(u_i)$, $\forall i$. Uma possível solução para o mapeamento analítico é escolher $F(u)$ tal que,

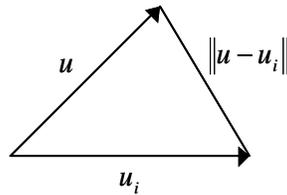
$$F(u) = w_0 + \sum_{i=0}^N w_i \varphi(\|u - u_i\|) \quad (2.18)$$

onde u é um determinado dado de entrada, u_i um ponto de referência, w_i são os coeficientes e $\{\varphi(\|u - u_i\|) / i = 1, 2, \dots, N\}$ é um conjunto de N funções de base radial. O operador $\|\cdot\|$ é a norma euclidiana, que mede a distância euclidiana da ponta de um vetor à sua origem (figura 2.4). Segundo Haykin (2001), quanto mais próximos de si

estiverem os vetores u e u_i , menor será a distância euclidiana e, portanto, maior será a similaridade entre eles.

Os pontos $\{u_i / i = 1, 2, \dots, N\}$ são tomados como centros das funções de base radial. Dessa forma, a ativação de uma unidade escondida (neurônio) em uma RN-RBF é determinada por uma função não-linear entre o vetor de entrada e um vetor de referência.

Figura 2.4 Ilustração da distância euclidiana entre dois vetores.



Escrevendo as condições de interpolação em forma matricial, obtemos o seguinte conjunto de equações lineares simultâneas para os coeficientes (pesos) desconhecidos $\{w_i\}$:

$$\begin{bmatrix} \varphi_{11} & \varphi_{12} & \dots & \varphi_{13} \\ \varphi_{21} & \varphi_{22} & \dots & \varphi_{23} \\ \vdots & \vdots & \dots & \vdots \\ \varphi_{N1} & \varphi_{N1} & \dots & \varphi_{NN} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \quad (2.19)$$

A matriz escrita em (2.19) é conhecida como matriz de interpolação. Considerando que ϕ representa uma matriz N -por- N com elementos φ_{ji} , pode ser escrita na forma compacta $\phi \cdot w = y$. Haykin (2001), Orr (1996) e De Castro e De Castro (2001) apresentam um conjunto de funções adequadas para o sistema de equações das incógnitas w_i para as quais existe uma única solução. O quadro 2.2 traz exemplos destas funções, comumente utilizadas como funções de base radial:

O parâmetro σ das funções de base radial determina quão rapidamente o valor da função de base radial cai a zero à medida que u se afasta do centro u_i . No caso da função do tipo gaussiana (a qual é a mais comumente usada em aplicações práticas), o parâmetro σ corresponde ao desvio padrão da função gaussiana. Assim, σ define a distância euclidiana média (raio médio) que mede o espalhamento dos dados representados pela função de base radial em torno do seu centro (De Castro e De Castro, 2001).

Quadro 2.2 Tipos principais de funções de base radial.

Lâmina <i>spline</i> fina	$\varphi(u) = \frac{u - u_i}{\sigma^2} \log\left(\frac{u - u_i}{\sigma}\right), \sigma > 0$
Multi-quadrática	$\varphi(u) = \sqrt{(u - u_i)^2 + \sigma^2}, \sigma > 0$
Multi-quadrática inversa	$\varphi(u) = \frac{1}{\sqrt{(u - u_i)^2 + \sigma^2}}, \sigma > 0$
Gaussiana	$\varphi(u) = \exp\left(-\frac{(u - u_i)^2}{2\sigma^2}\right), \sigma > 0$

As primeiras tentativas de implementação de RN-RBF utilizavam tantas funções de base radial quantos fossem os padrões do conjunto de dados representativos do problema em questão, objetivando uma aproximação mais exata. Entretanto, esta abordagem além de computacionalmente custosa, gerava o problema de *overfitting*, ou seja, a rede perdia sua capacidade de generalização. A solução deste problema foi proposta por Broomhead e Lowe (1988), os quais sugeriram que nem todos os vetores de entrada tivessem uma função de base radial associada. Isso faz com que o número de vetores pertencentes ao conjunto de treino da RN-RBF seja maior que o número de funções adotadas para a rede, tornando o sistema de equações lineares (cujas incógnitas são os pesos w_i) sobre-determinado, ou seja, terá mais equações que incógnitas. Essa metodologia resultou em significativa redução computacional e aumento da capacidade de generalização das RN-RBF.

A estrutura de uma RN-RBF é constituída basicamente por três camadas com papéis distintos. A camada de entrada é constituída por unidades sensoriais que conectam a rede ao seu ambiente, enquanto que a segunda (e única) camada oculta da rede aplica uma transformação não-linear do espaço de entrada para o espaço oculto.. A camada de saída é linear e composta por um único neurônio, fornecendo a resposta da rede ao padrão de ativação aplicado à camada de entrada. A figura 2.4 ilustra um exemplo genérico de RN-RBF.

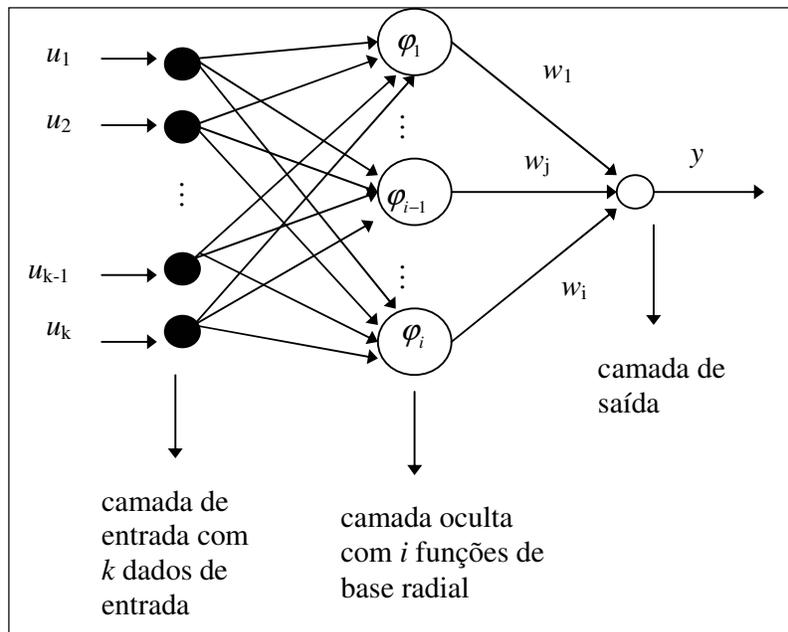
O mapeamento não-linear realizado pela camada oculta da RN-RBF, utilizando funções de ativação gaussianas é dado por

$$\varphi_i(n) = \varphi_i(u(n), u_i(n), \sigma_i^2(n)) = \exp\left(-\frac{\|u(n) - u_i(n)\|^2}{2\sigma_i^2}\right) \quad (2.20)$$

onde $u(n) \in \mathfrak{R}^M$ representa o vetor de entrada u no instante n , $u_i(n) \in \mathfrak{R}^M$ representa o vetor centro da i -ésima função de base radial $i=0, 1, \dots, i-1, i$ e $\sigma_i^2(n) \in \mathfrak{R}$ é a variância associada a cada uma das funções no instante n . O neurônio que compõe a camada de saída é definido como um combinador linear das funções de base radial. A saída y da rede é, portanto, a soma das saídas de cada função de ativação ponderadas pelos respectivos pesos sinápticos w , de tal forma que a combinação linear é expressa por¹⁰:

$$y = \sum_{i=0}^i w_i \varphi_i(u, u_i, \sigma_i^2) \quad (2.21)$$

Figura 2.5 RN função de base radial.



A implementação de uma RN-RBF envolve a determinação, através de um processo de aprendizagem, dos valores adequados aos parâmetros livres da rede, ou seja, as variâncias σ_i^2 , os centros u_i , e os pesos sinápticos w_i . O treinamento consiste em saídas y aproximem-se o máximo possível do conjunto de valores desejado.

Diferentes algoritmos podem ser usados para o ajuste dos parâmetros livres das RN-RBF, como, por exemplo, algoritmos genéticos, mapas de Kohonen, algoritmo de mínimos quadrados ortogonal, filtro de Kalman e o algoritmo *fuzzy c-means*. Dentre os mais comumente utilizados estão o algoritmo *k-médias* para início e/ou atualização dos

¹⁰ À equação 7 pode, em alguns casos, ser ainda acrescido um termo constante de polarização ou *bias*.

centros das funções de base radial e a metodologia de mínimos quadrados para cálculo e atualização dos pesos (Coelho, 2000; Haykin, 2001; De Castro e De Castro, 2001).

O ajuste dos parâmetros livres da RN-RBF utilizando algoritmo *k-médias* e o método de mínimos quadrados é descrito a seguir. Na primeira etapa, o algoritmo *k-médias* coloca os centros das funções de base radial apenas nas regiões do espaço de entrada onde dados significativos estão presentes, particionando os dados de entrada em subgrupos homogêneos. Para isso, escolhe-se inicialmente valores aleatórios para os centros u_i , com a restrição que esses valores iniciais sejam diferentes. Retira-se então um vetor x pertencente ao espaço de entrada e, a cada iteração, o algoritmo determina as distâncias euclidianas entre o vetor x e cada um dos centros u_i . Após n iterações, o centro escolhido é aquele que apresentar a maior similaridade com o vetor x , ou seja, aquele que apresenta a menor distância euclidiana,

$$k(x) = \arg \min_k \|x(n) - u_i(n)\| \quad (2.22)$$

onde $k(x)$ é o índice do centro com maior casamento com o vetor de entrada x e $k=1,2,\dots,i$ é o número de funções de base radial utilizadas. O algoritmo *k-médias* escolhe tantos centros quantos forem o número de funções de base radial determinado pelo usuário para o problema. O conjunto de dados é particionado de modo que para cada amostra x_i no espaço de entrada é atribuído um *cluster* com centro mais próximo.

O algoritmo de agrupamento *k-médias* segue o procedimento:

- i. Escolher aleatoriamente a posição inicial dos M_{ui} centros dos *clusters* para $j=1,\dots,M_{ui}$, aleatoriamente, no conjunto de dados de entrada;
- ii. Para $i=1,\dots,N$, onde N é o número de amostra do conjunto de dados, encontrar o centro u_k tal que a distância do vetor de entrada x_i seja mínima, ou seja, $\|x_i - u_k\| = \min_j \{\|x_i - u_j\|\}$, para $j=1,\dots,M_{ui}$ e atribuir a amostra i para o *cluster* k .
- iii. Calcular a média dos pontos de dados pertencentes a cada *cluster*:

$$\bar{x}_j = \frac{1}{N_j} \sum_{i \in u_j} x_i \quad , \text{ para } i=1,\dots,N \quad (2.23)$$

onde N_j é o número de amostras pertencente ao *cluster* j . O novo centro de cada *cluster* é então atualizado como sendo a média de seus pontos de dados, ou seja,

$$u_j = \bar{x}_j;$$

- iv. Repetir os passos (ii) e (iii) até que os centros dos *clusters* e a pertinência converjam.

Em seguida, o algoritmo heurístico *p-nearest neighbor* (Moody e Darken, 1989) é utilizado para a determinação da largura σ_j das funções de base radial. Após as posições dos centros da RN-RBF estarem fixas é aplicado o algoritmo *p-nearest neighbor* calcula a largura das funções como sendo a média sobre todas as distâncias euclidianas entre o centro de cada unidade de processamento (neurônio) n_i e o centro da unidade n_j mais próxima (Braga *et al.* 2000):

$$\sigma_j = \frac{1}{p} \sum_{k=1}^p \|u_j - u_k\| \quad (2.24)$$

onde u_k é o k -ésimo *cluster* mais próximo de u_j .

A próxima etapa consiste no cálculo do vetor de pesos da camada de saída. Considere a matriz de interpolação definida em (5) que pode ser escrita na forma compacta

$$\phi \cdot w = y \quad (2.25)$$

A solução de mínimos quadrados ordinários (ou pseudo-inversa de Penrose-Moore) para o vetor de pesos w_i é dada por

$$w = (\phi^T \phi)^{-1} \phi \cdot y \quad (2.26)$$

2.2.1.3 Redes neurais em previsão de séries temporais

Diversas metodologias têm sido propostas na literatura no intuito de utilizar redes neurais em previsão de séries temporais (Narendra e Parthasarathy, 1990; Chen e Billings, 1989). As mais utilizadas são as redes neurais com representação matemática do tipo NARX (*Non-linear Auto-Regressive with eXogenous inputs*) e representação matemática do tipo NARMAX (*Non-linear Auto-Regressive Moving Average with eXogenous inputs*), descritas a seguir:

RN-NARX

$$y(t) = f(y(t-1), \dots, y(t-n_y), u(t-1), \dots, u(t-n_u)) + e(t) \quad (2.27)$$

$$y(t) = f\left(\sum_{i=1}^{n_y} b_{yi} y(t-i) + \sum_{i=1}^{n_u} b_{ui} u(t-i)\right) + e(t) \quad (2.28)$$

RN-NARMAX

$$y(t) = f(y(t-1), \dots, y(t-n_y), u(t-1), \dots, u(t-n_u), e(t-1), \dots, e(t-n_e)) + e(t) \quad (2.29)$$

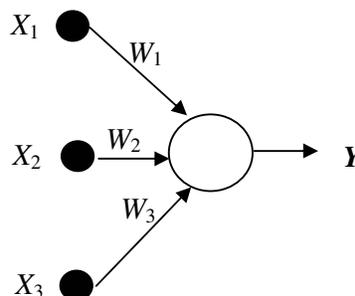
$$y(t) = f\left(\sum_{i=1}^{n_y} b_{yi} y(t-i) + \sum_{i=1}^{n_u} b_{ui} u(t-i) + \sum_{i=1}^{n_e} b_{ei} e(t-i)\right) + e(t) \quad (2.30)$$

Nas equações 2.27 a 2.30, $y(t)$, $u(t)$ e $e(t)$ são, respectivamente, as saídas do sistemas, as entradas e o componente média móvel; n_y , n_u e n_e são as defasagens máximas das saídas, das entradas e do componente média móvel. $\{e_i\}$ é assumido ser ruído branco. A função f é uma função não-linear que descreve o comportamento do sistema. No caso das RN-MLP, a função f representa a função de ativação, que pode ser sigmoideal. No caso das RN-RBF, a função f é uma função de base radial. O termo *exogenous inputs* deve-se ao fato de que as entradas $u(t)$ são originadas de fora do sistema, ou seja, são os próprios valores observados da série. Já as entradas $y(t)$ são originadas dentro do sistema, ou seja, a entrada $y(t)$ equivale à saída $y(t-1)$ da rede.

Fernandes *et al.* (1995) notam que os modelos de redes neurais têm vários pontos de contato com os modelos econométricos tradicionais, tais como o modelo de regressão, os modelos de séries de tempo ARMA/ARIMA (Box e Jenkins, 1976), e o modelo estrutural de séries de tempo. Os diferentes jargões técnicos adotados pelos economistas e engenheiros de controle de processos e/ou processamento de sinais impedem uma (desejável) sobreposição do conhecimento entre as áreas e uma unificação dos termos técnicos adotados.

No caso do modelo de regressão, por exemplo, temos uma variável endógena sendo explicada por uma ou mais variáveis exógenas. Nas RNs, as variáveis exógenas podem ser vistas como sinais da camada de entrada dos neurônios da rede enquanto que a variável endógena é vista como a saída desejada da rede, ou seja, ela é o objeto do aprendizado da RN. Supondo uma RN sem camada(s) intermediária(s), representada pela figura 2.6,

Figura 2.6 Rede neural sem camada intermediária



pode-se concluir que o vetor de pesos W_1 , W_2 , W_3 da RN nada mais é do que o vetor de parâmetros do modelo de regressão linear. Entretanto, quando introduzimos uma ou mais camadas intermediárias, estamos introduzindo não-linearidades no relacionamento entre as entradas X_1 , X_2 e X_3 e a variável de saída Y , já que as funções de ativação presentes nos neurônios das camadas intermediárias são intrinsecamente não-lineares.

Ao contrário dos modelos de regressão paramétrica, onde a forma funcional entre as variáveis dependentes e independentes é definida a priori, a abordagem das RNs caracteriza-se pela ausência (completa ou quase completa) de conhecimento a priori a respeito da forma funcional que está sendo estimada, o que faz com que as RNs sejam também consideradas um modelo de regressão não-paramétrica (Von Zuben e De Castro, 2001). Sendo assim, em RNs, mesmo que a função continue a ser estimada a partir do conjunto de parâmetros livres, o conjunto de “formas” que a função pode assumir (classe de funções que o modelo do estimador pode “aprender” e prever) é muito amplo.

Tomando um problema econométrico usual do tipo $y=f(x, \theta)$, a tarefa da rede neural consiste em encontrar uma forma funcional f e um conjunto de parâmetros θ (pesos sinápticos) que minimize o somatório do quadrado dos erros. A escolha da forma funcional f está associada ao modelo de rede e sua arquitetura, isto é, o número de camadas e número de neurônios nas camadas. A estimação do vetor de parâmetros θ depende, por sua vez, do método de treinamento da RN. Portugal e Fernandes (1996) afirmam que a grande novidade introduzida pelas RNs consiste em focalizar a atenção na forma funcional f , o que as torna extremamente úteis em problemas de classificação, reconhecimento de padrões, previsão etc, pois possuem a propriedade de serem “aproximadores universais” (Hornick *et al.*, 1989).

De acordo com Portugal e Fernandes (1996) os procedimentos comumente utilizados para determinação das variáveis exógenas relevantes como, por exemplo, funções de autocorrelação e autocorrelação parcial, testes para raízes unitárias, etc auxiliam na determinação do número de neurônios da camada de entrada, entretanto não ajudam muito no que diz respeito à determinação da arquitetura global da rede. Essa última questão é essencialmente empírica. A sugestão dada pelos autores é utilizar algum critério de poda, ou seja, começar-se com uma rede “grande” que sofre um processo sucessivo de poda até que se atinja uma arquitetura ideal, ou seja, elimina-se as conexões que mostrem pouca contribuição à capacidade preditiva da rede. Isso pode ser feito utilizando critérios de seleção de modelo, como o critério de informação Akaike (AIC) ou o critério Schwartz (SBC) (Enders, 1995).

Enquanto que nos modelos econométricos usuais, a estimação dos parâmetros é feita através de um processo que envolve minimização da soma do quadrado dos resíduos ou minimização de outro tipo de função custo, nas RNs a determinação dos pesos é feita através de um processo de aprendizado ou treinamento. Portugal e Fernandes (1996) afirmam que do ponto de vista matemático e estatístico a estimação dos pesos das RNs e a determinação dos coeficientes de modelos como o ARIMA, por exemplo, é a mesma coisa. Em ambos os casos, o que se procura atingir é a minimização do erro quadrado médio, ou qualquer outra função de custo escolhida. No caso das RNs, como foi visto, esse objetivo pode ser atingido através do algoritmo de retropropagação do erro (no caso das RN-MLP) ou através da minimização da distância euclidiana entre os vetores de entrada e saída (no caso das RN-RBF).

Atualmente, a principal deficiência da abordagem de RNs comparadas ao modelos econométricos tradicionais é a falta de uma base estatística pré-determinada. Enquanto que nos modelos econométricos estamos lidando com variáveis aleatórias que possuem uma determinada distribuição conjunta de probabilidade, nas RNs tem-se apenas sinais de entrada(s) e saída(s) da RN. A ausência de um modelo estatístico bem especificado impede, por exemplo, a construção de intervalos de confiança para as previsões geradas pelas RNs. Desta forma, as RNs geram previsões sempre pontuais, ao contrário do que ocorre com os modelos econométricos (Portugal e Fernandes, 1996).

2.2.1.4 Aplicações de redes neurais em Economia e Finanças

A literatura empírica que trata de aplicações de RNs em economia e finanças tem crescido exponencialmente nos últimos 10 anos, quando passaram a surgir os primeiros artigos sobre o tema publicados nos periódicos internacionais de primeira linha em economia e/ou econometria, como o *Econometrica*, *Journal of Econometrics* e o *Econometric Journal*, dentre outros. Wong e Selvi (1998) fazem uma revisão de parte da literatura empírica publicadas entre 1990 e 1996 (97 artigos) e concluem que os temas mais visados pelos pesquisadores são: previsão de falência de empresas, precificação de opções, previsão de séries temporais financeiras, seleção de ações, previsão da performance de portfólios, *rating* de títulos, sistemas de negociações e apoio à tomada de decisão, avaliação de risco de crédito, dentre outros.

Alguns trabalhos que utilizam redes neurais na previsão de taxas de câmbio devem ser destacados. Kaashoek e van Dijk (2002), por exemplo, concluem que a abordagem não-linear de redes neurais é superior aos modelos ARIMA na previsão de taxas de câmbio do Sistema Monetário Europeu. Resultados semelhantes foram encontrados por

Kuan e Liu (1995), utilizando taxas de câmbio da Libra inglesa, do Dólar canadense, do Marco alemão e do Iene japonês. Já Andrada-Félix *et al.*(2003) e Franses e Griesnven (1998) comparam a lucratividade de uma estratégia de negociação baseada em redes neurais com uma estratégia baseada em modelos ARIMA. Segundo os autores, a estratégia baseada em redes neurais conseguiu retornos ajustados ao risco superiores às estratégias baseadas em modelos lineares.

Portugal (1995) utilizando dados da produção industrial do estado do Rio Grande do Sul entre 1981 e 1983, investigou o poder preditivo do modelo auto-regressivo integrado de médias móveis (ARIMA), do modelo dos componentes não-observáveis (UCM) e do modelo de redes neurais. Os resultados do estudo mostram que o modelo de redes neurais é capaz de fazer previsões de curto prazo com maior precisão que o modelo ARIMA, ao passo que previsões prazo maior são superadas pelo modelo dos componentes não-observáveis.

Freitas e Souza (2002) utilizaram um modelo de redes neurais artificiais para a precificação de opções do mercado brasileiro, comparando com o modelo Black-Scholes. Foi utilizada uma rede com uma camada de entrada formada por cinco parâmetros (variáveis), uma camada intermediária com quatro neurônios e uma camada de saída com um neurônio. As conexões estabelecidas foram do tipo *feedforward*. Os autores concluíram que o modelo baseado em redes neurais foi capaz de superar o modelo Black-Scholes na precificação de opções fora do preço, no preço e dentro do preço.

Silva *et al.* (2002) estimam funções não-lineares de oferta e demanda por exportações do Brasil, utilizando a metodologia de RNs. Tendo aceitado a hipótese de presença de não-linearidades nas séries de exportação, os autores realizam uma análise de sensibilidade das séries de oferta de exportações fornecendo à rede já treinada um choque de 1% de crescimento da capacidade produtiva do País e um choque de 1% na taxa de câmbio. Para a série de demanda total de exportações, é fornecido à rede um choque de 1% de crescimento do PIB mundial e um choque de 1% no índice de competitividade dos países. As RNs foram capazes de identificar as elasticidades de todas essas variáveis em cada um dos anos analisados (1978 a 1999), proporcionando uma análise sobre como essas variáveis estão afetando as séries de demanda e oferta de exportações a cada ano.

Um modelo de previsão de retornos de ações do mercado norte-americano usando redes neurais e variáveis fundamentalistas é proposto por Eakins e Stansell (2003). Utilizando uma rede neural *feedforward* com uma camada intermediária e 7 variáveis

fundamentalistas (cada uma representando um neurônio na camada de entrada da rede) os autores concluem que a RN foi capaz de selecionar portfólios de ações com retornos e índices de Sharpe superiores aos índices de ações S&P500 e Dow Jones. De forma semelhante, Olson e Mossman (2003) concluem que as RNs foram superiores aos modelos de escolha discreta (LOGIT) e ao modelo de mínimos quadrados ordinários na seleção de ações através de variáveis fundamentalistas.

Oliveira e Ludemir (2002) analisam o poder preditivo de uma RN-MLP com representação matemática NARX para o IBOVESPA (índice da Bolsa de Valores de São Paulo) comparando com o poder preditivo de um modelo *random walk*. Os resultados indicaram que a RN com representação matemática NARX obteve um erro médio relativo de previsão (MRE – *mean relative error*) inferior ao modelo *random walk*, tanto na fase de treinamento quanto na fase de validação dos resultados.

Fernandes *et al.* (1995) comparam a metodologia dos componentes não-observáveis e a metodologia de RNs na previsão de uma série de índice mensal de passageiros de linhas aéreas. A metodologia dos componentes não-observáveis consiste basicamente em decompor a série em seus componentes sazonal, irregular e de tendência. Os autores conceberam uma RN em que os neurônios de entrada eram os próprios componentes não-observáveis da série. Os resultados indicaram que a metodologia tradicional conseguiu previsões um passo à frente mais acuradas, enquanto que para previsões doze passos à frente, a metodologia de RNs mostrou-se mais adequada.

Uma comparação entre vários modelos de previsão de índices de ações é feita por Leung *et al.* (2000). Os autores dividem os modelos em modelos de classificação (prevê, com base em probabilidade, a direção do movimento do índice) e modelos de previsão de nível (prevê a quantos pontos irá chegar determinado índice). Para o primeiro grupo foram analisados os modelos de escolha discreta LOGIT e PROBIT, análise discriminante linear e redes neurais. Para o segundo grupo foram analisados os modelos de ajustamento exponencial, funções de transferência multivariadas, modelos vetoriais autoregressivos e redes neurais. Além disso, foi avaliada uma estratégia de negociação baseada nas indicações (compra ou venda) de cada modelo de ambos os grupos. Os autores concluem que as estratégias de negociação baseadas em redes neurais conseguiram um retorno superior ao dos modelos concorrentes (de ambos os grupos) para todos os índices de ações analisados (S&P 500, FTSE 100 e Nikkei 225). A superioridade da abordagem de RNs em relação aos modelos de escolha discreta e de análise discriminante linear é também verificada em Dasgupta *et al.* (1994), os quais

analisam a capacidade desses modelos em identificar segmentos de consumo de produtos financeiros baseado na propensão dos consumidores em assumir riscos financeiros e adquirirem produtos de investimento não-tradicionais.

Kapetanios (2002) apresenta uma série de testes estatísticos para quebras estruturais em séries temporais usando redes neurais. A hipótese do autor é que a presença de não linearidade nas séries faz com que os modelos tradicionais de quebra estrutural para modelos lineares não tenham a performance desejada sob a hipótese nula de presença de quebra estrutural, em função de uma possível má especificação do modelo. A metodologia consiste em, primeiramente, aproximar a esperança condicional de uma variável dependente através de uma rede neural do tipo função de base radial e, posteriormente, usar o resíduo dessa aproximação para testar a presença de quebra estrutural usando testes tradicionais. Outro campo importante da aplicação de redes neurais consiste na formulação de testes estatísticos para presença de não linearidades em séries temporais (Blake e Kapetanios, 2003; Lee *et al.* 1993).

A comparação da capacidade preditiva do modelo de precificação de ativos APT (*Arbitrage Pricing Theory*) é comparado por Refenes *et al.* (1994) a uma RN-MLP treinada com algoritmo de retropropagação do erro. Novamente, a abordagem de RNs superou a abordagem tradicional com relação ao erro de previsão. Outras referências importantes de aplicações de redes neurais podem ser encontradas em teoria dos jogos (SgROI e Zizzo, 2002), modelagem e previsão de volatilidade (Donaldson e Kamstra, 1997; Hu e Tsoukalas, 1999), previsão de séries temporais utilizando indicadores de análise técnica (Yao *et al.* 1999), previsão de recessão usando indicadores macroeconômicos (Qi, 2001), aproximação de soluções para modelos macroeconômicos de crescimento estocásticos (Duffy e Mcnelis, 2001).

2.2.2 Sistemas nebulosos (*fuzzy systems*)

Dois dos principais aspectos da imperfeição da informação são a imprecisão e a incerteza. Estas duas características são intrinsecamente ligadas e opostas entre si: quanto mais se aumenta a incerteza mais se diminui a imprecisão e vice-versa (Sandri e Correa, 1999). Nesse sentido, a teoria dos conjuntos nebulosos foi desenvolvida a partir de 1965 por Lotfi Zadeh (Zadeh, 1965) para tratar do aspecto vago da informação.

As teorias mais conhecidas para tratar da imprecisão e da incerteza são a teoria clássica dos conjuntos e a teoria de probabilidades. Sandri e Correa (1999) observam que a teoria clássica dos conjuntos não é capaz de tratar o aspecto vago da informação e

a teoria de probabilidades, na qual a probabilidade de um evento determina completamente a probabilidade do evento contrário, é mais adequada para tratar de informações freqüentes do que aquelas fornecidas por seres humanos.

Formalmente, um conjunto nebuloso A do universo de discurso Ω é definido por uma função de pertinência $\mu_A(x) : \Omega \rightarrow [0,1]$. Essa função associa a cada elemento x de Ω o grau $\mu_A(x)$, com o qual x pertence a A . A função de pertinência $\mu_A(x)$ indica o grau de compatibilidade entre x e o conceito expresso por A :

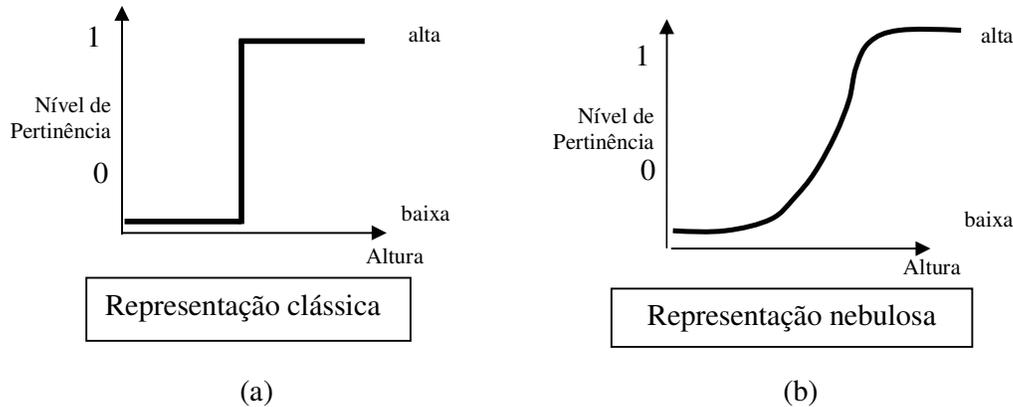
1. $\mu_A(x)=1$ indica que x é completamente compatível com A ;
2. $\mu_A(x)=0$ indica que x é completamente incompatível com A ;
3. $0 < \mu_A(x) < 1$ indica que x é parcialmente compatível com A , com grau $\mu_A(x)$.

Um conjunto A da teoria clássica dos conjuntos pode ser visto como um caso específico de conjunto nebuloso, denominado usualmente de “*crisp*”, para o qual $\mu_A(x) : \Omega \rightarrow \{0,1\}$, ou seja, a pertinência é do tipo “tudo ou nada”, “sim ou não”, e não gradual como para os conjuntos nebulosos (Sandri e Correa, 1999). Para ilustrar o problema dos conjuntos nebulosos, considere o problema da altura de um determinado indivíduo.

Enquanto que a lógica clássica é bivalente, ou seja, reconhece apenas dois valores (verdadeiro ou falso), a lógica nebulosa é multivalente, isto é, reconhece uma variedade de valores, assegurando que a verdade é uma questão de ponto de vista ou de graduação, definindo o grau de veracidade em um intervalo numérico $[0,1]$. Em outras palavras, o conjunto nebuloso é definido como sendo uma classe de objetos com contínuos graus de pertinência, a qual designa a cada objeto um grau de pertinência, definido entre o intervalo $[0,1]$.

Para ilustrar a idéia, é apresentado a seguir o problema da altura de determinada população. Supõe-se que os indivíduos “altos” recebem um grau de pertinência igual a 1 (verdadeiro), enquanto os indivíduos “baixos” recebem grau de pertinência 0 (falso). A representação gráfica dessa situação se dá de duas formas distintas (figura 2.7):

Figura 2.7 Representação clássica (a) e *fuzzy* (b) da altura de determinada população.



Caso seja definido que uma pessoa com altura de 1,90 metro como “alta”, de acordo com a representação clássica um indivíduo com altura de 1,88 metro receberia grau de pertinência 0 (falso), ou seja, seria considerado baixo. Entretanto, considerando a representação nebulosa, esse indivíduo poderia ser considerado alto, com nível de pertinência 0,9. Esse tipo de abordagem abre espaço para que considerações subjetivas sejam impostas, como do tipo “esse indivíduo é alto, mas *não muito*” ou então “indivíduo *um pouco* baixo”.

A teoria dos conjuntos nebulosos, quando utilizada em um contexto lógico, como o de modelos baseados em conhecimento, é conhecida como lógica nebulosa (do inglês *fuzzy logic*). Coelho (2000) e Sandri e Correa (1999) observam que a lógica nebulosa é uma das tecnologias atuais mais bem sucedidas para o desenvolvimento de modelos de identificação e previsão de sistemas, com aplicações que vão desde controle de eletrodomésticos ao controle de satélites, passando pelo mercado financeiro e a medicina.

Os fundamentos da lógica nebulosa (LN) levam em consideração o princípio de admitir incerteza e imprecisão na solução de problemas e permitir que se trabalhe com descrições simplificadas da realidade, possibilitando assim que a modelagem de sistemas complexos possa ser efetuada e que conclusões, possivelmente corretas, sejam extraídas. A definição dos conceitos da LN, a partir da extensão da teoria clássica de conjuntos, costuma ser apresentada como um simples trabalho de associação entre os conceitos da teoria dos conjuntos e conceitos da lógica. Logo, associa-se o conceito de pertinência ao valor-verdade de uma proposição lógica de união, interseção e complemento de conjuntos às operações *ou*, *e* e *não*, respectivamente. A inferência é vista como uma relação entre conjuntos (Coelho, 1997).

A formulação de modelos de identificação, controle e previsão de sistemas baseados em lógica nebulosa deu origem aos chamados sistemas nebulosos (Sandri e Correa, 1999), que são modelos compostos por um conjunto de regras do tipo **Se** <premissa> **Então** <conclusão> que definem um conjunto de ações de controle em função das diversas faixas de valores que as variáveis de estado do problema podem assumir.

2.2.2.1 Estrutura e classificação do sistemas nebulosos

Segundo Coelho (2000), os sistemas nebulosos (SNs) são usualmente compostos por quatro blocos principais:

- (i) *base de conhecimento*: definem as regras e os parâmetros das funções de pertinência;
- (ii) *procedimento de nebulização*: transformação das entradas “*crisp*” em conjuntos nebulosos;
- (iii) *unidade de inferência*: definem as operações de inferência sobre as regras;
- (iv) *procedimento de desnebulização*: transformação dos resultados nebulosos de inferência em saídas de dados “*crisp*”.

A *base de conhecimento* (i) é composta por uma *base de dados*, onde ficam armazenadas as definições sobre discretização e normalização dos universos de discurso, e as definições das funções de pertinência dos termos nebulosos e uma *base de regras*, formada por estruturas do tipo

Se <premissa> **Então** <conclusão>

como por exemplo:

se índice preço/lucro é alto
e empresa é altamente lucrativa
e vendas estão aumentando
então aprovação de crédito é muito segura

Segundo Babuska (2000) a parte <premissa> da regra é também denominada de proposição ou componente antecedente e é sempre um proposição nebulosa do tipo “ x é A ”, onde x é uma variável lingüística e A é um termo lingüístico constante. O valor verdadeiro da proposição (um número real entre 0 e 1) depende do grau de similaridade entre x e A . A parte da regra <conclusão> é chamada de proposição ou componente conseqüente.

A concepção de um modelo nebuloso é baseada em um mecanismo de “dividir para conquistar”, onde os antecedentes das regras nebulosas dividem o espaço de entrada em um número de regiões nebulosas locais, enquanto os conseqüentes descrevem o comportamento dentro de uma região através dos seus constituintes. Os componentes dos conseqüentes resultam em tipos diferentes de modelos (sistemas) nebulosos, mas seus antecedentes são essencialmente os mesmos.

O *procedimento de nebulização* (ii) faz a identificação dos valores dos dados de entrada, as quais caracterizam o estado do sistema (variáveis de estado), e as normaliza em um universo de discurso padronizado. Esses valores são então “nebulizados”, com a transformação da entrada “*crisp*” em conjuntos nebulosos para que possam se tornar instancias de variáveis lingüísticas.

Essas regras, juntamente com os dados de entrada, são processadas pela *unidade de inferência* (iii), a qual infere as ações de controle de acordo com o estado do sistema, aplicando o operador de implicação, se acordo com a seguinte seqüência:

- 1- verificação do grau de compatibilidade entre os fatos e as cláusulas nas premissas das regras;
- 2- determinação do grau de compatibilidade global da premissa de cada regra;
- 3- determinação do valor da conclusão, em função da compatibilidade da regra com os dados e a ação de controle constante na conclusão;
- 4- agregação dos valores obtidos como conclusão nas varias regras, obtendo-se uma ação de controle global

O *procedimento de desnebulização* (iv) é utilizado para obter uma única ação de controle precisa. Segundo Babuska (2000), um método usual para desnebulização é conhecido como *centro de gravidade*, dado por

$$y' = \frac{\sum_{j=1}^F \mu_{B'}(y_j) y_j}{\sum_{j=1}^F \mu_{B'}(y_j)} \quad (2.31)$$

onde $\mu_{B'}(y_i)$ representa a função de pertinência do conjunto nebuloso de saída (após agregação de todas as regras) e y_j os elementos do domínio da variável sobre a qual deseja-se fazer a inferência.

Os componentes dos conseqüentes resultam em tipos diferentes de modelos nebulosos, mas seus antecedentes são essencialmente os mesmos. Uma classificação

baseada nos componentes dos conseqüentes pode ser usada para classificar os sistemas nebulosos em baseados em (Coelho e Santos, 2003):

- (i) *modelos lingüísticos*: são baseados em regras *se-então* apresentando predicados vagos e utilizando raciocínio nebuloso. Nestes modelos, as quantidades nebulosas são associadas aos termos lingüísticos, e o modelo nebuloso é essencialmente uma expressão qualitativa do sistema. Os modelos, deste tipo, formam a base de modelagem qualitativa, que descreve o comportamento do sistema através da utilização da linguagem natural. O representante mais significativo desta classe é o modelo ou sistema do tipo Mamdani (Mamdani e Assilian, 1975);
- (ii) *modelos interpolativos*: são formados por regras lógicas por uma parte nebulosa antecedente e um conseqüente funcional. Estes modelos são, essencialmente, uma combinação de modelos nebulosos e não-nebulosos. O representante mais significativo dos modelos interpolativos é o modelo do tipo Takagi-Sugeno (Takagi e Sugeno, 1985). A representação do conhecimento, nestes modelos, caracterizam-se pelo potencial de expressar informações quantitativas. Entretanto, nos modelos interpolativos perde-se a informação qualitativa.

O sistema nebuloso apresentado no exemplo anteriormente pode ser classificado como um modelo lingüístico. Entretanto, para problemas envolvendo séries temporais, modelos interpolativos são mais adequados, em função das características acima descritas. Dessa forma, nesse trabalho será utilizado o sistema nebuloso do tipo Takagi-Sugeno (TS), descrito a seguir.

2.2.2.2 Sistema nebuloso de TS

Os modelos lingüísticos apresentados anteriormente eram formados por regras **SE-ENTÃO** com termos lingüísticos na parte antecedente e conseqüente. O modelo Takagi-Sugeno, por outro lado, usa funções “*crisp*” na parte conseqüente. Dessa forma, pode ser visto como uma combinação entre conhecimento lingüístico (parte antecedente) e regressão matemática (parte conseqüente) de tal forma que os antecedentes descrevem regiões nebulosas no espaço de entrada nas quais as funções conseqüentes são válidas (Babuska, 2000). As regras do modelo TS são estruturadas da seguinte forma:

$$R_i : SE x_1 \acute{E} A_{i1} E \dots E x_n \acute{E} A_{in} ENT\tilde{A}O \hat{y}_i = f_i(x_1, \dots, x_n), \quad i = 1, 2, \dots, K \quad (2.32)$$

onde $f_i(x_1, \dots, x_n)$ é uma função “*crisp*” no conseqüente. A função $f_i(x_1, \dots, x_n)$ é usualmente um polinômio com variáveis de entrada x (Coelho, 2000).

Uma maneira comum de representar a função f é através da função afim, que é linear nos parâmetros, gerando regras do tipo

$$R_i : SE x_1 \acute{E} A_{i1} E \dots E x_n \acute{E} A_{in} ENT\tilde{A}O \hat{y}_i = \mathbf{a}_i^T \mathbf{x} + b_i, \quad i = 1, 2, \dots, K \quad (2.33)$$

onde \mathbf{a}_i^T é um vetor de parâmetros e b_i é um escalar; R_i é a i -ésima regra, $\mathbf{x} = [x_1, \dots, x_n]^T$ é o vetor das variáveis de entrada das regras (antecedentes), A_{i1}, \dots, A_{in} são conjunto nebulosos definidas no espaço dos antecedentes, e y_i é a saída da regra. A variável K denota o número de regras na base de regras, e a saída agregada do modelo \hat{y} é calculada pela média ponderada dos conseqüentes da regra

$$\hat{y} = \frac{\sum_{i=1}^K \beta_i(\mathbf{x}) \hat{y}_i}{\sum_{i=1}^K \beta_i(\mathbf{x})} \quad (2.34)$$

ou, equivalentemente,

$$\hat{y} = \frac{\sum_{i=1}^K \beta_i(\mathbf{x}) (\mathbf{a}_i^T \mathbf{x} + b_i)}{\sum_{i=1}^K \beta_i(\mathbf{x})} \quad (2.35)$$

onde $\beta_i(\mathbf{x})$ é o grau de ativação da i -ésima regra, calculado por

$$\beta_i(\mathbf{x}) = \prod_{j=1}^n \mu_{A_{ij}}(x_j), \quad i = 1, 2, \dots, K \quad (2.36)$$

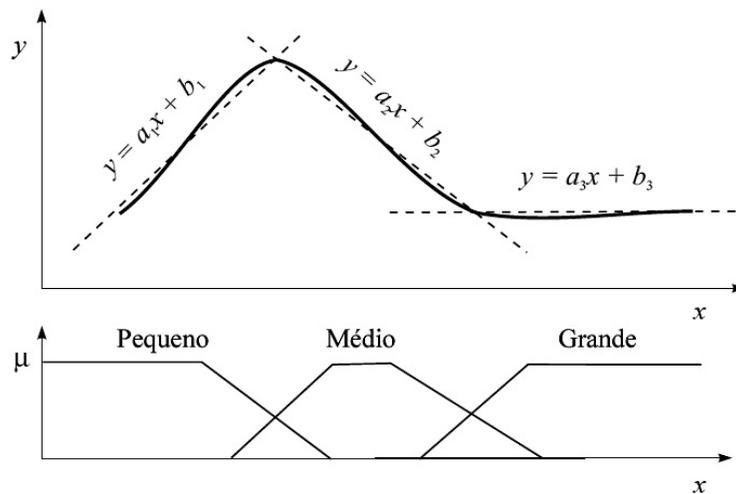
onde $\mu_{A_{ij}}(x_j) : \mathfrak{R} \rightarrow [0, 1]$ é a função de pertinência do conjunto nebuloso A_{ij} no antecedente de R_i . A cada conjunto nebuloso do antecedente A_{ij} é associada uma função de pertinência $\mu_{A_{ij}}(x_i)$ descrita por

$$\mu_{A_{ij}}(x_i) = \exp\left[-\frac{1}{2} \frac{(x_i - m_{ij})^2}{\sigma_{ij}^2}\right] \quad (2.37)$$

onde m_{ij} e σ_{ij} são respectivamente o centro e a largura função de pertinência do tipo Gaussiana, determinados através do algoritmo de *clustering k-médias* (equação 2.23) e do algoritmo heurístico *p-nearest-neighbor* (equação 2.24). A união de todos estes parâmetros define o conjunto de parâmetros da premissa.

Babuska (2000) nota que nos modelos nebulosos de TS os parâmetros a_i e b_i correspondem a uma linearização (aproximação) local de uma função não-linear (*smoothed piece-wise approximation*), como pode ser visto na figura 2.8.

Figura 2.8 Modelo nebuloso de Takagi-Sugeno visto como uma aproximação linear local de uma função não-linear (adaptada de Babuska, 2000)



O sistema nebuloso de TS trata-se de um equivalente funcional da rede neural de base radial. Neste caso, quando a rede neural apresenta o mesmo número de funções de ativação na camada oculta que o número de regras do modelo nebuloso TS de ordem zero (Wu e Tam, 1999). Um aspecto relevante do sistema de TS é o seu poder de representação, especialmente para a descrição de sistemas complexos. Este sistema nebuloso permite a decomposição de um sistema complexo em subsistemas simples.

Uma metodologia de previsão de séries temporais utilizando o modelo nebuloso TS, proposta por Roubos *et al.* (1999), e utilizada neste trabalho consiste na utilização

das representações matemáticas NARX e NARMAX, definidas anteriormente (equação 2.34). Supondo, por exemplo, uma representação NARX do tipo

$$y(t) = f(y(t-1), \dots, y(t-n_y), u(t-1), \dots, u(t-n_u)), \quad (2.38)$$

onde $y(t)$, $u(t)$ são, respectivamente, as saídas do sistemas e as entradas; n_y , n_u são as defasagens máximas das saídas e das entradas, podemos definir o conjunto de regras de um sistema nebuloso com representação matemática NARX por

$$\begin{aligned} R_i : SE \quad & y(t) \acute{E} A_{i1} \ E \ y(t-1) \acute{E} A_{i2} \ E \dots \ y(t-n-1) \acute{E} A_{in} \ E \ u(t) \acute{E} B_{i1} \\ & E \ u(t-1) \acute{E} B_{i2} \ E \dots \ u(t-n-1) \acute{E} B_{in} \ ENT\tilde{A}O \quad y(t+1) \acute{E} C_1. \end{aligned} \quad (2.39)$$

Babuska (2000) nota que a seleção da estrutura dos modelos nebulosos envolve as seguintes escolhas:

- (i) *Variáveis de entrada e saída*: nos modelos de previsão univariados, como aqueles representados matematicamente pelas estruturas NARX e NARMAX, a escolha das variáveis de entrada e saída significa escolher o número de regressores de acordo com suas defasagens n_y e n_u ;
- (ii) *Estrutura das regras*: o formato dos antecedentes e conseqüentes das regras está diretamente relacionado ao tipo de modelo nebuloso utilizado (lingüístico, interpolativo);
- (iii) *Número e tipo de função de pertinência para cada variável de entrada*: essa escolha determina o nível de detalhe (granularidade) do modelo;
- (iv) *Tipos de mecanismo de inferência, operadores conectivos e método de desnebulização*: essas escolhas são restritas de acordo com modelo nebuloso utilizado.

Após a estrutura estar fixa, a performance do modelo nebuloso pode ser majorada pelo ajuste dos parâmetros dos termos antecedentes e conseqüentes (no caso do modelo TS, os parâmetros das funções de pertinência e os parâmetros a e b da função conseqüente, respectivamente). Para a estimativa desses parâmetros, a técnica de agrupamento de dados ou clusterização nebulosa é utilizada para o reconhecimento de padrões nos dados de entrada. Baseado em alguma medida adequada de grau de

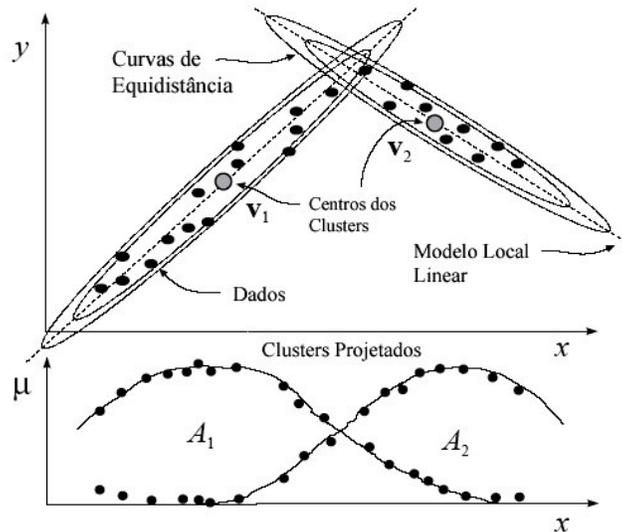
similaridade (distância euclidiana, por exemplo), vetores são clusterizados de modo que os dados dentro um determinado *cluster* sejam o mais similar possível e dados de diferentes *clusters* sejam o menos similar possível (Babuska, 2000).

A figura 2.9 ilustra a idéia de clusterização nebulosa, onde os dados são divididos em dois grupos de centros v_1 e v_2 usando o critério de distância euclidiana. A partição dos dados é expressa na matriz de partição nebulosa cujos elementos μ_{ij} são graus de pertinência dos pares de dados $[x_i, y_i]$ projetados em cada *cluster* nebuloso A_1 e A_2 .

Regras do tipo **SE-ENTÃO** podem ser extraídas projetando-se os *clusters* nos eixos.

$$\begin{aligned} \text{Se } x \text{ é } A_1 \text{ Então } y &= a_1x + b_1 \\ \text{Se } x \text{ é } A_2 \text{ Então } y &= a_2x + b_2 \end{aligned} \quad (2.40)$$

Figura 2.9 *Clusters* nebulosos em formato elipsoidal (adaptada de Babuska, 2000)



A figura 2.9 mostra um conjunto de dados com dois *clusters* associados a duas regras nebulosas. Cada *cluster* obtido é representado por uma regra no modelo TS. Dessa forma, a partir de cada *cluster*, funções de pertinência dos termos antecedentes e os parâmetros dos termos conseqüentes podem ser obtidos para o modelo TS.

Neste trabalho, utilizou-se o algoritmo de agrupamento ou clusterização Gustafson-Kessel (Gustafson e Kessel, 1979) para a determinação dos *clusters* nebulosos no conjunto de dados. O algoritmo Gustafson-Kessel caracteriza-se por

clusterizar os dados em elipsóides, com formato determinado de modo adaptativo (Babuska, 2000).

Os parâmetros do conseqüente de cada regra são obtidos por uma estimativa através do algoritmo dos mínimos quadrados. Considerando-se que X_e denota a matriz $[X; \mathbf{1}]$; Γ_i é a matriz diagonal em $\Re^{N \times N}$ tendo grau de pertinência normalizado

$$\gamma(\mathbf{x}_k) = \frac{\beta_i(\mathbf{x}_k)}{\sum_{j=1}^K \beta_j(\mathbf{x}_k)}$$

como seu k -ésimo elemento diagonal. Além disso, significa que X'

é a matriz em $\Re^{N \times KN}$ composta das matrizes Γ_i e X_e , onde

$$X' = [(\Gamma_1 X_e); (\Gamma_2 X_e); \dots; (\Gamma_K X_e)] \quad (2.41)$$

Significa θ' , o vetor em $\Re^{K(n+1)}$ dado por

$$\theta' = [\theta_1^T; \theta_2^T; \dots; \theta_K^T] \quad (2.42)$$

onde $\theta_i^T = [a_i^T; b_i]$ para $1 \leq i \leq K$. O resultado do problema dos mínimos quadrados $\mathbf{y} = X\theta' + \xi$, onde ξ é erro de aproximação, tem a solução

$$\theta' = [(X')^T X']^{-1} (X')^T \mathbf{y} \quad (2.43)$$

Da equação 2.49 os parâmetros a_i e b_i são obtidos por

$$a_i = [\theta'_{q+1}, \theta'_{q+2}, \dots, \theta'_{q+n}]^T, \quad b_i = [\theta'_{q+n+1}] \quad (2.44)$$

onde $q = (i - 1)(n + 1)$.

2.3 Síntese conclusiva

Este capítulo teve como objetivo apresentar os modelos econométricos de previsão utilizados neste trabalho. Inicialmente foi apresentada uma definição formal de modelo linear e não-linear, e em seguida introduzidos os modelos lineares ARMA e ARMA-GARCH, suas estruturas e procedimentos de estimação dos parâmetros. Posteriormente, foi apresentado o modelo não-linear de redes neurais, incluindo os dois tipos de rede

utilizados neste trabalho: redes neurais *perceptron* multicamadas e redes neurais função de base radial. São discutidos os procedimentos de treinamento e a estrutura de cada um dos modelos.

Por último, foi apresentada a abordagem não-linear dos sistemas nebulosos. Nessa seção, discute-se inicialmente as principais características desse tipo de modelo e, em seguida, apresenta-se o sistema nebuloso TS, utilizado neste trabalho.

3

Fundamentos práticos da previsão de séries temporais

Este capítulo discute alguns aspectos práticos relacionados a previsão de séries financeiras, como a escolha da estrutura dos modelos, identificação de não-linearidades e avaliação da acurácia de previsão.

3.1 Introdução

Os principais ingredientes ao problema da previsão de séries temporais são (Coelho, 2000):

- (i) Um conjunto de dados de entrada(s);
- (ii) Uma classe de modelos candidatos e suas respectivas estruturas;
- (iii) Um critério de adequação entre os dados e o modelo;
- (iv) As rotinas para validação e aceitação dos modelos resultantes.

Definidas essas questões, deve-se proceder à resolução das seguintes etapas (Coelho, 2000):

- (i) *Detecção de não-linearidades*: existe não-linearidades na série? Qual o critério usado para avaliar?
- (ii) *Escolha da estrutura*: quais termos estão no modelo?
- (iii) *Estimação de parâmetros*: quais são os valores dos coeficientes desconhecidos?
- (iv) *Validação do modelo*: o modelo está correto?
- (v) *Previsão*: qual a resposta do modelo para um instante futuro?

3.2 Detecção de não-linearidades

A especificação e estimação de modelos lineares de séries temporais, baseados em modelos univariados ARIMA, são procedimentos bem estabelecidos. Entretanto, a teoria econômica frequentemente sugere um relacionamento não-linear entre as variáveis, e muitos economistas aparentam crer que o sistema econômico como um todo é não-linear (Lee *et al.*, 1993). Dessa forma, é interessante testar se uma determinada série econômica aparenta ser gerada por um modelo linear ou se é melhor reproduzida por um modelo não-linear.

A maioria dos testes de não-linearidade tem como primeiro passo a extração de uma estrutura linear através do uso de um filtro estimado. Tipicamente, um modelo AR(p) é ajustado à série e então o teste é aplicado aos resíduos obtidos. Para otimizar o processo, um valor particular de p é escolhido (usualmente $p=1$ ou $p=2$), entretanto é possível escolher p através de algum critério de informação (Lee *et al.*, 1993).

Neste trabalho, utilizou-se o teste BDS (Brock *et al.*, 1986) para a detecção de não-linearidades nas séries. A partir de pesquisas relacionadas à presença de caos em dados econômicos, Brock *et al.* (1986) desenvolveram um teste apropriado para detectar não-linearidades. Para uma série y_t , define-se

$$C_m(\varepsilon) = n^{-2} [\text{número de pares } (i, j) \text{ tal que} \quad (3.1)$$

$$|y_i - y_j| < \varepsilon, |y_{i+1} - y_{j+1}| < \varepsilon, \dots, |y_{i+m-1} - y_{j+m-1}| < \varepsilon]$$

tal que y_i, \dots, y_{i+m-1} e y_j, \dots, y_{j+m-1} são dois segmentos de uma série com m observações de tal forma que todos os pares de pontos diferem uns dos outros pelo tamanho ε . A estatística do teste é $BDS = n^{1/2} [C_m(\varepsilon) - C_1(\varepsilon)^m]$. Sob a hipótese nula de que a série é independente e identicamente distribuída, BDS é assintoticamente normalmente distribuída com média 0 e variância σ .

Para Lee *et al.* (1993) o teste BDS é interessante, pois parte de considerações diferentes dos outros testes, além de ser robusto nos casos de séries que apresentem variância condicional dos resíduos (ARCH). Na implementação do teste, a série y_t deve ser substituída pela série dos resíduos estimados na regressão linear. Outras metodologias para verificação de não-linearidades são propostas por Keenan (1985), Tsay (1986), Ramsey (1969), Hinich (1982) e Lee *et al.* (1993).

Bolleslerv *et al.* (1992) argumentam que a aparente não-linearidade presente em series econômicas pode ser explicada por um fato estilizado na literatura em finanças: a heteroscedasticidade condicional auto-regressiva dos resíduos (efeitos ARCH). Segundo os autores, removendo os efeitos ARCH das séries, testes de não-linearidade tendem a aceitar a hipótese nula, ou seja, rejeitar a presença de não-linearidade na série. Dessa forma, a não-linearidade pode estar significando simplesmente uma característica comum ao segundo momento dos retornos das series financeiras.

Essa hipótese é corroborada por Abhyankar *et al.* (1995), os quais, analisando series de alta frequência do índice da bolsa de valores do Reino Unido (FTSE-100), concluem que efeitos ARCH estariam explicando em grande parte a presença de não-linearidades nas series. A metodologia dos autores consistiu em aplicar os testes de não-linearidade aos resíduos obtidos pela estimativa de um modelo GARCH, e não de um modelo AR, como é usualmente realizado.

3.3 Escolha da estrutura dos modelos

Os procedimentos estimação de parâmetros e determinação da estrutura dos modelos lineares estão bem sedimentados, como, por exemplo, a metodologia Box-Jenkins para determinação da estrutura dos modelos ARIMA. As dificuldades existem em se aplicar esses procedimentos aos modelos não-lineares (Coelho, 2000).

A escolha adequada dos termos pertencentes ao modelo evita efeitos dinâmicos espúrios. O número de termos para representar a dinâmica do sistema deve ser suficientemente pequeno para evitar a instabilidade numérica provocada por sobreparametrização (Coelho, 2000). A escolha da ordem (numero de defasagens incluídas) deve considerar um compromisso entre a capacidade de representação das dinâmicas do processo e a simplicidade do modelo, que representa um menor esforço computacional para os algoritmos de estimação.

Este trabalho utilizou o critério de informação Akaike (AIC) (Enders, 1995) para a escolha da estrutura dos modelos. A estatística AIC procura avaliar o *trade-off* entre somatório dos quadrados dos resíduos e número de parâmetros estimados, penalizado modelos pouco parcimoniosos. Quanto menor o número de parâmetros estimados e menor o somatório do quadrado dos resíduos, menor é a estatística AIC e, portanto, melhor é o modelo. A estatística AIC é dada por

$$AIC = T \ln(SQR) + 2n \quad (3.2)$$

onde SQR é o somatório dos quadrados dos resíduos do modelo, n é o número de parâmetros estimados e T é o número de observações. Através desse método, é possível julgar se o incremento na ordem do modelo traz uma contribuição significativa na capacidade de representar a serie temporal analisada (Coelho, 2000).

3.5 Avaliação do desempenho preditivo dos modelos

Baseado nos estudos de Chu e Zhang (2003), Terui e van Dijk (2002), Meade (2002) e Fair (1999), este trabalho utilizou os seguintes critérios de avaliação do desempenho preditivo dos modelos:

- *Root Mean Squared Error* (RMSE): Procura avaliar o tamanho do erro de previsão. Quanto maior o RMSE, maior o erro de previsão do modelo. O RMSE é dado por

$$RMSE = \sqrt{\frac{\sum_{j=1}^m e_{t+j}^2}{m}}, \quad (3.3)$$

onde e é o erro de previsão calculado para a observação $t+j$ e m é o número de observações.

- *Índice de Desigualdade de Theil* (U-Theil): O índice U-Theil procura avaliar o ajustamento da série prevista à série original. Quanto mais próximo a 1 significa que a série estimada pelo modelo tem pouco ajustamento à série original, ou seja, o modelo não conseguiu fazer boas previsões. Entretanto, quanto mais próximo a zero, maior é a similaridade da série prevista com a série original, indicando que o modelo consegue fazer boas previsões. O índice U-Theil é definido por

$$U - Theil = \frac{\sqrt{\sum_{t=T+1}^{T+m} (\hat{y}_t - y_t)^2 / m}}{\sqrt{\sum_{t=T+1}^{T+m} (\hat{y}_t)^2 / m + \sum_{t=T+1}^{T+m} (y_t)^2 / m}}, \quad (3.4)$$

onde \hat{y}_t é o valor previsto (estimado) para o período t , y_t é o valor observado e m é o número de observações.

- *Percentual de sinais corretamente previstos (S.C.P.):* O trabalho de Fernandez-Rodriguez *et al.* (2000) sugere que seja avaliado a capacidade dos modelos em prever a direção correta dos movimentos da série observada. Nesse sentido, deve ser avaliado o percentual de sinais corretamente previstos (S.C.P.). Quanto maior for o número de S.C.P., melhor será o modelo. Em outras palavras, quanto maior for o número de vezes em que o modelo consegue prever a direção do movimento, ou seja, quando o retorno do ativo será positivo ou negativo, melhor será seu desempenho. O percentual de S.C.P. é crucial para nos dizer se a estratégia de negociação baseada no modelo tenderá a ser mais rentável que uma estratégia simplista do tipo *buy-and-hold*.

- *Estatística de falha de previsão Pesaran-Timmermann:* A validação estatística dos modelos de previsão de séries financeira pode ser feita utilizando-se a estatística não-paramétrica Pesaran-Timmermann (Pesaran e Timmermann, 1992), também conhecida como teste de falha de previsão. A estatística Pesaran-Timmermann (PT) baseia-se na análise da previsão correta da direção da mudança da variável em questão. A estatística PT é particularmente interessante quando a distribuição de probabilidade (intervalos de confiança) das previsões é de difícil derivação por métodos analíticos, como no caso de um grande número de modelos dinâmicos não-lineares (Pesaran e Timmermann, 1992). A estatística PT é dada por

$$PT_n = \frac{\hat{P} - \hat{P}_*}{\{\text{var}(\hat{P}) - \text{var}(\hat{P}_*)\}^{1/2}} \sim N(0, 1) \quad (3.5)$$

onde \hat{P} é o percentual de vezes em que o sinal de mudança da variável foi previsto corretamente, \hat{P}_* é o valor esperado de \hat{P} , dado por $\hat{P}_* = P_-P_+ + (1 - P_-)(1 - P_+)$ onde P_- é o percentual calculado de sinais negativos e P_+ é o percentual calculado de sinais positivos. $\text{var}(\hat{P})$ e $\text{var}(\hat{P}_*)$ são dados por:

$$\text{var}(\hat{P}) = n^{-1}\hat{P}_*(1 - \hat{P}_*) \quad (3.6)$$

$$\begin{aligned} \text{var}(\hat{P}_*) &= n^{-1}(2\hat{P}_- - 1)^2\hat{P}_+(1 - \hat{P}_+) + n^{-1}(2\hat{P}_+ - 1)^2\hat{P}_-(1 - \hat{P}_-) \\ &\quad + 4n^{-2}\hat{P}_-\hat{P}_+(1 - \hat{P}_-)(1 - \hat{P}_+) \end{aligned} \quad (3.7)$$

A estatística P.T. nos diz se o modelo utilizado prevê, com significância estatística, a direção do movimento da série observada. A hipótese nula é que as séries estimada e observada são independentes. Dessa forma, a rejeição da hipótese nula indica que o modelo consegue prever a direção do movimento da série observada, ou seja, a previsão feita pelo modelo é estatisticamente significativa.

A utilização de critérios de desempenho baseados apenas no tamanho do erro de previsão é criticada por Leitch e Tanner (1991). A análise empírica feita pelos autores mostra pouca relação entre o erro de previsão avaliado e a lucratividade proporcionada pela estratégia de negociação baseada nas previsões dos modelos utilizados. Diante disso, Leitch e Tanner propõem que, em paralelo à avaliação do tamanho do erro de previsão, seja estabelecida uma estratégia de negociação baseada nas respostas dos modelos. A estratégia de negociação estabelecida neste trabalho é detalhada na seção 4.3.1.

3.6 Síntese conclusiva

Este capítulo tem como objetivo discutir alguns aspectos práticos relacionados à previsão de séries temporais. Dentre esses aspectos, destacam-se: i) um procedimento para detecção de não-linearidades nas séries, ii) o estabelecimento de critérios para a escolha correta da estrutura dos modelos, de modo a evitar a sobreparametrização e iii) a escolha dos critérios que irão avaliar o desempenho de previsão dos modelos.

Com relação ao aspecto (i), será utilizada a estatística BDS para identificação de possíveis não-linearidades presentes nas séries temporais. Para a escolha da estrutura dos modelos (aspecto (ii)) será utilizada o critério de informação de Akaike, que procura combinar parcimônia e minimização do somatório dos quadrados dos resíduos. Por fim, com relação ao aspecto (iii), serão utilizados os seguintes critérios para avaliação da acurácia de previsão dos modelos: *root mean squared error* (RMSE), índice de desigualdade U-Theil, percentual de sinais corretamente previstos e a estatística Pesaran-Timmermann de falha de previsão.

4

Análise empírica

Este capítulo traz os resultados da estimação e previsão realizadas pelos modelos lineares e não-lineares abordados neste trabalho. O capítulo está organizado da seguinte forma. Na seção 4.1 algumas características acerca das séries temporais utilizadas são apresentadas. Na seção 4.2 são discutidas algumas questões relativas aos procedimentos de estimação e os critérios utilizados na avaliação da performance preditiva. A seção 4.3 apresenta os resultados da estimação e avaliação do desempenho dos modelos, bem como a estratégia de negociação estabelecida com base nas previsões dos modelos. Por último, na seção 4.4, algumas conclusões do capítulo são apresentadas.

4.1 Séries temporais utilizadas e testes de especificação

A série histórica usada neste trabalho para a estimativa dos modelos de previsão foi obtida a partir da série de preços do último negócio da taxa de câmbio brasileira Real/Dólar (R\$/US\$), coletada em diferentes frequências. As séries de alta frequência, ou seja, aquelas em que as cotações são obtidas dentro de um dia de negociação, possuem observações coletadas a cada 15, 60 e 120 minutos. O período analisado vai de 01/01/2002 a 01/01/2003. A cada dia de negociação, as observações foram coletadas na faixa de horário que vai das 9:30 AM até as 16:30 PM. Além das séries de alta frequência, utilizaram-se séries de fechamento da taxa oficial de câmbio Real/Dólar com periodicidade diária e semanal, coletadas no período que vai de 01/01/2000 a 01/01/2004. A motivação para utilizar séries com diferentes frequências vem dos trabalhos de Abhyankar *et al.* (1995) e Campbell *et al.* (1997), os quais afirmam que a previsibilidade pode estar relacionada com a frequência com que a série é observada. As séries históricas foram coletadas a partir do *software* de acompanhamento do mercado financeiro Apligraf®.

As séries de retornos foram obtidas a partir da logaritmização dos valores observados, com o objetivo de linearização de tendências exponenciais e estabilização de variâncias. Deste modo, ao invés de calcular o retorno de forma discreta, sendo

$$R_t = \frac{P_t}{P_{t-1}} - 1, \quad (4.1)$$

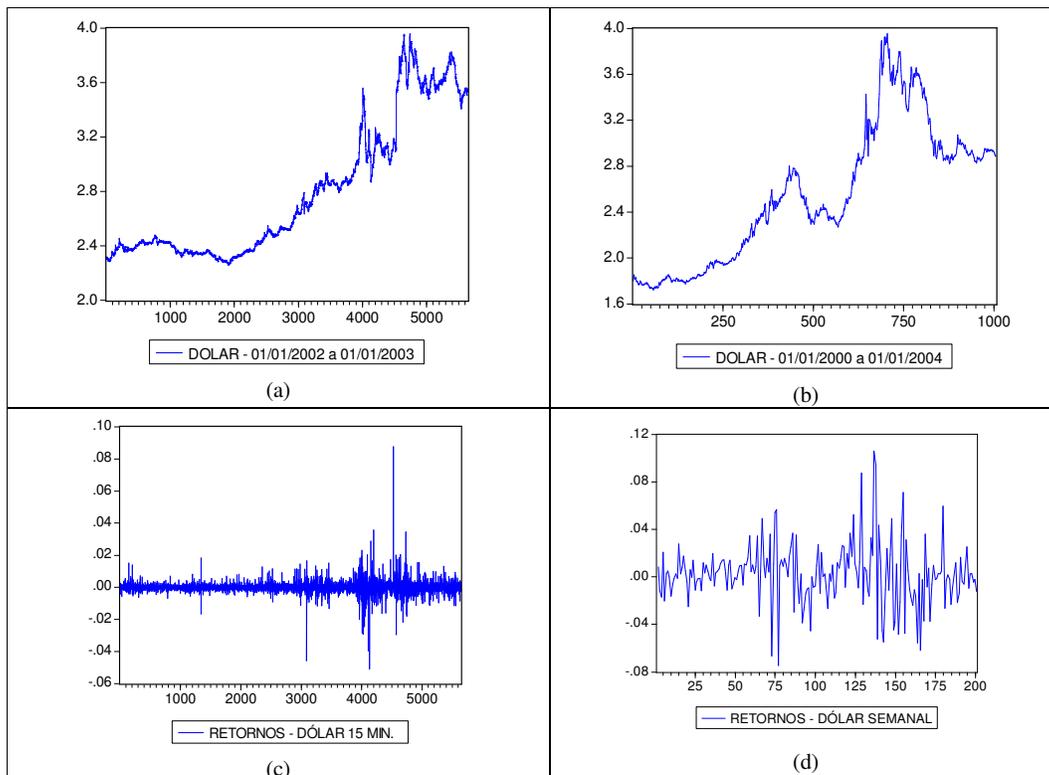
calculou-se na forma composta continuamente:

$$R_r = \ln(P_r) - \ln(P_{r-1}), \quad (4.2)$$

onde R_t é o retorno obtido na data t e P_t e P_{t-1} é o preço do ativo na data t e na data $t-1$, respectivamente. Esse tipo de transformação é usualmente utilizada para tornar a série estacionária e eliminar a presença de raízes unitárias, condição necessária para que as previsões convirjam para a média incondicional da série (Enders, 1995).

A figura 4.1 abaixo traz os gráficos da série da taxa de câmbio em nível para o período que vai de 01/01/2002 a 01/01/2003 (gráfico “a”) e para o período que vai de 01/01/2000 a 01/01/2004 (gráfico “b”). Os gráficos “c” e “d” trazem os retornos da taxa de câmbio para a série de menor frequência (15 min.) e para a série de maior frequência (semanal), respectivamente. Percebem-se nos gráficos “a” e “c” um aumento substancial na volatilidade da taxa de câmbio na segunda metade da amostra, fato ocorrido em função da proximidade das eleições presidenciais no Brasil, o chamado “Efeito Lula”.

Figura 4.1 Gráficos da série da taxa de câmbio Real/Dólar em nível e dos retornos da taxa de câmbio para a série de 15 min. e para a série semanal.



A tabela 4.1 traz a estatística descritiva das séries analisadas. O resultado do teste Jarque-Bera mostra que a hipótese de normalidade na distribuição dos retornos foi rejeitada em todas as séries em função da presença de elevada curtose e assimetria positiva (com exceção da série diária). O teste Dickey-Fuller Aumentado, por sua vez, mostrou que a hipótese nula de presença de raízes unitárias foi rejeitada em todas as séries, indicando que todas são estacionárias. O teste para presença de heteroscedasticidade condicional auto regressiva (ARCH), baseado no multiplicador de Lagrange, indicou a presença de ARCH em todas as séries.

Tabela 4.1 Estatística descritiva para as séries de retornos da taxa de câmbio Real/Dólar, de acordo com a frequência observada.

	15 min.	60 min.	120 min.	Diário	Semanal
Média	0,00008	0,0003	0,0004628	0,00047	0,002361
Desvio-padrão	0,00350	0,0067	0,0087056	0,010562	0,025464
Assimetria	2,07798	0,9037	0,9797082	-0,720195	0,432551
Curtose	101,10366	34,5214	25,784635	12,04787	5,704347
Jarque-Bera	2.266.981	61.224	19.895	3.515	67,182
Prob.	0,000	0,000	0,000	0,000	0,000
Dickey-Fuller Aumentado	-13,616	-7,300	-7,699	-14,886	-13,804
Prob.	0,000	0,000	0,000	0,000	0,000
ARCH LM	58,574	23,571	9,757	174,954	25,178
Prob.	0,000	0,000	0,001	0,000	0,000
N. Observações	5.643	1.474	913	1.005	200
Período observado	01/01/2002 a 01/01/2003			01/01/2000 a 01/01/2004	

A tabela 4.2 traz os resultados do teste BDS para presença de não-linearidade nas séries de retorno da taxa de câmbio Real/Dólar. O teste BDS é tradicionalmente realizado com base numa série de resíduos obtidos a partir da estimativa de um modelo AR(1) entretanto, caso essa série de resíduos apresente efeitos ARCH, é possível estar cometendo um erro do tipo I, já que a não-linearidade pode estar “mascarada” pela heteroscedasticidade condicional autoregressiva dos resíduos (Abhyankar *et al.*, 1995). Dessa forma, o teste BDS foi realizado nas séries de resíduos obtidos a partir da estimativa de um modelo AR(1)-GARCH(1,1), com o objetivo de “filtrar” o efeito ARCH.

A hipótese alternativa de presença de não-linearidade foi aceita em todas as séries analisadas. Deve-se destacar, entretanto, que quanto maior é a frequência com que a série é observada, maior é a sua não-linearidade. Isso fica evidente quando se compara o

escore Z na série com frequência de 15 min. com o escore Z da série semanal – quanto maior é a frequência da série, maior é a rejeição da hipótese nula de linearidade (com exceção da série diária).

Tabela 4.2 Teste BDS para presença de não-linearidade nas séries de retorno da taxa de câmbio Real/Dólar.

15 min.			
Dimensão	Estat. BDS	Escore Z	Prob.
2	0,025	16,643	0,000
3	0,049	20,906	0,000
4	0,072	25,187	0,000
5	0,085	28,798	0,000
6	0,092	32,038	0,000
60 min.			
Dimensão	Estat. BDS	Escore Z	Prob.
2	0,032	11,510	0,000
3	0,061	13,586	0,000
4	0,082	15,289	0,000
5	0,096	17,230	0,000
6	0,102	18,953	0,000
120 min.			
Dimensão	Estat. BDS	Escore Z	Prob.
2	0,034	9,572	0,000
3	0,067	11,820	0,000
4	0,090	13,314	0,000
5	0,106	14,968	0,000
6	0,113	16,504	0,000
Diária			
Dimensão	Estat. BDS	Escore Z	Prob.
2	0,036	11,986	0,000
3	0,074	15,391	0,000
4	0,103	17,971	0,000
5	0,124	20,811	0,000
6	0,136	23,512	0,000
Semanal			
Dimensão	Estat. BDS	Escore Z	Prob.
2	0,032	4,292	0,000
3	0,066	5,484	0,000
4	0,085	5,851	0,000
5	0,091	6,010	0,000
6	0,098	6,640	0,000

4.2 Estimação dos modelos

Os modelos utilizados neste trabalho foram estimados utilizando 80% dos dados disponíveis em cada série histórica. Os 20% restantes foram utilizados para previsão e validação fora-da-amostra dos resultados obtidos na fase de estimação. Dessa forma, os resultados apresentados referentes à performance de previsão dos modelos e avaliação da estratégia de negociação referem-se sempre aos 20% últimos dados das séries. Esse tipo de procedimento é comumente utilizado para verificar se os resultados (parâmetros) dos modelos obtidos na fase de estimação são válidos ou consistentes para serem

utilizados com dados que não fazem parte da amostra. A tabela 4.3 traz a distribuição dos dados entre as etapas de estimação e validação fora-da-amostra.

Tabela 4.3 Distribuição do número de observações de cada série histórica de acordo com as etapas de estimação e validação fora-da-amostra.

Série	15 min.	60 min.	120 min.	Diária	Semanal
Estimação	4.515	1.180	730	804	180
Validação	1.128	294	183	201	20
Total	5.643	1.474	913	1.005	200

Este trabalho utilizou o teste de previsão de Chow para avaliar a hipótese de que os parâmetros obtidos na fase de estimação podem ser utilizados para fazer previsões com dados fora-da-amostra. A idéia do teste de previsão de Chow é estimar dois modelos – um utilizando toda a amostra disponível e outro utilizando apenas o conjunto de dados da fase de estimação. Se houver uma grande diferença entre os resultados obtidos pelos dois modelos, significa que pode não haver uma relação de estabilidade entre o período de estimação e o período de validação, ou seja, houve uma mudança no processo gerador de dados. A hipótese nula é que existe uma relação de estabilidade entre os períodos e, portanto, o modelo é capaz de fazer previsões para dados fora-da-amostra. O teste de previsão de Chow é dado por

$$Chow = \frac{(SQR_T - SQR_E)/n_2}{SQR_E/(n_1 - k)} \sim F(n_2, n_1 - k), \quad (4.3)$$

onde SQR_T é o somatório dos quadrados dos resíduos obtido a partir da estimativa do modelo utilizando todo o conjunto de dados, SQR_E é o somatório dos quadrados dos resíduos obtido a partir da estimativa do modelo utilizando apenas os dados da fase de estimação, n_2 é o número total de observações, n_1 é o número de observações da fase de estimação e k é o número de parâmetros estimados.

A tabela 4.4 traz o resultado do teste de previsão de Chow obtido a partir da estimativa de um modelo AR(1) para todas as séries de retorno da taxa de câmbio, de acordo com a divisão entre dados de estimação e validação apresentada na tabela 4.3. A hipótese nula de que existe uma relação de estabilidade entre a fase de estimação e a fase de validação foi aceita em todas as frequências observadas. Esse resultado nos diz que não há uma mudança significativa no processo gerador de dados entre os períodos, de modo que os modelos podem fazer previsões com dados fora-da-amostra.

Tabela 4.4 Resultados do teste de previsão de Chow obtidos a partir da estimativa de um modelo AR(1) para as séries de retorno da taxa de câmbio Real/Dólar.

Série	15 min.	60 min.	120 min.	Diária	Semanal
Estat. F	1,055	1,082	0,757	0,479	0,731
Prob.	0,138	0,215	0,977	0,999	0,881

Por último, é importante salientar que a avaliação da performance dos modelos se deu sobre previsões um-passo-a-frente, ou seja, avaliou-se no período t a capacidade de cada modelo em fazer previsões sobre o período $t+1$.

4.2.1 Modelos lineares

A estimação e avaliação do desempenho preditivo dos modelos lineares deu-se da seguinte forma. Foram inicialmente estimados e avaliados os modelos AR(1), MA(1) e ARMA(1,1). Em seguida, foram estimados e avaliados os modelos AR, MA, ARMA e ARMA-GARCH que combinassem menor AIC (mais parcimônia), menor RMSE (menor erro de previsão) e menor índice U-Theil (série prevista mais similar à série original). Para alcançar esse objetivo, utilizou-se inicialmente o correlograma das séries (função autocorrelação e função autocorrelação parcial) para estimar modelos com um grande número de defasagens em seus termos. A partir de sucessivas simulações foram eliminados os termos que não traziam contribuição em termos de performance preditiva, até que fosse encontrado o modelo mais parcimonioso que maximizasse a capacidade de previsão¹¹. Todos os modelos foram estimados para fornecerem previsões um-passo-a-frente.

4.2.2 Modelos não-lineares

Em função de sua maior complexidade computacional, o procedimento de estimação dos modelos não-lineares utilizados neste trabalho difere do processo de estimação dos modelos lineares apresentados na seção 4.2.1. A maioria dos pacotes econométricos ainda não disponibiliza a estimação de modelos não-lineares de um modo *user friendly*, cabendo ao usuário a implementação das rotinas computacionais em algum *software* ou linguagem computacional, o que aumenta significativamente o tempo de obtenção, análise e tabulação dos resultados. Utilizou-se neste trabalho o

¹¹ Neste trabalho, utilizou-se seguinte convenção: um modelo AR(2) inclui apenas a segunda defasagem da série observada, enquanto que um modelo AR(1-3) inclui a primeira, a segunda e a terceira defasagem da série. Esta mesma convenção vale para os modelos MA, ARMA e ARMA-GARCH.

software MATLAB® para implementação das rotinas computacionais de estimação dos modelos.

As representações matemáticas NARX e NARMAX foram utilizadas para a estimação dos modelos não-lineares. Dessa forma, a partir da definição da representação matemática dos modelos, sucessivas simulações do tipo *loops* (sucessivas iterações) foram feitas variando-se os parâmetros livres de cada modelo, até que fosse encontrado o melhor resultado. No caso das RNs-MLP, as simulações foram feitas variando-se o número de neurônios na camada oculta. No caso das RNs-RBF, as simulações foram feitas variando-se o número de centros pertencentes a cada função de base radial e a sua área de atuação (abertura). Finalmente, para o sistema nebuloso TS, as simulações foram feitas variando-se o número de funções de pertinência.

As RNs-MLP foram estimadas considerando uma camada de entrada (neurônios sem funções de ativação), uma camada oculta (neurônios com funções de ativação do tipo sigmóide) e uma camada de saída (um neurônio com função de ativação linear). O treinamento foi realizado através o algoritmo de segunda ordem Levenberg-Marquardt. Para cada *loop* realizado pela RN na busca do melhor resultado, o número de épocas de treinamento foi fixado em dez. Em geral, o nível de erro mínimo entre a saída desejada e a saída estimada da RN era atingido a partir da quinta época de treinamento.

As RNs-RBF foram estimadas considerando uma camada de entrada (neurônios sem funções de base radial), uma camada oculta (neurônios com funções de base radial do tipo gaussianas) e uma camada de saída (um neurônio que faz a estimativa de mínimos quadrados dos pesos da RN). A estratégia de treinamento da rede foi a técnica do algoritmo *k-médias* e o método de mínimos quadrados em batelada, demonstrada anteriormente. Os sistemas nebulosos TS (SN-TS) foram estimados utilizando o algoritmo de agrupamento de dados ou clusterização nebulosa Gustafson-Kessel (Gustafson e Kessel, 1979) para a determinação dos *clusters* nebulosos no conjunto de dados e a técnica de mínimos quadrados para a estimativa dos parâmetros dos consequentes de cada regra.

4.3 Apresentação dos resultados

A tabela 4.5 traz os resultados obtidos a partir da estimativa dos modelos lineares e não-lineares de previsão para a série de retornos da taxa de câmbio com frequência de 15 min. Dentre os modelos lineares estimados, o modelo ARMA-GARCH foi o que conseguiu o menor valor para o índice U-Theil (0,867), ou seja, gerou a série prevista mais semelhante à série observada. Esse modelo conseguiu prever corretamente 44% da

direção dos movimentos da taxa de câmbio, entretanto, o valor da estatística P.T. indica que não houve significância estatística nessas previsões. Além disso, todos os modelos lineares apresentaram o mesmo RMSE (0,005), apesar de estimados com diferentes termos e defasagens, indicando que esse critério de avaliação foi pouco útil para escolher qual é o melhor modelo de previsão.

Dentre os modelos não-lineares, aquele de melhor desempenho foi o modelo RN-MLP-NARMAX, que obteve um índice U-Theil de 0,786 e conseguiu prever corretamente 54% da direção dos movimentos da taxa de câmbio. O valor da estatística P.T. indica que houve significância estatística nas previsões realizadas por esse modelo.

Tabela 4.5 Resultados dos modelos lineares e não-lineares de previsão para a série da taxa de câmbio Real/Dólar com frequência de 15 min.

MODELO	AIC	RMSE	U-THEIL	S.C.P.	P.T.
Modelos Lineares					
AR(1)	-8,742	0,005	0,966	43%	-1,92
MA(1)	-8,742	0,005	0,967	43%	-1,92
ARMA(1,1)	-8,743	0,005	0,956	44%	-2,72
AR(3-6)	-8,746	0,005	0,931	44%	-3,08
MA(3-6)	-8,746	0,005	0,930	44%	-3,01
ARMA(3-6,3-6)	-8,750	0,005	0,894	42%	-5,35
ARMA(1-2,1-2)-GARCH(1-2,1-2)	-9,300	0,005	0,867	44%	-4,74
Modelos Não-Lineares					
RN-MLP-NARX	-9,337	0,005	0,783	53%	2,19*
RN-MLP-NARMAX	-9,337	0,005	0,786	54%	3,62*
RN-RBF-NARX	-8,728	0,005	0,828	51%	1,91
RN-RBF-NARMAX	-8,728	0,005	0,815	48%	1,56
SN-TS-NARX	-29,551	0,03	0,957	54%	1,14
SN-TS-NARMAX	-29,551	0,04	0,940	51%	0,03

* Estatisticamente significativa a 95% de confiança.

No caso da série da taxa de câmbio com frequência de 15 min., todos os modelos lineares apresentaram um baixo poder de previsão. O índice U-Theil esteve próximo a 1 em todos os modelos, indicando pouca similaridade da série prevista com a série original, enquanto que o percentual de sinais corretamente previstos esteve em torno de apenas 44%. A estatística P.T. não foi significativa em nenhum dos casos¹², mostrando que os modelos estimados não conseguiram prever de modo estatisticamente consistente a direção do movimento da série observada. Os modelos não-lineares conseguiram um desempenho superior, com valores inferiores do índice U-Theil, percentuais de S.C.P. superiores e estatisticamente significantes no caso do modelo RN-MLP.

¹² A estatística P.T. é passível de assumir valores negativos. Isso acontece quando o valor observado dos sinais corretamente previstos é inferior ao seu valor esperado, indicando que o modelo possui um baixo poder de previsão. Para que o modelo seja um bom previsor, a estatística P.T. deve ser superior a 1,96 (valor tabelado a 95% de confiança).

Os resultados dos modelos de previsão para a série da taxa de câmbio com frequência de 60 min. são apresentados na tabela 4.6. Todos os modelos retornaram o mesmo erro de previsão (RMSE igual a 0,01), entretanto foram os modelos AR(5-9) e ARMA(8-9,5-9) que conseguiram o maior número de sinais corretamente previstos (51%). Esse resultado não foi estatisticamente significativo, como mostra a estatística P.T. Os índices U-Theil obtidos (todos acima de 0,8) indicam que as séries previstas têm pouca similaridade com as séries observadas. Pode-se concluir dessa forma que, no caso da série com frequência de 60 min., os modelos lineares não obtiveram um bom desempenho preditivo.

Dentre os modelos não-lineares, o modelo RN-MLP-NARMAX foi o que obteve o menor valor para o índice U-Theil (0,640) e o maior número de sinais corretamente previstos (60%). Ao contrário dos modelos lineares, todos os modelos não-lineares obtiveram significância estatística em suas previsões. Com relação ao tamanho do erro de previsão, todos os modelos não-lineares (com exceção do modelo SN-TS) obtiveram um RMSE inferior aos modelos lineares.

Tabela 4.6 Resultados dos modelos lineares e não-lineares de previsão para a série da taxa de câmbio Real/Dólar com frequência de 60 min.

MODELO	AIC	RMSE	U-THEIL	S.C.P.	P.T.
Modelos Lineares					
AR(1)	-7,52	0,01	0,974	45%	-1,09
MA(1)	-7,52	0,01	0,975	45%	-1,09
ARMA(1,1)	-7,52	0,01	0,93	50%	0,63
AR(5-9)	-7,54	0,01	0,843	51%	0,67
MA(5-9)	-7,55	0,01	0,830	50%	0,19
ARMA(8-9,5-9)	-7,55	0,01	0,817	51%	0,62
ARMA(8-9,5-9)-GARCH(1-4,1-4)	-8,17	0,01	0,877	49%	-0,47
Modelos Não-Lineares					
RN-MLP-NARX	-9,62	0,001	0,652	59%	3,56*
RN-MLP-NARMAX	-9,61	0,001	0,640	60%	3,31*
RN-RBF-NARX	-9,23	0,001	0,740	59%	3,10*
RN-RBF-NARMAX	-9,23	0,001	0,725	60%	3,21*
SN-TS-NARX	-2,12	0,07	0,953	54%	2,69*
SN-TS-NARMAX	-2,12	0,07	0,900	55%	2,03*

* Estatisticamente significativa a 95% de confiança.

A tabela 4.7 traz os resultados dos modelos de previsão para a série da taxa de câmbio com frequência de 120 min. Dentre os modelos lineares, o modelo AR(1,3,5,6,9-11) foi o que conseguiu maximizar o número de sinais corretamente previstos (51%). Dado que a estatística P.T. foi de apenas 0,67 para esse resultado, não houve significância estatística quanto ao número de sinais corretamente previstos. Apesar da diferença de termos e defasagens nos modelos, o RMSE foi o mesmo para

todos os modelos. A análise do índice U-Theil (todos acima de 0,8) indica que os modelos lineares não conseguiram fazer boas previsões para a série da taxa de câmbio com frequência de 120 min., resultado corroborado pelo valor estatisticamente não significativo da estatística P.T.

Os modelos não-lineares, por sua vez, conseguiram um desempenho preditivo superior aos modelos lineares. O número de sinais corretamente previstos foi superior em todos os casos, de maneira estatisticamente significativa. O modelo RN-MLP-NARX obteve o melhor desempenho, prevendo corretamente 61% dos movimentos da taxa de câmbio e atingindo o menor índice de similaridade U-Theil (0,617).

Tabela 4.7 Resultados dos modelos lineares e não-lineares de previsão para a série da taxa de câmbio Real/Dólar com frequência de 120 min.

MODELO	AIC	RMSE	U-THEIL	S.C.P.	P.T.
Modelos Lineares					
AR(1)	-7,058	0,013	0,890	50%	0,17
MA(1)	-7,059	0,013	0,886	50%	0,02
ARMA(1,1)	-7,060	0,013	0,871	49%	-0,17
AR(1,3,5,6,9-11)	-7,080	0,013	0,782	55%	1,46
MA(1,3,5,6,9-11)	-7,100	0,013	0,762	54%	1,01
ARMA(1,3,5,6,9-11)	-7,110	0,013	0,735	49%	-0,20
ARMA(1,3,5,6,9-11)-GARCH(2,2)	-7,668	0,013	0,844	53%	0,70
Modelos Não-Lineares					
RN-MLP-NARX	-8,65	0,01	0,617	63%	3,74*
RN-MLP-NARMAX	-8,65	0,01	0,637	61%	3,19*
RN-RBF-NARX	-7,89	0,01	0,753	58%	2,42*
RN-RBF-NARMAX	-7,89	0,01	0,714	58%	2,32*
SN-TS-NARX	-1,55	0,09	0,938	60%	3,79*
SN-TS-NARMAX	-1,55	0,09	0,837	55%	2,17*

* Estatisticamente significativa a 95% de confiança.

Os resultados dos modelos de previsão para a série da taxa de câmbio com frequência diária são mostrados na tabela 4.8. O modelo MA(1) foi o que conseguiu melhor desempenho, prevendo corretamente 56% dos sinais. O valor da estatística P.T. de 2,41 mostra que esse resultado é estatisticamente significativo. Além do modelo MA(1), os modelos ARMA(1,1) e AR(1,2,5,7,8,10) também conseguiram um valor estatisticamente significativo para suas previsões. Dentre os modelos não-lineares, o modelo RN-MLP-NARX conseguiu um índice de similaridade U-Theil de 0,7 e um percentual de sinais corretamente previstos de 61%. Com exceção do modelo SN-TS-NARX, todos os modelos não-lineares conseguiram significância estatística em suas previsões.

Tabela 4.8 Resultados dos modelos lineares e não-lineares de previsão para a série da taxa de câmbio Real/Dólar com frequência diária.

MODELO	AIC	RMSE	U-THEIL	S.C.P.	P.T.
Modelos Lineares					
AR(1)	-6,182	0,007	0,829	54%	1,55
MA(1)	-6,193	0,007	0,787	56%	2,41*
ARMA(1,1)	-6,198	0,008	0,798	54%	2,11*
AR(1,2,5,7,8,10)	-6,200	0,007	0,781	52%	1,96*
MA(1,2,4,5,7)	-6,220	0,008	0,787	51%	1,49
AR(1,2,5,7,8,10) MA(1,2,4,5,7)	-6,210	0,008	0,771	47%	0,19
AR(1,2,7)MA(1,2,4,5,7)-GARCH(2,2)	-6,740	0,008	0,867	54%	1,55
Modelos Não-Lineares					
RN-MLP-NARX	-20,56	0,001	0,700	61%	3,85*
RN-MLP-NARMAX	-19,25	0,001	0,701	58%	3,21*
RN-RBF-NARX	-15,24	0,001	0,821	60%	3,14*
RN-RBF-NARMAX	-8,65	0,01	0,732	58%	2,42*
SN-TS-NARX	2,36	0,10	0,953	57%	1,17
SN-TS-NARMAX	2,36	0,10	0,772	59%	2,34*

* Estatisticamente significativa a 95% de confiança.

A tabela 4.9 traz os resultados dos modelos lineares de previsão para a série da taxa de câmbio com frequência semanal. A análise da tabela indica que todos os modelos geraram o mesmo erro de previsão (RMSE igual a 0,013), entretanto o modelo MA(3-5,8) foi o que conseguiu melhor desempenho, prevendo corretamente 67% do total de sinais que indicam a direção do movimento da série observada. O modelo ARMA(3-5,8) obteve o segundo melhor desempenho, prevendo corretamente 62% dos sinais. Ambos os modelos obtiveram significância estatística em suas previsões. O valor do índice U-Theil de 0,654 para o modelo MA(3-5,8) e de 0,624 corrobora esse resultado, indicando que esses modelos superaram os demais na tarefa de fazer uma boa previsão para série da taxa de câmbio com frequência semanal.

O modelo não-linear de melhor desempenho foi o modelo RN-MLP-NARMAX, que obteve um índice U-Theil de 0,458 e conseguiu prever corretamente 82% da direção dos movimentos da taxa de câmbio, ou seja, obteve um desempenho preditivo largamente superior ao melhor modelo linear de previsão. Todos os modelos não-lineares conseguiram prever corretamente mais de 60% dos movimentos da taxa de câmbio e, com exceção do modelo RN-RBF-NARMAX, todos obtiveram significância estatística em suas previsões.

Tabela 4.9 Resultados dos modelos lineares e não-lineares de previsão para a série da taxa de câmbio Real/Dólar com frequência semanal.

MODELO	AIC	RMSE	U-THEIL	S.C.P.	P.T.
Modelos Lineares					
AR(1)	-4,424	0,01	0,883	38%	-0,01
MA(1)	-4,430	0,01	0,882	38%	-0,01
ARMA(1,1)	-4,417	0,01	0,880	38%	-0,01
AR(3-5,8)	-4,412	0,01	0,681	57%	1,78
MA(3-5,8)	-4,466	0,01	0,654	67%	2,33*
ARMA(3-5,8)	-4,417	0,01	0,624	62%	2,06*
ARMA(3-5,8)-GARCH(2,2)	-4,660	0,01	0,658	57%	0,74
Modelos Não-Lineares					
RN-MLP-NARX	-4,86	0,02	0,469	75%	3,27*
RN-MLP-NARMAX	-4,86	0,02	0,458	82%	4,32*
RN-RBF-NARX	-3,87	0,02	0,698	62%	2,11*
RN-RBF-NARMAX	-2,65	0,13	0,796	65%	1,50
SN-TS-NARX	-1,55	0,13	0,910	62%	2,48*
SN-TS-NARMAX	-1,55	0,13	0,904	60%	2,00*

* Estatisticamente significativa a 95% de confiança.

A análise dos resultados dos modelos de previsão para as séries de retornos da taxa de câmbio permite extrair algumas conclusões importantes. Em primeiro lugar, pode-se concluir que os modelos não-lineares obtiveram um desempenho preditivo superior aos modelos lineares em todas as séries analisadas. O modelo não-linear NN-MLP foi o que obteve melhor desempenho em todas as séries. As evidências empíricas encontradas neste trabalho vêm, portanto, a corroborar com os estudos anteriores que atestam a superioridade dos modelos não-lineares na previsão de séries temporais financeiras.

Em segundo lugar, ficou claro que a qualidade da previsão feita tanto pelos modelos lineares quanto os não-lineares está relacionada com a frequência com que a série é observada. Os resultados aqui apresentados (sintetizados na tabela 4.10) mostram que quanto menor é a frequência com que a série é observada maior é a capacidade do modelo em fazer previsões.

Esse fato fica evidente pela análise do percentual de sinais corretamente previstos do modelo de melhor desempenho (tabela 4.10), que evoluiu de 54% na série de 15 min. para 82% na série semanal. O índice U-Theil, por sua vez, foi de 0,786 para a série de 15 min. e 0,458 para a série semanal, mostrando que quanto maior é a frequência da série, maior é a semelhança entre a série prevista pelo modelo e a série observada.

Tabela 4.10 Resumo dos resultados dos melhores modelos lineares de previsão para a série de retornos da taxa de câmbio Real/Dólar de acordo com a frequência observada.

Série	15 min.	60 min.	120 min.	Diária	Semanal
Modelo de melhor desempenho	RN-MLP-NARMAX	RN-MLP-NARMAX	RN-MLP-NARX	RN-MLP-NARMAX	RN-MLP-NARMAX
U-Theil	0,786	0,640	0,617	0,700	0,458
S.C.P.	54%	60%	63%	61%	82%
Estat. P.T.	3,62*	3,31*	3,74*	3,85*	4,32*

* Estatisticamente significante a 95% de confiança.

Uma possível explicação para esse resultado está relacionada à presença de não-linearidade nas séries. Os resultados da tabela 4.2 mostram que quanto maior é a frequência com que a série é observada, maior é a presença de não-linearidade. Dessa forma, a não-linearidade das séries pode estar influenciando diretamente a qualidade da previsão, em função de que a não-linearidade significa um processo gerador de dados da série mais complexo e, por conseqüência, menos previsível. Nessa perspectiva, o desempenho superior dos modelos não-lineares pode ser justificado pela sua maior capacidade em capturar não-linearidades presentes nas séries analisadas.

Dentre os modelos não-lineares analisados, aquele que obteve o melhor desempenho geral foi o modelo RN-MLP. É interessante destacar que, graças ao algoritmo utilizado, o tempo de treinamento e obtenção dos resultados foi semelhante ao tempo de obtenção de resultados usando uma rede RN-RBF. Cada *loop* (período de treinamento com 10 épocas, no caso da RN-MLP) foi efetuado num tempo médio de 15 seg., 11 seg., 8 seg., 6 seg., e 4 seg. para as séries com frequências de 15 min., 60 min., 120 min., diária e semanal, respectivamente, utilizando-se um microcomputador PC doméstico. O modelo não-linear SN-TS foi, dentre todos os modelos não-lineares, o que levou mais tempo para gerar os resultados.

Por último, o tamanho do erro de previsão (RMSE) mostrou-se um critério pouco útil para escolher qual é o modelo de melhor desempenho. No caso, por exemplo, das séries com frequência de 15 min. e semanal, o RMSE dos modelos lineares foi maior que o RMSE dos modelos não-lineares. Esse resultado corrobora o estudo de Leitch e Tanner (1991), os quais concluíram que, além do tamanho erro de previsão, outros critérios devem ser levados em conta na escolha de qual é o melhor modelo de previsão.

Os resultados encontrados neste trabalho corroboram com outras evidências empíricas anteriores, como, por exemplo, os trabalhos de Kaashoek e van Dijk (2002), Kuan e Liu (1995) e Boero e Marrocu (2002). De modo semelhante, esses trabalhos

concluíram que a abordagem não-linear foi superior à abordagem linear na previsão de taxas de câmbio.

4.3.1 Estratégia de negociação

Esta seção apresenta os resultados da estratégia de negociação estabelecida com base nas previsões geradas pelos modelos. A motivação para a análise de uma estratégia de negociação em paralelo à avaliação estatística da previsão vem do trabalho de Leitch e Tanner (1991), os quais argumentam que o critério mais adequado para avaliar a previsão de um modelo é a sua capacidade de gerar lucros, e não uma medida indireta feita através do tamanho do erro de previsão.

O foco da estratégia consiste na capacidade dos modelos em prever de forma correta a direção do movimento dos preços, ou seja, prever quando o retorno no próximo período será negativo ou positivo. Sendo assim, uma estratégia é lucrativa quando ela consegue prever corretamente o maior número de vezes em que o retorno será positivo ou negativo. Além disso, considera-se a possibilidade de se fazer venda a descoberto, de modo que se possa obter lucros também nos movimentos de queda do dólar. Com relação aos custos de transação, assume-se que esses custos estejam embutidos no *spread* entre a compra e a venda. A estratégia pode ser resumida nas seguintes regras:

- Regra 1: Se o valor previsto para o próximo período for positivo, computa-se uma compra;
- Regra 2: Se o valor previsto para o próximo período for negativo, computa-se uma venda.

A rentabilidade obtida pela estratégia de negociação será comparada com a rentabilidade de uma estratégia passiva do tipo *buy-and-hold* (comprar e reter), que consiste numa estratégia simplista de negociação em que o investidor não lança mão de técnicas de operações que busquem otimizar os pontos de entrada e saída do mercado. Além disso, será calculado o índice de Sharpe de modo que possa comparar o retorno obtido com o risco (desvio-padrão) assumido em cada estratégia. O índice de Sharpe (IS) pode ser definido como

$$IS = \frac{E(R_i)}{\sigma_i} \quad (4.6)$$

onde $E(R_i)$ é o retorno esperado do investimento i e σ_i é o desvio padrão dos retornos do investimento. Quanto maior for o valor do IS, melhor é a estratégia de negociação.

A tabela 4.11 traz os resultados das estratégias de negociação para a série da taxa de câmbio Real/Dólar com frequência de 15 min. Nenhuma estratégia baseada nos modelos lineares conseguiu superar a estratégia *buy-and-hold*, a qual obteve um retorno de 13,01% um IS de 3,26. A estratégia baseada no modelo AR(3-6) foi, dentre os modelos lineares a que obteve melhor desempenho, com um retorno de 12,55% e índice IS de 3,20.

Todas as estratégias baseadas em modelos não-lineares, com exceção dos modelos RN-RBF-NARX e SN-TS-NARMAX, conseguiram retornos e IS superiores à estratégia *buy-and-hold*. A estratégia baseada no modelo RN-MLP-NARMAX foi a que obteve o maior retorno (48,05%). Entretanto, foi a estratégia baseada no modelo RN-MLP-NARX que obteve o maior IS (4,52).

Pode-se afirmar que, no que se refere à série de retornos da taxa de câmbio com frequência de 15 min, as estratégias baseadas nos modelos lineares não superaram - em termos de risco e retorno - uma estratégia simplista de negociação. Alguns dos modelos não-lineares analisados, entretanto, conseguiram superar a estratégia *buy-and-hold* e todas as outras estratégias baseadas em modelos lineares.

Tabela 4.11 Rentabilidade acumulada, desvio padrão e índice de Sharpe das estratégias de negociação aplicadas à série de retornos da taxa de câmbio Real/Dólar com frequência de 15 min.

Estratégia	Retorno acumulado	Desvio padrão	Índice de Sharpe
<i>Buy-and-hold</i>	13,01%	3,98%	3,26
Modelos Lineares			
AR(1)	-0,69%	4,34%	-0,16
MA(1)	-0,69%	4,34%	-0,16
ARMA(1,1)	0,02%	2,60%	0,62
AR(3-6)	12,55%	3,92%	3,20
MA(3-6)	8,12%	3,76%	2,16
ARMA(3-6,3-6)	-13,13%	5,08%	-2,58
ARMA(1-2,1-2)-GARCH(1-2,1-2)	-7,08	3,85%	-1,84
Modelos Não-lineares			
RN-MLP-NARX	30,73%	6,81%	4,52
RN-MLP-NARMAX	48,05%	11,62%	4,13
RN-RBF-NARX	10,30%	5,53%	1,86
RN-RBF-NARMAX	21,47%	5,11%	4,20
SN-TS-NARX	31,20%	9,41%	3,31
SN-TS-NARMAX	21,27%	6,85%	3,11

A tabela 4.12 traz os resultados das estratégias de negociação aplicadas à série de retornos da taxa de câmbio com frequência de 60 min. A estratégia *buy-and-hold* obteve

um retorno de 13,6% e um IS de 2,6. Ao contrário do observado na série com frequência de 15 min., os modelos lineares ARMA(1,1), AR(5-9) e MA(5-9) conseguiram uma rentabilidade superior à estratégia passiva (36,55%, 35,10% e 19,05%, respectivamente). Esses modelos também alcançaram uma melhor relação risco-retorno, medida pelo IS (4,21, 4,00, e 2,76, respectivamente).

No caso dos modelos não-lineares, todos eles superaram – em termos de retorno e risco – a estratégia simplista de negociação. O modelo RN-MLP-NARMAX atingiu a maior rentabilidade (92,76%), entretanto foi também o modelo que atingiu maior desvio-padrão (28,21%), resultando num IS de 3,28. Dentre todos os modelo não-lineares, foi o modelo SN-TS-NARMAX que conseguiu o maior IS (4,06). Todos os modelos não-lineares conseguiram rentabilidades acima de 30%, bem acima da rentabilidade atingida pela estratégia *buy-and-hold*.

Tabela 4.12 Rentabilidade acumulada, desvio padrão e índice de Sharpe das estratégias de negociação aplicadas à série de retornos da taxa de câmbio Real/Dólar com frequência de 60 min.

Estratégia	Retorno acumulado	Desvio padrão	Índice de Sharpe
<i>Buy-and-hold</i>	13,60%	5,23%	2,60
Modelos Lineares			
AR(1)	12,35%	5,16%	2,40
MA(1)	12,35%	5,16%	2,40
ARMA(1,1)	36,55%	8,67%	4,21
AR(5-9)	35,10%	8,77%	4,00
MA(5-9)	19,05%	6,89%	2,76
ARMA(8-9,5-9)	6,17%	4,61%	1,34
ARMA(8-9,5-9)-GARCH(1-4,1-4)	5,84%	4,28%	1,36
Modelos Não-Lineares			
RN-MLP-NARX	83,75%	20,9%	4,01
RN-MLP-NARMAX	92,76%	28,21%	3,28
RN-RBF-NARX	53,38%	13,66%	3,91
RN-RBF-NARMAX	83,15%	24,63%	3,37
SN-TS-NARX	46,56%	12,44%	3,74
SN-TS-NARMAX	30,04%	7,4%	4,06

Os resultados das estratégias de negociação aplicadas à série de retornos da taxa de câmbio com frequência de 120 min. são apresentados na tabela 4.13. Com exceção do modelo ARMA(1,3,5,6,9-11), todos os modelos lineares conseguiram uma rentabilidade acumulada superior à estratégia passiva. O modelo MA(1) foi o que obteve melhor desempenho, com uma rentabilidade de 25,18% e um IS de 4,03, ao passo que a estratégia passiva obteve um retorno de 12,4% e um IS de 2,54. Apesar de o modelo AR(1,3,5,6,9-11) ter obtido a maior rentabilidade, o desvio padrão obtido foi também o mais elevado (13,21%), resultando num IS de 2,12.

Todos os modelos não-lineares obtiveram rentabilidades superiores à estratégia passiva de negociação e às estratégias baseadas em modelos lineares. De forma semelhante ao ocorrido na série com frequência de 60 min., o modelo RN-MLP-NARMAX conseguiu a maior rentabilidade e maior desvio-padrão (80,25% e 24,3%. Respectivamente), entretanto foi o modelo SN-TS-NARMAX que conseguiu o maior IS (4,07).

Tabela 4.13 Rentabilidade acumulada, desvio padrão e índice de Sharpe das estratégias de negociação aplicadas à série de retornos da taxa de câmbio Real/Dólar com frequência de 120 min.

Estratégia	Retorno acumulado	Desvio padrão	Índice de Sharpe
<i>Buy-and-hold</i>	12,40%	4,88%	2,54
Modelos Lineares			
AR(1)	19,13%	4,86%	3,94
MA(1)	25,17%	6,24%	4,03
ARMA(1,1)	21,82%	8,01%	2,70
AR(1,3,5,6,9-11)	28,02%	13,21%	2,12
MA(1,3,5,6,9-11)	20,66%	12,00%	1,72
ARMA(1,3,5,6,9-11)	8,24%	7,76%	1,06
ARMA(1,3,5,6,9-11)-GARCH(2,2)	16,44%	9,36%	1,76
Modelos Não-Lineares			
RN-MLP-NARX	72,31%	22,51%	3,21
RN-MLP-NARMAX	80,25%	24,30%	3,30
RN-RBF-NARX	53,03%	13,68%	3,87
RN-RBF-NARMAX	40%	10,38%	3,85
SN-TS-NARX	49%	12,13%	4,03
SN-TS-NARMAX	41,1%	10%	4,07

A tabela 4.14 apresenta os resultados das estratégias de negociação aplicadas à série de retornos da taxa de câmbio com frequência diária. Todos os modelos lineares superaram em termos de risco e retorno a estratégia passiva de investimento, a qual acumulou uma rentabilidade de -18,11% e um IS de -4,27. O modelo AR(1) foi o que obteve melhor desempenho, com um retorno de 36,98% e IS de 3,40.

Assim como os modelos lineares, todos os modelos não-lineares superaram a estratégia simplista de negociação. A maior lucratividade foi alcançada pelo modelo RN-MLP-NARMAX (50%), ao passo que o maior IS foi alcançado pelo modelo SN-TS-NARX (4,1%).

Tabela 4.14 Rentabilidade acumulada, desvio padrão e índice de Sharpe das estratégias de negociação aplicadas à série de retornos da taxa de câmbio Real/Dólar com frequência diária.

Estratégia / Modelo	Retorno acumulado	Desvio padrão	Índice de Sharpe
<i>Buy-and-hold</i>	-18,11%	-4,24%	-4,27
Modelos Lineares			
AR(1)	36,98%	10,23%	3,40
MA(1)	34,62%	11,30%	3,06
ARMA(1,1)	28,87%	9,61%	3,00
AR(1,2,5,7,8,10)	22,68%	7,62%	2,98
MA(1,2,4,5,7)	28,37%	8,50%	3,34
ARMA(1,2,4,5,7)	-12,33%	5,94%	-2,08
ARMA(1,2,4,5,7)-GARCH(2,2)	4,38%	4,51%	0,99
Modelos Não-Lineares			
RN-MLP-NARX	49,42%	15,77%	3,14
RN-MLP-NARMAX	50%	15,77%	3,13
RN-RBF-NARX	45%	12,41%	3,62
RN-RBF-NARMAX	38,12%	10,43%	3,65
SN-TS-NARX	25,47%	6,21%	4,10
SN-TS-NARMAX	47,23%	14,06%	3,36

Os resultados das estratégias de negociação aplicadas à série de retornos da taxa de câmbio com frequência semanal são apresentados na tabela 4.15. A estratégia *buy-and-hold* acumulou uma rentabilidade de -6,20% e um IS de -4,44. O modelo linear de melhor desempenho foi o modelo ARMA-GARCH, com uma rentabilidade de 6,36% e IS igual a 2.

Tabela 4.15 Rentabilidade acumulada, desvio padrão e índice de Sharpe das estratégias de negociação aplicadas à série de retornos da taxa de câmbio Real/Dólar com frequência semanal.

Estratégia / Modelo	Retorno acumulado	Desvio padrão	Índice de Sharpe
<i>Buy-and-hold</i>	-6,20%	1,40%	-4,44
Modelos Lineares			
AR(1)	-6,20%	1,40%	-4,44
MA(1)	-6,20%	1,40%	-4,44
ARMA(1,1)	-6,20%	1,40%	-4,44
AR(3-5,8)	4,15%	2,93%	1,41
MA(3-5,8)	4,95%	3,48%	1,42
ARMA(3-5,8)	2,09%	2,57%	0,81
ARMA(3-5,8)-GARCH(2,2)	6,36%	3,20%	2,00
Modelos Não-Lineares			
RN-MLP-NARX	50%	15,78%	3,17
RN-MLP-NARMAX	55,26%	17,25%	3,20
RN-RBF-NARX	14,80%	3,81%	3,81
RN-RBF-NARMAX	22,13%	6,78%	3,26
SN-TS-NARX	21,93%	6,10%	3,60
SN-TS-NARMAX	17,47%	4,91%	3,55

Todos os modelos não-lineares conseguiram superar – em termos de risco e retorno – todas as outras estratégias de negociação. O modelo RN-MLP-NARMAX, por

exemplo, obteve uma rentabilidade de 55,26%, um desvio padrão de 17,25% e um IS de 3,2. O maior IS foi obtido pelo modelo RN-RBF-NARX (3,81%).

A análise dos resultados da estratégia de negociação permite extrair duas conclusões. Em primeiro lugar, é possível estabelecer uma ligação entre a avaliação estatística das previsões (seção 4.3.1) e a significância econômica das previsões (seção 4.3.3). De um modo geral, os modelos que obtiveram uma boa avaliação estatística conseguiram superar, em termos de retorno e risco, a estratégia passiva de investimento. Da mesma forma, aqueles modelos que não alcançaram uma significância estatística das suas previsões não conseguiram superar a estratégia passiva. Isso sugere que a utilização de múltiplos critérios estatísticos de avaliação do desempenho preditivo, em paralelo à análise de retorno e risco, é um procedimento válido para a escolha de modelos de previsão.

Em segundo lugar, é possível concluir que os modelos não-lineares obtiveram um desempenho superior – em termos de risco e retorno – em relação às outras estratégias de negociação analisadas. Em todas as séries observadas, houve um modelo não-linear que obteve rentabilidade e IS superiores à estratégia simplista de negociação e às estratégias baseadas em modelos lineares. Esse resultado corrobora com evidências empíricas anteriores, como os trabalhos de Franses e Griensven (2003), Andrada-Felix *et al.* (2003) e Boero e Marrocu (2002), os quais concluíram que estratégias de negociação baseadas em modelos não-lineares superaram em termos de risco e retorno estratégias de negociação baseadas em modelos lineares tradicionais.

4.4 Síntese conclusiva

Este capítulo teve como objetivo apresentar os resultados das previsões geradas pelos modelos lineares e não-lineares aplicadas às séries de retorno da taxa de câmbio brasileira Real/Dólar com frequências de 15 min., 60 min., 120 min., diária e semanal. Inicialmente, são mostrados os resultados de testes estatísticos de especificação que visam identificar características pertinentes às séries analisadas, como presença de não-linearidade, presença de efeitos ARCH, presença de raízes unitárias e aderência à curva Normal. Os resultados mostram que todas as séries são estacionárias e que possuem efeitos ARCH. Além disso, a hipótese de presença de não-linearidade foi aceita em todas as séries. Constatou-se que quanto maior é a frequência da série, maior é o seu grau de não-linearidade.

A avaliação do desempenho preditivo indicou que os modelos não-lineares superaram os modelos lineares. Em todas as séries analisadas, os modelos não-lineares conseguiram previsões mais acuradas que os modelos lineares. Além disso, constatou-se que a acurácia de previsão está intimamente ligada com a frequência observada da série. Quanto menor é a frequência da série, maior é a capacidade dos modelos em gerarem boas previsões. Nesse sentido, os maiores índices de sucesso nas previsões foram obtidas nas séries diária e semanal. Uma possível explicação para esse fato pode estar relacionada à presença de não-linearidades nas séries, ou seja, quanto menor é a presença de não-linearidade, maior é a capacidade dos modelos em obterem sucesso em suas previsões.

Os resultados da estratégia de negociação estabelecida com base nas previsões geradas pelos modelos indicaram que os modelos não-lineares obtiveram um desempenho superior aos modelos lineares e à estratégia simplista de negociação. Além disso, foi possível estabelecer uma relação entre a significância estatística das previsões e sua significância econômica, mensurada pelo binômio retorno-risco. Todos os modelos que conseguiram significância estatística em suas previsões conseguiram também superar a estratégia passiva de investimento, indicando que os critérios de avaliação do desempenho preditivo utilizados são adequados para escolha dos melhores modelos de previsão.

5

Conclusões

O surgimento de novas metodologias econométricas que buscam identificar padrões ou regularidades no processo gerador de dados das séries temporais financeiras, aliado à crescente evidência empírica de presença de complexas estruturas não-lineares nestas séries tem, ao mesmo tempo, reacendido o debate acerca da previsibilidade dos retornos dos ativos financeiros e, posto em xeque o uso de modelos lineares tradicionais na descrição e previsão da dinâmica dos preços. Alguns autores, como, por exemplo, Aguirre e Aguirre (1997), argumentam que nesses casos a utilização de modelos não-lineares de previsão mostra-se mais apropriado, em função da impossibilidade dos modelos lineares em reproduzir de modo consistente a dinâmica das séries financeiras.

Pesquisadores da área de séries temporais têm procurado investigar o benefício da utilização de modelos não-lineares em detrimento de modelos lineares, no que se refere a sua acurácia de previsão. Enquanto que parcela dos estudos empíricos mostra que modelos não-lineares conseguem uma previsão fora-da-amostra mais acurada que modelos lineares (Kaashoek e van Dijk 2002; Andrada-Félix *et al.*, 2003; Kuan e Liu, 1995), outra parcela dos estudos chega à conclusão oposta (Stock e Watson, 1999; Neely e Weller, 2003; Swanson e White, 1995), tornando a questão controversa.

É dentro dessa linha de pesquisa que está situado este trabalho. A idéia central foi fornecer mais uma contribuição a essa questão ainda carente de consenso. Nesse sentido, o objetivo principal do trabalho foi avaliar a hipótese de que modelos não-lineares de previsão de séries temporais são capazes de gerar uma previsão fora-da-amostra mais precisa que modelos lineares tradicionais. Utilizaram-se séries de retorno (primeira diferença dos logaritmos dos preços) da taxa de câmbio Real/Dólar (R\$/US\$) coletadas nas frequências de 15 min., 60min. e 120 min. durante o período que vai de 01/01/2002 a 01/01/2003 e séries com frequência diária e semanal durante o período que vai de 01/01/2000 a 01/01/2004.

O desempenho preditivo dos modelos foi avaliado segundo o tamanho do erro de previsão, mensurado pelo *root mean squared error* (RMSE), o índice de desigualdade U-Theil, o percentual de sinais corretamente previstos e a estatística de falha de previsão Pesaran-Timmermann. Além disso, foi estabelecida uma estratégia de

negociação baseada nas previsões dos modelos, avaliando-se a lucratividade e risco de cada estratégia, tomando como *benchmark* uma estratégia simplista de negociação do tipo *buy-and-hold*.

Os testes de especificação indicaram que todas as séries utilizadas são estacionárias e todas possuem efeitos ARCH, além de rejeitarem a hipótese de aderência à curva Normal. O teste de não-linearidade, baseado na estatística BDS, indicou que todas as séries aceitaram a hipótese de presença de não-linearidade em seus mecanismos geradores de dados. Verificou-se que as séries de alta frequência possuem um grau de não-linearidade superior às séries de baixa frequência, ou seja, quanto maior a frequência com que a série foi observada, maior foi o seu grau de não-linearidade.

Os modelos lineares de previsão utilizados foram os modelos auto-regressivos (AR), modelos média móvel (MA), modelos auto-regressivos de média móvel (ARMA) e modelos ARMA considerando heteroscedasticidade condicional auto-regressiva dos resíduos (ARMA-GARCH). Os procedimentos de estimação desses modelos, baseados na metodologia proposta por Box e Jenkins (1976), estão bem estabelecidos na área de séries temporais. Os modelos não-lineares utilizados foram as redes neurais *perceptron* multi-camadas, redes neurais função de base radial e os sistemas nebulosos do tipo Takagi-Sugeno. As estruturas matemáticas NARX e NARMAX (Chen e Billings, 1989) foram utilizadas para a representação dinâmica não-linear das séries temporais.

Apesar da disponibilidade de uma série de outros modelos não-lineares, os modelos de redes neurais e os sistemas nebulosos encontram-se entre os principais candidatos ao problema da previsão de séries temporais. O modelo de redes neurais possui propriedades interessantes, como a capacidade de realizar um aprendizado, através de treinamento específico, acerca do ambiente ou sistema no qual está sendo utilizado. Os sistemas nebulosos, por sua vez, possuem a capacidade de lidar com a incerteza, tanto do ponto de vista qualitativo quanto quantitativo. No caso do sistema nebuloso utilizado neste trabalho, a idéia fundamental é “dividir para conquistar”, ou seja, decompor um sistema complexo em subsistemas simples, tornando mais simples o problema da previsão. Além disso, as redes neurais e os sistemas nebulosos possuem grande aplicabilidade quando se trata de modelagem e controle de processos não-lineares em âmbito multidisciplinar, passando pela engenharia elétrica, engenharia mecânica, engenharia da produção, química, física, biologia e, mais recentemente, economia e finanças (Coelho, 2000).

Os resultados da avaliação da previsão encontrados neste trabalho atestam em favor dos modelos não-lineares. Em todas as séries temporais analisadas, os modelos

não-lineares superaram os modelos lineares na obtenção de uma previsão mais acurada. O modelo não-linear RN-MLP foi o modelo de melhor desempenho em todas as séries.

Deve-se destacar que o desempenho preditivo dos modelos esteve intimamente relacionado com a frequência com que a série temporal é observada. Os maiores índices de sucesso nas previsões foram obtidos nas séries de menor frequência, tanto para os modelos lineares quanto para os modelos não-lineares. A série com frequência semanal foi a que apresentou o maior grau de previsibilidade. O modelo linear de melhor desempenho conseguiu prever corretamente 67% da direção dos movimentos dessa série, ao passo que o melhor modelo não-linear conseguiu prever corretamente 82% dos movimentos da série. Do lado oposto, a série que apresentou o menor grau de previsibilidade foi a série com frequência de 15 min. Dessa forma, quanto maior a frequência com que a série foi observada, maior foi o seu grau de previsibilidade.

Uma possível explicação para esse resultado está relacionada à presença de não-linearidade nas séries. O seu grau mais elevado nas séries de alta frequência pode ter contribuído para reduzir sua previsibilidade, enquanto que nas séries de menor frequência, a maior previsibilidade pode estar relacionada a um grau menor de não-linearidade. A justificativa para essa afirmação reside no fato de que uma maior não-linearidade significa um mecanismo gerador de dados da série mais complexo, o que torna mais difícil a tarefa de se fazer uma boa previsão. Sob essa perspectiva, é possível justificar o desempenho superior dos modelos não-lineares em função da sua maior capacidade em capturar não-linearidades presentes nas séries temporais analisadas.

A avaliação da estratégia de negociação indicou que as estratégias baseadas em modelos não-lineares superaram – em termos de retorno e risco – as estratégias baseadas em modelos lineares e a estratégia passiva de investimento. Em geral, todos os modelos que obtiveram uma boa avaliação estatística, também obtiveram um bom desempenho na estratégia de negociação.

Como sugestões para trabalhos futuros, é interessante o estudo comparativo de outros modelos de redes neurais e sistemas nebulosos, além de outras abordagens não-lineares, como os modelos STAR, SETAR, *nearest-neighbor*, algoritmos genéticos, modelos híbridos neuro-nebulosos-evolutivos, dentre outras abordagens. A utilização de séries temporais de outros ativos financeiros, como ações, índices de ações, contratos futuros, commodities, dentre outros é também recomendada. Por último, é interessante avaliar não apenas a capacidade de previsão um-passo-a-frente, como também a capacidade de previsão k -passos-a-frente.

6

Referências

ABHYANKAR, A., COPELAND, L., WONG, W. Nonlinear dynamics in real-time equity markets indices: evidence from United Kingdom. The Economic Journal. V. 105, N. 431, pp.864-880, 1995.

AGUIRRE, L. Introdução à identificação de sistemas: técnicas lineares e não-lineares aplicadas a sistemas reais. 2. ed. Editora UFMG: Belo Horizonte, 2004.

AGUIRRE, A., AGUIRRE, L. A tutorial introduction to nonlinear dynamics in economics. Nova Economia. V. 7, N. 2, pp. 9-47, 1997.

ANDRADA-FÉLIX, J., FERNANDEZ-RODRIGUEZ, F., GARCIA-ARTILES, M., SOSVILLA-RIVERO, S. An empirical evaluation of non-linear trading rules. Studies in Nonlinear Dynamics and Econometrics. V.7, N.3, pp.1-30, 2003.

BABUSKA, R. Fuzzy systems, modeling and identification. Delft: Delft University of Technology, 2000. (Textos para discussão).

BLAKE, A., KAPETANIOS, G. A radial basis function artificial neural network test for ARCH. Economics Letters. N. 69 pp.15–23, 2000.

BOERO, G., MARROCU, E. The performance on non-linear exchange rates models: a forecasting comparison. Journal of Forecasting. V.21, pp. 513-542, 2002.

BOLDRIN, M., DENECKERE, R. Simple macroeconomic models with very complicated dynamics. Los Angeles: University of California, EUA, 1989. (Textos para discussão).

BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics. V. 31, pp. 307-327, 1986.

BOLLERSLEV, T., CHOU, R., KRONER, K. ARCH modeling in finance: a review of the theory and empirical work. Journal of Econometrics. V. 52, pp. 5-59, 1992.

BOX, G., JENKINS, G. Time series analysis: forecasting and control. Holden Day: San Francisco, 1976.

BRAGA, A., DE CARVALHO, A., LUDEMIR, T. Redes neurais artificiais: teoria e aplicações. LTC: Rio de Janeiro, 2000.

BROCK, W., DECHERT, W., SCHEINKMAN, J. A test for independence based upon the correlation dimension. Wisconsin: University of Wisconsin-Madison, EUA, 1987. (Texto para discussão).

- BROOMHEAD, D.; LOWE, D. Multivariable functional interpolation and adaptative networks. Complex Systems, V.2, pp. 321–355, 1988.
- CAMPBELL, J., LO, A., MACKINLAY, A. The econometrics of financial markets. 1.ed. Princeton: Princeton university press, 1997.
- CASTRO, M. Predição não-linear de séries temporais usando redes neurais RBF por decomposição em componentes principais. 2001. Tese (Doutorado em Engenharia Elétrica), Universidade de Campinas - UNICAMP, 2001.
- CHEN, S., BILLINGS, S. Representation of nonlinear systems: the NARMAX model. International Journal of Control. V.49, pp.1012-1032, 1989.
- CHU, C., ZHANG, G. A comparative study of linear and nonlinear models for aggregate retail sales forecasting. International Journal of Production Economics. V.86, pp. 217-231, 2003.
- CLEMENTS, M., FRANSES, P., SWANSON, N. Forecasting economic and financial time-series with non-linear models. International Journal of Forecasting.v.20, pp. 169-183, 2004.
- COELHO, L. Metodologias da inteligencia computacional em identificação e controle de processos: abordagem nebulosa evolutiva e neural. 1997. 144f. Dissertação (Mestrado em Ciências da Computação) - Universidade Federal de Santa Catarina, Florianópolis, 1997.
- _____. Identificação e controle de processos multivariáveis via metodologias avançadas e inteligência computacional. 2000. 342 f. Tese (Doutorado em Engenharia Elétrica) - Universidade Federal de Santa Catarina. Florianópolis, 2000.
- COELHO, L., SANTOS, A. Sistema Híbrido Inteligente Baseado em Sistemas Nebulosos e Estratégias Evolutivas Aplicado à Previsão do Índice da BOVESPA: Fundamentos e Estudo de Caso. In: Encontro Nacional de Administração – ENANPAD, Atibaia. Anais do XXVII Encontro Nacional de Administração – ENANPAD. Atibaia:ANPAD, pp. 215, 2003.
- COX, E. The fuzzy systems handbook. 1. ed. Academic Press: Cambridge, 1994.
- DA SILVA, S., MATSUSHITA, R., GLERIA, I., FIGUEREDO, A., RATHIE, P. International finance, Lévy distributions, and the econophysics of exchange rates. Communications in Nonlinear Science and Numerical Simulation. V.10, N.4, pp. 365-393, 2005.
- DASGUPTA, C., DISPENSA, G., GHOSE, S. Comparing the predictive performance of a neural network model with some traditional market response models. International Journal of Forecasting. V.10, pp. 235-244, 1994.
- DE CASTRO, F., DE CASTRO, M. Redes neurais artificiais. Porto Alegre: Pontifícia Universidade Católica do Rio Grande do Sul, 2001.(Textos para discussão)
- DIEBOLD, F., NASON, J. Non-parametric exchange rate prediction. Journal of International Economics. N.28, pp. 315-332, 1990.

DIJK, D., FRANSES, P. Selecting a nonlinear time series model using weighted tests of equal forecast accuracy. Rotterdam: Erasmus University, 2003. (Textos para discussão)

DONALDSON, R., KAMSTRA, M. An artificial neural network-GARCH model for international stock return volatility. Journal of Empirical Finance. V.4, pp. 17-46, 1997.

DUFFY, J., MCNELIS, P. Approximating and simulating the stochastic growth model: parameterized expectations, neural networks and the genetic algorithm. Journal of Economic Dynamics & Control. V. 25, pp. 1273-1303, 2001.

EAKINS, S., STANSELL, S. Can value-based stock selection criteria yield superior risk-adjusted returns: an application of neural networks. International Review of Financial Analysis. V.12, N.1, pp. 83-97, 2003.

ENDERS, W. Applied econometric time series. 1 ed. John Wiley & Sons. 1995.

ENGLE, R. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. Econometrica. N.50, pp. 987-1008, 1982

EITRHEIM, O., TERASVIRTA, T. Testing the adequacy of smooth transition autoregressive models. Journal of Econometrics. N. 74, pp. 59-75, 1996.

FAIR, R. Evaluating the predictive accuracy of models. In: GRILICHEST, Z., INTRILIGATOR, M. Handbook of Econometrics, 1. ed, V. 3, North Holland, 1999.

FAMA, E. F. Efficient capital markets: a review of theory and empirical work. Journal of Finance. V. 2, N. 25, pp. 383-417, 1970.

_____ Efficient capital markets: II. Journal of Finance V.46, N.5, pp. 1575-1617, 1991.

FAMA, E.; FRENCH, K. The cross section of expected stock returns. Journal of Finance, V.47 N.2, pp. 427-465, 1992.

FANG, Y., XU, D. The predictability of asset returns: an approach combining technical analysis and time series forecasts. International Journal of Forecasting. V. 19, N. 3, pp. 369-386, 2003.

FERNANDES, L., PORTUGAL, M., NAVAUX, P. Previsão de séries de tempo: redes neurais artificiais e modelos estruturais. In: Encontro Brasileiro de Econometria. Anais do XVII Encontro Brasileiro de Econometria, Salvador, 1995.

FERNANDEZ-RODRIGUEZ, F., GONZALEZ-MARTEL, C., SOSVILLA-RIVERO, S. On the profitability of technical trading rules based on artificial neural networks: evidence from the Madrid stock market. Economics Letters. N.69, pp. 89-94, 2000.

FRANSES, P., GRIENSVEN, K. Forecasting exchange rates using neural networks for technical trading rules. Studies in nonlinear dynamics and econometrics. V.2, N.4, pp.109-114, 2003.

FREITAS, S., SOUZA, A. Utilização de um modelo baseado em redes neurais para a precificação de opções. In: Encontro Nacional de Administração – ENANPAD. Anais do XX encontro nacional de administração – ENANPAD. Salvador, 2002.

GALVÃO, R., BECERRA, V., ABOU-SEADA, M. A wavelet network classifier applied to financial distress prediction. In: Congresso Brasileiro de Redes Neurais. Anais do VI congresso brasileiro de redes neurais, São Paulo, 2003.

GLEISSER, I. Caos e complexidade. 1. ed. Rio de Janeiro:Campus, 2002.

GLEICK, J. Caos: a criação de uma nova ciência. 9. ed. Rio de Janeiro:Campus, 1989.

GLERIA, I., MATSUSHITA, R., DA SILVA, S. Scaling power laws in the Sao Paulo Stock Exchange. Economics Bulletins. V.7, N.3, pp.1-12, 2002.

GRANGER, C. Modeling nonlinear relationships between extended-memory variables. Econometrica. V. 63, N. 2, pp. 265-279, 1995.

GUERRA, F., COELHO, L. Identificação de sistema dinâmico caótico usando rede neural *perceptron* multicamadas. In: Simpósio sul-brasileiro de matemática e informática. Anais do 1º simpósio sul-brasileiro de matemática e informática, Curitiba, 2002.

GUSTAFSON, D. E.; KESSEL, W. C. Fuzzy Clustering with a Fuzzy Covariance Matrix. In: IEEE Conference on Control and Decision. Proceedings of IEEE Conference on Control and Decision, San Diego, CA, USA, pp. 761-766. 1979.

HAMILTON, J. Time series analysis. Princeton University Press: Princeton, 1994.

HAYKIN, S. Redes Neurais: Princípios e Prática. 2. ed, Bookman, Porto Alegre, 2001.

HINICH, M.J. Testing for gaussianity and linearity of a stationary time series. Journal of Time Series Analysis, V. 3, pp. 169-176, 1982.

HORNIK, K., STINCHCOMBE, M., WHITE, H. Multilayer feedforward networks are universal approximators. Neural Networks. V.2, pp.359-366, 1989.

HU, M., TSOUKALAS, C. Combining conditional volatility forecasts using neural networks: an application to the EMS exchange rates. Journal of International Financial Markets, Institutions and Money. V.9, pp. 407-422, 1999.

INTRILIGATOR, M. Economic and econometric models In: INTRILIGATOR, M., GRILICHES, Z. Handbook of Econometrics, 1. ed, V. 1, North Holland, 1983.

IVAKHNENKO, A. G., IVAKHNENKO, G.A. The review of problems solvable by algorithms of the group method of data handling (GMDH). Pattern recognition and image analysis. V. 5, N. 4, pp. 527-535, 1995.

JANG, J.-S. R., SUN, C.-T. Neuro-fuzzy modeling and control. Proceedings of the IEEE. V. 83, No. 3, pp. 378-406, 1995.

- KAASHOEK, J., VAN DIJK, H. Neural network pruning applied to real exchange rate analysis. Journal of Forecasting. V.21, pp. 559-577, 2002.
- KAPETANIOS, G. Testing for structural breaks in nonlinear dynamic models using neural networks. Londres: University of London, 2002. (Textos para discussão)
- KEENAN, D.M. A Tukey nonadditivity type test for time series nonlinearity. Biometrika. V.72, pp. 39-44, 1985.
- KENDALL, M. Time series. Griffin London, 1973.
- KROSE, B., VAN DER SMAGT, P. An introduction to neural networks. 8 ed. Amsterdã: Universidade de Amsterdã, 1996.
- KUAN, C., LIU, T. Forecastin Exchange rates using feedforward and recurrent neural networks. Journal of Applied Econometrics. V.10, pp. 347-364, 1995
- LAKONISHOK, J., A. SHLEIFER, R., W. VISHNY, Contrarian investment, extrapolation, and risk. Journal of Finance, V.49 N.5, pp. 1541-78. 1994.
- LEBARON, B. Chaos and nonlinear forecastability in economics and finance. Wisconsin: University of Wisconsin-Madison. 1994. (Textos para discussão)
- LEE, T., WHITE, H., GRANGER, C. Testing for neglected nonlinearity in time series models: a comparison of neural network methods and alternative tests. Journal of Econometrics. N. 56, pp. 269-290, 1993.
- LEITCH, G. TANNER, E. Economic forecast evaluation: profits versus the conventional error measures. American Economic Review. V.81, N. 3, pp. 580-590, 1991.
- LEÓN, T.; LIERN, V.; VERCHER, E. Viability of infeasible portfolio selection problems: a fuzzy approach, European Journal of Operational Research, V. 139, pp. 178-189, 2002.
- LEUNG, M., DAOUK, H., CHEN, A. Forecasting stock indices: a comparison of classification and level estimation models. International Journal of Forecasting. V.10, pp.173-190, 2000.
- LEVENBERG, K. A method for the solution of certain non-linear problems in least squares. Quart. Applied Mathematics, V. 2, no. 2, pp. 164-168, 1944.
- LO, A., MACKINLAY, A.C. Stock prices do not follow random walks: evidence from a simple specification test. Review of Financial Studies. V.1, pp. 41-66, 1998.
- MALHOTRA, R. MALHOTRA, D. Differentiating between good credits and bad credits using neuro-fuzzy systems. European Journal of Operational Research. V.136, pp. 190-211, 2002.
- MARQUARDT, D. W. An algorithm for least-squares estimation of nonlinear parameters. SIAM J. of Applied Mathematics. V. 11, no. 2, pp. 431-441, 1963.

- MCCULLOUGH, W., PITTS, W. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. N.5, pp. 115-133, 1943.
- MEADE, N. A comparison of the accuracy of short term foreign exchange forecasting methods. International Journal of forecasting. N. 18, pp. 67-83, 2002.
- MEDEIROS, M., TERASVIRTA, T. RECH, G. Building neural network for time series: a statistical approach. Estocolmo: Stockholm School of Economics, 2002. (Texto para discussão).
- MOODY, J.; DARKEN, C. Fast learning in networks of locally-tuned processing units. Neural Computation. V.1, pp. 281-294, 1989.
- NARENDRA, K., PARTHASARATHY, K. Identification and control of dynamical systems using neural networks. IEEE Transactions on Neural Networks, V. 1, N.1, pp. 4-27. 1990.
- NEELY, C. WELLER, P. Intraday technical trading in the foreign exchange market. Journal of International Money and Finance. V.22, N. 2, pp. 223-237, 2003.
- NELSON, C. Applied time series analysis. Holden Day: San Francisco, 1973.
- OCAL, N. Nonlinear models for UK macroeconomic time series. Studies in nonlinear dynamics and econometrics. N. 3, V. 4, pp. 123-135, 2000.
- OHASHI, A. Testing for chaotic dynamic in stock returns. Recife: Universidade Federal de Pernambuco, 2000. (Texto para discussão)
- OLIVEIRA, E., LUDEMIR, T. Forecasting the IBOVESPA using NARX networks and random walk model. In: Congresso Brasileiro de Redes Neurais. Anais do VII Congresso Brasileiro de Redes Neurais, Recife, 2002.
- OLSON D., MOSSMAN C. Neural network forecasts of Canadian stock returns using accounting ratios. International Journal of Forecasting. N. 3, V. 19, pp. 453-465, 2003.
- ORR, M. Introduction to radial basis function network. Edimburgo: University of Edinburgh, 1996. (Textos para discussão).
- PESARAN, M., TIMMERMANN, A. A simple nonparametric test of predictive performance. Journal of Business & Economic Statistics. V.10, N.4, pp. 561-565, 1992.
- PINDYCK, R., RUBINFELD, D. Econometric models and economic forecasts. 3. ed, McGraw-Hill, 1991.
- PORTUGAL, M. Neural networks versus time series methods: a forecasting exercise. Revista Brasileira de Economia. V. 49, pp. 611-629, 1995.
- PORTUGAL, M., FERNANDES, L. Redes neurais artificiais e previsão de séries econômicas: uma introdução. Porto Alegre: Universidade Federal do Rio Grande do Sul, 1996. (Textos para discussão).
- QI, M. Predicting US recessions with leading indicators via neural network models. International Journal of Forecasting. V.17, pp. 383-401, 2001.

RAMSEY, J. Tests for specification errors in classical linear least squares regression analysis. Journal of the Royal Statistical Society. V. 31, pp. 350-371. 1969.

_____. If nonlinear models cannot forecast, what use are they? Studies in nonlinear dynamics and econometrics. V. 1, N.2, pp. 65-87, 1996.

REFENES, A., ZAPRANIS, A., FRANCIS, G. Stock performance modeling using neural networks: a comparative study with regression models. Neural Networks, V. 7, N. 2, pp. 375-388, 1994.

RISTESKI, D., KULAKOV, A., DAVCEV, D. Single exponential smoothing method and neural network in one method for time series prediction. Proceedings of the IEEE Conference on Cybernetics and Intelligent Systems, Singapura, pp. 740-744, 2004.

ROUBOS, J., MOLLOV, R., BABUSKA, R., VERBRUGGEN, H. Fuzzy model-based predictive control using Takagi-Sugeno models. International Journal of Approximate Reasoning. V.23, pp.3-30, 1999.

RUBIO, O., RIVERO, S., RODRIGUEZ, F. Non-linear forecasting models: some applications to analysis of financial time series. Madri: Fundacion de Estudios de Economía Aplicada – FEDEA, 2002. (Textos para discussão).

SANDRI, S., CORREA, C. Lógica nebulosa. In: Escola de Redes Neurais. Anais da V Escola de Redes Neurais, São José dos Campos, pp. 17-34, 1999.

SCHEINKMAN, J. Nonlinearities in economics dynamics. Economic Journal. V.100, pp. 33-48, 1990.

SGROI D., ZIZZO, J. Strategy Learning in 3x3 Games by Neural Networks. Cambridge: Cambridge University, 2002. (Textos para discussão).

SILVA, A., CECHIN, A., SILVA, J., PORTUGAL, M. Redes neurais artificiais e análise de sensibilidade aplicadas às exportações brasileiras. In: Encontro de Economia da Região Sul. Anais do V Encontro de Economia da Região Sul – ANPEC SUL, Florianópolis, 2002.

STEINER, M., ADAMOVICZ, E., NETO, P. Redes neurais na análise econômico-financeira de empresas para concessão de crédito. In: Congresso brasileiro de redes neurais. Anais do VI congresso brasileiro de redes neurais, São Paulo, 2003.

STOCK, J., WATSON, M. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. NBER working paper, Cambridge, USA, N. 6607, 1999. Disponível em <www.nber.org/papers/w6607>. Acesso em 20 junho 2003.

SWANSON N., WHITE H. A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. Journal of Business & Economic Statistics. V. 13, N.3, pp. 265-275, 1995.

TAKAGI, T., SUGENO, M. Fuzzy identification of systems and its application to modeling and control. IEEE Transactions on Systems, Man, and Cybernetics, V.15, pp.116-132, 1985.

- TERASVIRTA, T., TJOSTHEIN, D., GRANGER, C. Aspects of modeling nonlinear time series. In: ENGLE, R., MCFADDEN, D. Handbook of Econometrics, 1. ed, V. 4, North Holland, 1999.
- TEROUI, N., VAN DIJK, H. Combined forecasts from linear and nonlinear time series models. International Journal of Forecasting. V.18, pp. 421-438, 2002.
- TSAIH, R., HSU, Y. LAI, C. Forecasting S&P 500 stock index futures with a hybrid AI System. Decision Support Systems. V. 23, pp., 161-174, 1998.
- TSAY, R.S. Nonlinearity tests for times series. Biometrika, V.73, pp. 461-466, 1986.
- TSENG, F.-M.; TZENG, G.-H. A fuzzy seasonal ARIMA model for forecasting, Fuzzy Sets and Systems, V. 126, pp. 367-376. 2002.
- VAN DIJK, D., TERASVIRTA, T. Smooth transition autoregressive models - a survey of recent developments. Rotterdam: Erasmus University, 2000.(Textos para discussão).
- VON ALTROCK, C. Applying fuzzy logic to business and finance, Optimus, no. 2, pp. 38-39, 2002.
- VON ZUBEN, F., DE CASTRO, L. Regressão paramétrica e não-paramétrica. Campinas: Unicamp, 2001. Disponível em <<http://www.dca.fee.unicamp.br/~vonzuben/courses/ia353.html>>. Acesso em 8 agosto 2003.
- WONG, K.; SELVI, Y. Neural network applications in finance: a review and analysis of literature (1990-1996). Information and Management. V. 34, N. 3, pp. 129-139, 1998.
- WU, A., TAM, P. K. S. A Simplified Model of Fuzzy Inference System Constructed by Using RBF Neurons. Proceedings of IEEE International on Fuzzy Systems Conference. V. 1, Seoul, Korea, pp. 50-54, 1999.
- XU, Y. Small levels of predictability and large economic gains. Journal of Empirical Finance. V.11, pp. 247-275, 2004.
- YAO, J., TAN, C., POH, H. Neural networks for technical analysis: a study on KLCI. International Journal of Theoretical and Applied Finance. V. 2, N. 2 pp. 221-241, 1999.
- YING, H. General SISO Takagi-Sugeno Fuzzy Systems with Linear Rule Consequent are Universal Approximators. IEEE Transactions on Fuzzy Systems. V. 6, No. 4, pp. 582-587, 1998.
- ZADEH, L. Fuzzy sets. Information and Control. V. 8, pp. 338-353, 1965.