

"A FEA e a USP respeitam os direitos autorais deste trabalho. Nós acreditamos que a melhor proteção contra o uso ilegítimo deste texto é a publicação online. Além de preservar o conteúdo motiva-nos oferecer à sociedade o conhecimento produzido no âmbito da universidade pública e dar publicidade ao esforço do pesquisador. Entretanto, caso não seja do interesse do autor manter o documento online, pedimos compreensão em relação à iniciativa e o contato pelo e-mail bjbfea@usp.br para que possamos tomar as providências cabíveis (remoção da tese ou dissertação da BDTD)."

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ECONOMIA

**TRÊS ENSAIOS COM APLICAÇÕES DE REDES NEURAS EM
SÉRIES FINANCEIRAS**

Rodrigo Octávio Marques de Almeida

Orientadora: Prof^a Dr^a Vera Lúcia Fava

São Paulo
2004

Reitor da Universidade de São Paulo
Prof.Dr. Adolpho José Melfi

Diretor Faculdade de Economia, Administração e Contabilidade
Profa. Dra. Maria Tereza Leme Fleury

Chefe do Departamento de Economia
Profa. Dra. Elizabeth Maria Mercier Querido Farina

Coordenador do Programa de Pós-Graduação em Economia
Prof.Dr. José Paulo Zeetano Chahad

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ECONOMIA

**TRÊS ENSAIOS COM APLICAÇÕES DE REDES NEURAIS EM
SÉRIES FINANCEIRAS**

Rodrigo Octávio Marques de Almeida

Orientadora: Prof^a Dr^a Vera Lúcia Fava



20600026641

Tese apresentada ao Departamento de
Economia, Administração e
Contabilidade da Universidade de São
Paulo, como requisito parcial para a
obtenção do título de Doutor em
Economia

São Paulo
2004

AGRADECIMENTOS

À Prof. Vera Lúcia Fava pelos ensinamentos, pela paciência e pela dedicada orientação no curso de doutorado e durante a elaboração desta tese.

Ao professores Maria Dolores Diaz e Rogério Rosenfeld, membros da Banca de Qualificação pelas alterações sugeridas.

A Alessandra Maielo e Camilla Roberta pela ajuda em diferentes partes da pesquisa.

Um agradecimento especial a Claudia Viegas.

RESUMO

A Hipótese da Eficiência de Mercados (HEM) postula que os preços dos ativos nos mercados financeiros devem refletir toda a informação disponível: como consequência, os preços devem ser consistentes com seus fundamentos. Esta tese examina a evidência empírica da HEM usando a abordagem das redes neurais.

Muitos estudos têm mostrado que Redes Neurais Artificiais têm a capacidade de aprender a mecânica dos mercados acionários. A tese está dividida em três ensaios. O primeiro ensaio aplica os modelos de redes neurais para prever os três principais mercados acionários latino-americanos (Brasil, Argentina e México). O ensaio dois, foca o papel da análise técnica na sinalização dos pontos de entrada e saída do mercado. O ensaio três, estende os ensaios um e dois e aplica as redes recorrentes e de saltos de conexão. Os três ensaios concluem que as redes neurais são uma boa ferramenta para o “market timing” da alocação de ativos.

ABSTRACT

The Efficient Market Hypothesis (EMH) states that asset prices in financial markets should reflect all available information: as consequence, price should always be consistent with “fundamentals”. In this dissertation, we examine the empirical evidence of EMH using a neural network approach.

Many studies have shown that the artificial neural networks have the capability to learn the underlying mechanics of stock markets. The dissertation is divided in three essays. First essay applies neural network models to predict the three main stock markets in Latin America (Brazil, Argentina and Mexico) Essay two focuses on the role of technical analysis in signaling the timing of stock market entry and exit. Essay three extends essays one and two and applies both recurrent and jump connections networks. The three essays conclude that neural networks are a good tool to market timing asset allocation

SUMÁRIO

Introdução	1
Ensaio 1: Previsibilidade Não-Linear dos Retornos dos Mercados Acionários Latinos Americanos com Redes Neurais	
1.1. Introdução	6
1.2. Resenha da Pesquisa sobre Eficiência de Mercados	7
1.2.1 Revisão dos Argumentos Teóricos	7
1.2.2. Argumentos Empíricos	9
1.2.3. Testes de Previsibilidade	11
1.3. Descrição do Procedimento do Teste	13
1.3.1. Aspectos Teóricos.	13
1.3.2. Implementação do Procedimento	18
1.3.2.1. Descrição das Séries	18
1.3.2.2. Escolha da Função de Ativação	25
1.3.2.3. Escolha do Algoritmo de Aprendizado	27
1.3.2.4. Escolha do número de neurônios escondidos e dos números de defasagens relevantes	28
1.4. Resultados	30
1.5. Lucratividade	35
1.6. Conclusão	42
Apêndice 1 : Características das Arquiteturas das Redes Neurais	44
Apêndice 2: Desempenho da RNA (1,1,1) nos conjuntos de treinamento e teste	46
Ensaio 2: Regras de Análise Técnica e a Previsibilidade dos Retornos com Redes Neurais	
2.1. Introdução	48
2.2. Padrões e Indicadores Técnicos	49
2.3. Resenha da Literatura	63
2.4. Metodologia dos Testes	68
2.5. Resultados	72
2.6. Conclusão	76

Ensaio 3: Testando a Capacidade de Previsão de Diferentes Arquiteturas de Redes Neurais Artificiais: Uma Comparação entre diferentes extensões do Algoritmo de Retropropagação	
3.1. Introdução	77
3.2. Aspectos Teóricos das Diferentes Arquiteturas	82
3.3. Estimação e Resultados.	87
3.4. Conclusões	91
Conclusão Final	94
Bibliografia	95

Índice de Tabelas

Tabela 1.1: Resumo das Estatísticas e dos Testes de Não-Linearidade	22
Tabela 1.2 :Estatísticas dentro da amostra (3/01/93 a 22/5/02)	32
Tabela 1.3 : Estatísticas fora da amostra (23/05/02 a 24/07/02)	35
Tabela 1.4 : Brasil: Avaliação fora da amostra das previsões 1 passo à frente das arquiteturas de RNAs usando a comparação entre estratégias passivas e ativas de investimento	38
Tabela 1.5 : Argentina: Avaliação fora da amostra das previsões 1 passo à frente das arquiteturas de RNAs usando a comparação entre estratégias passivas e ativas de investimento	40
Tabela 1.6: México: Avaliação fora da amostra das previsões 1 passo à frente das arquiteturas de RNAs usando a comparação entre estratégias passivas e ativas de investimento	42
Tabela 2.1 : Resultados Resultados dentro da Amostra considerando os sinais de compra e venda gerados pelas variáveis de Análise Técnica	72
Tabela 2.2: Resultados Fora da Amostra considerando os sinais de compra e venda gerados pelas variáveis de Análise Técnica	73
Tabela 2.3: Avaliação das Previsões Um Passo à Frente das Arquiteturas de RNAs Comparando a Estratégia Passiva com as Diferentes Estratégias Ativas	75

Tabela 3.1: Estatísticas dentro da amostra- Redes recorrentes e com Saltos	88
Tabela 3.2: Estatísticas fora da amostra-Redes Recorrentes e com Saltos	88
Tabela 3.3: Avaliação das Previsões Um Passo à Frente das Arquiteturas de RNAs Comparando a Estratégia Passiva com as Diferentes Estratégias Ativas	90
Tabela 3.4. Avaliação das Previsões Um Passo à Frente das Arquiteturas de RNAs Comparando a Estratégia Passiva com Diferentes Estratégias Ativas considerando a geração de sinais da variável de análise técnica	92

Índice de Gráficos

Gráfico 1.1: Série do Índice Bovespa denominado em dólares	
– diário (3/01/93-24/07/02)	19
Gráfico 1.2: Série dos Retornos do Índice Bovespa denominado em dólares	
- diário (3/01/93-24/7/02)	19
Gráfico 1.3: Série do Índice Merval denominado em dólares	
– diário (03.01/93-24/07/02)	20
Gráfico 1.4: Série dos Retornos do Índice Merval denominado em dólares	
– diário (03/01.93-24/07/02)	20
Gráfico 1.5.. Série do Índice (MexBolsa) denominado em dólares	
- diário (03/01/93-24/07/02)	21
Gráfico 1.6. Série dos Retornos do Índice MexBolsa denominado em dólares	
– diário (03/01/93-24/07/02)	21
Gráfico 1.7: Brasil- Resultado das Diferentes Estratégias	
para US\$ 100 Investidos	37
Gráfico 1.8: Argentina- Resultado das Diferentes Estratégias	
para US\$100 Investidos	39
Gráfico 1.9. México - Resultado das Diferentes Estratégias	41
Gráfico 1.10. Desempenho da Rede RNA(1,1,1) no conjunto de treinamento	46
Gráfico 1.11 Desempenho da rede RNA(1,1,1) no conjunto de teste	47

Gráfico 2.1 : Exemplo dos principais padrões utilizando a série diária do Índice Bovespa denominado em dólares entre janeiro de 1993 e agosto de 2002.	55
Gráfico 2.2: Exemplo dos sinais de compra e venda emitidos por uma regra de médias móveis aritméticas	57
Gráfico 2.3: Exemplo dos sinais de compra e venda emitidos pela regra de MACD	58
Gráfico 2.4: Exemplo da geração de sinais de compra e venda do estocástico	60
Gráfico 2.5: Exemplo dos sinais de compra e venda gerados pelo ADX no Ibovespa em dólares	62
Gráfico 2. 6: Resultado de Alguns Estudos de Análise Técnica em relação à Estratégia Passiva	74
Gráfico 3.1: Resultado das Diferentes Redes para cada \$100 investidos	89
Gráfico 3.2: Resultados das Diferentes Redes com uma Regra de Análise Técnica para cada 100 dólares investidos	91

Índice de Figuras

Figura 2.1: Formação Ombro-Cabeça-Ombro	50
Figura 2.2: Formação Ombro-Cabeça-Ombro Invertido	50
Figura 2.3: Topos Crescentes	51
Figura 2.4: Fundos Decrescentes	52
Figura 2.5: Triângulos	52
Figura 2.6: Topos Duplos	54
Figura 2.7: Fundos Duplos	54
Figura 2.8: Determinação visual do Average Directional Index (ADX)	61
Figura 3.1: RR(1)	78
Figura 3.2: RR(2)	79
Figura 3.3: RR(3)	80
Figura 3.4: Rede com 3 camadas e saltos	81

INTRODUÇÃO

O objetivo geral do presente trabalho é apresentar, por meio de três ensaios aplicações de redes neurais artificiais a séries financeiras e, simultaneamente testar a hipótese de que os mercados financeiros são eficientes.

Redes Neurais Artificiais são redes compostas por conexões simples capazes de processar informação em resposta a inputs externos. As Redes Neurais Artificiais (RNAs) foram originalmente apresentadas como modelos do sistema nervoso humano. Logo depois da segunda guerra mundial, os cientistas descobriram que a fisiologia do cérebro era similar ao processamento eletrônico dos computadores. No caso dos computadores, a unidade elementar de processamento é o bit, que está no estado “on” ou “off”. No caso do cérebro humano, os neurônios se incumbem do processamento de dados. Os neurônios são células que seguem um padrão binário de emissão de sinais. Quando um neurônio está ativado ele emite um sinal para um outro neurônio através de uma rede de sinapses.

Portugal e Fernandes(1996) usam a definição de Hecht-Nielsen (1988, p.2 e 3) segundo a qual uma RNA é *“uma estrutura de processamento de informação distribuída e paralela. Ela é formada por unidades de processamento, comumente chamadas de nós, neurônios ou células, interconectadas por arcos unidimensionais, também chamados de ligações, conexões ou sinapses. Os nós possuem memória local e podem realizar operações de processamento de informação localizada. Cada célula possui uma única*

saída, a qual pode se ramificar em muitas ligações colaterais (cada ramificação possuindo o mesmo sinal de saída do neurônio). Todo o processamento que se realiza em cada unidade deve ser completamente local, isto é, deve depender apenas dos valores correntes dos sinais de entrada que chegam dos neurônios através das conexões. Estes valores atuam sobre os valores armazenados na memória local”

No início da década de 1940, McCulloch e Pitts (1943) estudaram o potencial de interconexão de um modelo de um neurônio. Os pesos em um neurônio são dispostos de tal forma que os neurônios efetuam uma função lógica simples, com diferentes neurônios efetuando diferentes funções.

No final dessa mesma década, Donald Hebb mostra, que quando dois neurônios são ativados simultaneamente o resultado é a memória biológica. A conexão sináptica de neurônios sincronizados é reforçada e tem preferência sobre neurônios que não são ativados simultaneamente. O nível de preferência é medido como um valor ponderado. O reconhecimento de padrões, é baseado nas ponderações das conexões entre os vários pares de neurônios que são ativados simultaneamente. Escrito de outra forma, quando dois neurônios são simultaneamente ativados, uma mudança ocorre no sistema nervoso tal que suas ativações tornam-se mais correlacionadas no futuro.¹

A idéia apresentada por Hebb(1949) era a de desenvolver um modelo de computador que replicasse a maneira como os neurônios formavam conexões no cérebro humano. Roseblatt(1958) cunhou o nome “perceptron”, que é uma classe de redes neurais. O “perceptron” mais típico consiste em uma camada de inputs conectada com pesos fixos a neurônios de associação. Os pesos nas conexões se ajustam de forma a

¹ Golden (1996, p.17)

buscar convergência aos “pesos verdadeiros”, ou seja, os pesos que irão resolver o problema.

Fausett(1994) explica que após a considerada “idade de ouro” das redes neurais, correspondendo às décadas de 1950 e 1960, a década de 1970 viu uma redução do entusiasmo por essa linha de pesquisa. As duas principais razões seriam a incapacidade de perceptrons de uma camada em resolver problemas simples como a chamada função XOR² e a ausência de um método geral para o treinamento de uma rede com camadas múltiplas.

Rumelhart, Hinton e Williams(1986) introduziram o algoritmo de retropropagação à pesquisa com redes neurais. Uma rede que utiliza o algoritmo de retropropagação, em geral, consiste de três conjuntos de unidades conectados por camadas de pesos. As variáveis explicativas ou inputs enviam “estímulos” para as outras camadas através da chamada “função de ativação”. As “unidades escondidas” recebem os estímulos da primeira camada e geram “outputs” que serão “inputs” na última camada onde o problema é resolvido. Os pesos entre as conexões vão se alterando durante o processo de aprendizagem.

Segundo Golden (1996) o sucesso das arquiteturas de redes neurais que utilizam a retropropagação se deve a três fatores distintos: (1) a descoberta de que os algoritmos das RNAs podem ser vistos como algoritmos de otimização não linear, que procuram um mínimo de alguma função objetivo; (2) a descoberta do poder das estruturas de múltiplas

² Segundo Haykin (1999), uma vez que o perceptron elementar não possui camada oculta, este não pode classificar padrões de entrada que não sejam linearmente separáveis. No problema XOR (Exclusive OR) ocorrem padrões não linearmente separáveis. Este problema pode ser resolvido com a inclusão de uma camada oculta.

camadas e (3) a importância das unidades escondidas com funções não lineares e diferenciáveis.

McNelis (2003 p.4) fornece uma justificativa exemplar para o uso das redes neurais em séries financeiras e que é o cerne deste trabalho.

“The appeal of the neural network approach lies in its assumption of “bounded rationality”: when we forecast in financial markets, we are forecasting the forecasts of others, or “approximating” the expectations of others. Financial market participants are thus engaged in a learning process, adapting our prior subjective beliefs from past mistakes.”

De fato, como argumentam Frances e VanDijk (2000) as séries financeiras apresentam características que tornam próprio o uso de modelos não-lineares como forma de previsão. Estas características são: (1) retornos extremos em termos absolutos ocorrem mais frequentemente do que poderia ser esperado sob a hipótese de que a série tivesse distribuição normal. (2) estes retornos muito grandes em termos absolutos tendem a se agrupar, o que segundo os autores indica a possibilidade da presença de volatilidade variante em relação ao tempo; (3) a distribuição dos retornos tende a ser assimétrica; (4) períodos voláteis tendem a antecipar retornos negativos.

Este trabalho está dividido em três ensaios interconectados.

No primeiro ensaio, aplica-se o modelo de previsão de redes neurais para mostrar a ineficiência de três mercados de ações: Brasil, Argentina e México. São testadas diferentes arquiteturas e diferentes funções de ativação. Como resultado obteve-se mais de uma rede capaz de “bater o mercado”, ou seja, ter um desempenho bem acima de uma estratégia passiva também chamada de “compra e segura” (buy and hold).

O segundo ensaio é uma extensão do primeiro. Escolheu-se uma das arquiteturas com bom desempenho para o índice Bovespa³, no ensaio 1 e foram testados como inputs adicionais, regras de análise técnica. A análise técnica vem ganhando destaque no meio acadêmico depois de décadas de utilização no mercado financeiro. O objetivo do ensaio 1 também é mantido, ou seja, através dos sinais de compra e venda de uma rede neural artificial procura-se uma estratégia de decisão de investimentos que vença a estratégia passiva.

O terceiro ensaio amplia a discussão para as redes recorrentes e com saltos nas conexões. A “rede benchmark”, ou seja, a referência, é sempre a mesma, ou seja a RNA(1,1,1) do índice Bovespa denominado em dólares. Redes recorrentes vem ganhando popularidade uma vez que incorporam algumas características de modelos de séries de tempo, como a memória em um ambiente não linear. No entanto, apesar da “popularidade” espelhada nas indicações de aprofundamento das pesquisas no final de muitos artigos que tratam do algoritmo de retropropagação, existe pouca literatura aplicada ao mercado financeiro sobre este assunto.

Este ensaio também testa o que acontece quando um input de análise técnica serve como variável explicativa além das defasagens dos retornos. Nota-se que a capacidade de melhora depende da arquitetura da rede.

³ Esta será uma rede usada como “benchmark”, apenas uma referência em função da capacidade de generalização.

Ensaio 1

Previsibilidade Não Linear dos Retornos dos Mercados Acionários Latinos Americanos com Redes Neurais Artificiais

1.1 Introdução:

Pode-se argumentar que em função da extrema volatilidade dos retornos dos mercados emergentes, é difícil a existência de algum grau de previsibilidade. A despeito desta volatilidade, Harvey (1995) documenta que existe mais previsibilidade nos retornos de mercados emergentes do que em mercados desenvolvidos. Uma hipótese é que os mercados acionários de países emergentes são menos eficientes do que os mercados desenvolvidos, e este grau de ineficiência pode ser explorado por um modelo de previsão.

Dessa forma, mercados emergentes fornecem uma boa oportunidade para o teste de técnicas de previsão não-lineares¹. Este ensaio investiga se os retornos dos mercados acionários de alguns países latino americanos podem ser previstos pelo meio das redes neurais artificiais (RNAs).

O ensaio segue a seguinte divisão. Depois desta introdução, a segunda seção resenha seletivamente o atual estágio da pesquisa relacionada à eficiência de mercados e sua extensão para mercados emergentes. A terceira seção, descreve a modelagem das redes neurais artificiais. A quarta parte, discute os resultados nos países que serão testados (Brasil, Argentina e México) enquanto a quinta discute a questão da lucratividade desta estratégia. A última parte será dedicada às conclusões.

¹ Ceretta e Costa Júnior (2002) e Torres, Bonomo e Fernandes (2002) sugerem a utilização de modelos não-lineares para a previsão do mercado acionário brasileiro.

1.2 .Resenha da Pesquisa sobre Eficiência de Mercados

1.2.1. Revisão dos Argumentos Teóricos

A primeira aplicação da hipótese de passeio aleatório para mercados financeiros era baseada na hipótese de que a aleatoriedade no preço dos ativos era devida a grandes grupos de investidores continuamente à procura de uma maneira de aumentar a sua riqueza.² Nesta visão, acreditava-se que o movimento dos preços de mercado era uma resposta direta a informações não antecipadas. Isto conduziu ao conceito de “mercado eficiente”³ em que o preço dos ativos reage rapidamente a novas informações, de forma que qualquer oportunidade sistemática de trading lucrativo é eliminada.

Um dos principais argumentos contra a hipótese de mercado eficiente foi escrito por Grossman (1976) e Grossman e Stiglitz (1980). Eles argumentam que mercados eficientes só podem existir na ausência de custos para adquirir informações. Se existe um custo de coleta, então não haveria razão econômica para fazê-lo, uma vez que os preços incorporariam toda a informação corrente. Também haveria pouca justificativa para se comprar e vender ativos, uma vez que, nenhum lucro poderia ser feito e assim pequeno acréscimo nos custos de transação poderia conduzir o mercado ao colapso.

Em função das lacunas empíricas da hipótese de mercados eficientes surgiram algumas teorias revisionistas.

² Samuelson(1965).

³ Fama (1965).

Uma das mais importantes é chamada de Hipótese da Eficiência de Mercado Relativa.

White (1988p. 470) resume o postulado desta corrente revisionista:

"Despite the strength of the simple efficient markets hypothesis, it is still only a theory, and any theory can be refuted with appropriate evidence. It may be that techniques capable of finding such evidence have not yet been applied. Furthermore, the theory is realistically mitigated by bounded rationality arguments. Such arguments hold that humans are inherently limited in their ability to process information so that efficiency can hold only to limits of human information processing. If a new technology (such as neural networks methods) suddenly becomes available for processing available information, then profit opportunities may arise for the processor of that technology. The technology effectively allows creation of a form of inside information".

Uma versão mais recente da hipótese da eficiência de mercados relativa declara que não é possível sistematicamente ganhar retornos em excesso sem algum tipo de vantagem competitiva (Lo e MacKinlay 1999). Neste trabalho, os autores argumentam que o funcionamento do mercado financeiro é semelhante à comercialização de novos produtos em outras indústrias onde patentes provêm um tempo em que se pode gerar retornos excessivos até que os outros participantes passem a copiar o produto. Apesar dos produtos no mercado financeiro não serem patenteados, existe uma defasagem em que é possível se apropriar de estratégias que gerem lucros. Assim, considera-se que este efeito torna o mercado relativamente eficiente enquanto que os retornos potenciais associados à

inovação provêem suficiente motivação para o desenvolvimento de novas formas de vantagem competitiva.

1.2.2. Argumentos empíricos

Depois que as hipóteses de passeio aleatório e eficiência de mercados foram apresentadas na década de 60, a pesquisa empírica começou a tentar estabelecer a validade destas teorias e a testar a viabilidade das estratégias de trading mais usadas. Uma vasta literatura tem-se acumulado desde então, e pretende-se fazer uma breve resenha ilustrativa a respeito deste assunto. A hipótese de passeio aleatório implica que nenhuma regularidade deva existir em um mercado financeiro que funcione “corretamente”.

Existem três formas de eficiência relatadas pela literatura. Fraca, semiforte e forte. A forma fraca da teoria dos mercados eficientes considera informação apenas de preços históricos, a semiforte considera toda informação pública disponível e a forte considera toda a informação pública e privada disponíveis.

A maior parte dos testes tem se concentrado na forma fraca da teoria dos mercados eficientes, os se dividem em duas categorias. A primeira implementa testes estatísticos em dados históricos para identificar padrões significativos e a segunda examina regras de trading para determinar se estas podem gerar retornos anormais depois dos custos de transação. Um tópico em particular que foi foco de muitos dos trabalhos iniciais considerou a análise dos efeitos da sazonalidade de diferentes dias da semana e meses do ano como forma de comportamento previsível. Uma das primeiras anomalias da

forma fraca da Teoria dos Mercados Eficientes (TME) foi identificada como “efeito fim de semana” (French(1980)).

Em outra área de pesquisa, foram conduzidos testes simples de previsibilidade para examinar se retornos passados podem prever retornos futuros. Um número grande de estudos examina a correlação de primeira ordem entre os retornos ao longo do tempo, entre 1 e 3 meses em vários mercados acionários(Fama (1965), Cootner (1974)). Em geral, no entanto, os resultados não mostraram uma significativa correlação e estudos argumentam que correlações não podem ser usadas para determinar eficiência pela influência de “outliers” (Fama (1965), Jennergren e Korvold (1975))

Em contraste com estes resultados negativos, a hipótese de passeio aleatório tem sido testada comparando-se os estimadores de variância em diferentes freqüências de tempo (Lo e MacKinlay (1988 e 1989). Os resultados para retornos semanais de ações em um período de 23 anos (1962-1985) fortemente rejeitam a hipótese de passeio aleatório para vários índices agregados e portfolios. Atribui-se este fenômeno ao comportamento de ações de valor pequeno e necessariamente ao pequeno número de transações de ações ilíquidas ou a volatilidades variantes no tempo.

Um outro grupo de testes empíricos advoga que é possível pesquisar relações não lineares nos retornos de forma indireta, desenvolvendo complexas regras baseadas em movimentos de preços históricos. Um exemplo comum é a quebra de suportes e resistências (Fama e Blume (1966) Jennergren e Korsold (1975) e Curcio e Goodhart (1992)). Estes formulam uma estratégia de trading que vende quando os preços rompem uma barreira inferior (suporte) e compram quando os preços rompem uma barreira superior (resistência). Os resultados mostram a evidência de alguma lucratividade. Outras

regras, como ombro-cabeça-ombro e regras de força relativa também tem sido testadas⁴ e existem evidências de que geram lucros acima do benchmark de mercado.⁵

1.2.3. Testes de previsibilidade

A hipótese de passeio aleatório para os preços dos ativos pode ser decomposta em três diferentes versões dependendo da forma das hipóteses subjacentes. A forma forte da hipótese assume que os retornos dos ativos têm incrementos independentes e identicamente distribuídos. A hipótese de independência implica que os incrementos são não correlacionados e que as funções não lineares dos incrementos também são não correlacionadas. No caso das ações, os preços dos ativos são sempre positivos assim a definição precisa de modificações. Assume-se que os incrementos dos preços log normais seguem um passeio aleatório.

No entanto, este modelo de preços dos ativos não é realista quando comparado com o comportamento passado dos mercados. Por exemplo, no longo prazo os mercados adaptam-se através do tempo quando mudanças econômicas e políticas acontecem. Isto pode ser visto observando-se a heterocedasticidade dos retornos dos ativos. Este modelo de incrementos identicamente distribuídos é uma hipótese considerada muito forte da versão de passeio aleatório para ser aplicada a dados financeiros.

Em uma segunda forma, a hipótese de passeio aleatório assume que os incrementos são independentes, mas não identicamente distribuídos. No entanto, parece extremamente difícil testar a independência estatística sem assumir distribuições

⁴ Para uma resenha atualizada ver o ensaio 2

⁵ Ver Brock et al. (1992) e Le Baron (1996)

idênticas.⁶ Alguns testes não paramétricos têm sido desenvolvidos (rank tests) mas ainda precisam de alguma hipótese restritiva a respeito da distribuição. A falta de teste estatísticos potentes para esta versão da hipótese do passeio aleatório tem conduzido ao aumento da pesquisa em torno do desenvolvimento de testes “econômicos de previsibilidade”. O âmago destes testes é indiretamente medir a previsibilidade analisando-se a performance de simples estratégias de trading para a previsão dos retornos de ativos.

Em geral, estes testes de previsibilidade têm tomado a forma de um filtro simples que é aplicado a séries de retorno de um ativo. O retorno total gerado por uma estratégia de alocação de ativos, implementada com base em uma regra de trading pode então ser considerado como medida da previsibilidade dos retornos dos ativos. Em estudos recentes formas mais avançadas de regras tem sido desenvolvidas numa classe de gestão de ativos chamada de análise técnica. A hipótese básica desta estratégia é que preços históricos e outras estatísticas de mercados exibem regularidades que formam padrões que podem ser explorados lucrativamente e permitem extrapolar movimentos futuros de preços.

Uma terceira versão da hipótese de passeio aleatório assume que os preços dos ativos podem ter incrementos dependentes, mas não correlacionados em todas as defasagens possíveis. Isto é testado sob a hipótese nula de que todos os coeficientes de autocorrelação são zero.

⁶ Campbell, Lo & MacKinlay (1997) p.41

1.3. Descrição do Procedimento do Teste

1.3.1. Aspectos Teóricos

As redes neurais artificiais (RNAs) são uma classe de modelos não paramétricos generalizados inspirados em estudos do cérebro e do funcionamento do sistema nervoso. A vantagem comparativa de uma RNA em relação a modelos econométricos convencionais é que ela pode aproximar qualquer função não linear (ou linear) com um arbitrário grau de precisão com um número razoável de unidades escondidas através da composição de uma rede com funções relativamente simples.⁷

Com relação à pesquisa no campo da macroeconomia, Qi (2001) examina a relevância de diferentes indicadores econômicos e financeiros na previsão das recessões americanas através de modelos de redes neurais, com a hipótese de que os ciclos de negócios são assimétricos e não podem ser adequadamente previstos por modelos lineares. Shaaf (2000) compara os modelos de RNAs com os tradicionais de econometria para prever recessões utilizando a curva de juros americana e conclui que a utilização de RNAs fornece uma previsão mais acurada com menor erro e variação do que uma regressão.

No entanto, apesar dos bons resultados obtidos, outras pesquisas indicam o oposto. Stock e Watson (1998) testam diferentes modelos não lineares na previsão de 215 séries macroeconômicas americanas mensais e chamam a atenção para a pobreza das previsões em séries econômicas por parte das RNAs.

⁷ White(1990) e White, Galant, Hornik, Stinchcombe e Wooldridge (1992).

As RNAs têm um forte apelo em aplicações financeiras em que existem fortes não linearidades nas séries e poucos modelos que sejam capazes de lidar com esta questão. White (1988) descobriu, ao tentar prever os retornos diários da IBM, que apesar de um bom ajustamento dentro da amostra, os resultados fora da amostra das estimações com RNAs, eram pobres. Para retornos mensais do índice da New York Stock Exchange, Chuah (1993) descobriu que as RNA não tinham poder de acertar as oscilações de mercado (market timing) bem como que seus erros não diferiam muito dos gerados por modelos lineares. Estes resultados positivos são encontrados em Martin e Tan (1992), Gençay (1998), Gençay e Stengos (1998) e Frances e VanDijk (2000), cuja a resenha do ensaio 2 detalha com mais cuidado.

A contradição dos resultados de alguns trabalhos pode ser creditada às diferentes metodologias implementadas na escolha da função de ativação e do número de neurônios na camada escondida. No entanto, as previsões de uma rede podem ser capazes de gerar ganhos extraordinários se comparados com suas contrapartes lineares bem como com estratégias passivas (buy and hold).

A despeito de várias características desejáveis, a construção de uma rede neural para uma aplicação particular não é trivial. Os principais problemas para a implementação de uma RNA são os comuns aos outros modelos não-paramétricos, ou seja, a possibilidade de overfitting⁸. No caso mais específico da RNA, existe um número de parâmetros relativamente grande a serem determinados pela não linearidade inerente nessas especificações; a função objetiva possivelmente não será globalmente convexa e

⁸ Como definido em Campbell, Lo & Mackinlay (1997) p.523, "*Overfitting occur when a model fit "too well", in the sense that the model has captured both random noise as well as genuine nonlinearities. Heuristically, the primary source of overfitting is having too few "degrees of freedom" or too many parameters relative to the number of data points, and a typical symptom is an excellent in-the-sample fit but poor out-of-sample performance*".

pode ter vários mínimos locais. A construção de uma RNA envolve dessa forma, a escolha de uma arquitetura apropriada (número de camadas, número de neurônios na camada escondida), a seleção das funções de transferência das camadas escondida e de saída, o algoritmo de treinamento e os pesos iniciais. Para este ensaio será usada uma rede com três camadas e variações do algoritmo de retropropagação.

A apresentação padrão do algoritmo de retropropagação pode seguir Haykin (1999) ou Frances e VanDijk (2000)⁹ e que pode ser resumida da seguinte forma.

Seja um modelo RNA(p,q,l)

$$y_i = x_i' \phi + \sum_{j=1}^q \beta_j G(x_j' \gamma_j) + \varepsilon_i \quad (1.1)$$

onde y_i é a variável endógena (output)

$x_i' \phi$ é o vetor de inputs na camada de entrada

$G(.)$ função de ativação

β pesos gerados na camada escondida

Seus parâmetros podem ser estimados minimizando-se a função da soma dos quadrados dos resíduos.

$$Q_n(\theta) = \sum_{i=1}^n [y_i - F(x_i; \theta)]^2, \quad (1.2)$$

onde

$$F(x_i; \theta) = x_i' \phi + \sum_{j=1}^q \beta_j G(x_j' \gamma_j)$$

⁹ Estes últimos seguem Kuan e White(1994).

e θ é o vetor que consiste de $p+1+q(p+2)$ parâmetros em $\phi, \beta_1, \dots, \beta_q, \gamma_1, \dots, \gamma_q$.

Sob condições gerais, $\hat{\theta}_n$, converge para θ^* definido como

$$\theta^* = \arg \min_{\theta} E\left[\sum_i (y_i - F(x_i; \theta))^2\right], \quad (1.3)$$

com o tamanho da amostra n aumentando sem limites.

O estimador normalizado $\sqrt{n}(\hat{\theta}_n - \theta^*)$ converge para a distribuição normal multivariada com média zero e uma matriz de covariância que pode ser estimada consistentemente.

Qualquer algoritmo convencional de mínimos quadrados pode ser empregado para obter $\hat{\theta}_n$. Dadas as estimativas na r -ésima interação, $\hat{\theta}_n^{(r)}$, computa-se a soma dos quadrados dos resíduos $Q_n(\hat{\theta}_n^{(r)})$ e o gradiente

$$\nabla Q_n(\hat{\theta}_n^{(r)}) = \frac{\partial Q_n(\hat{\theta}_n^{(r)})}{\partial \theta} \quad (1.4)$$

para obter a nova estimativa $\hat{\theta}_n^{(r+1)}$

$$\hat{\theta}_n^{(r+1)} = \hat{\theta}_n^{(r)} - \lambda A(\hat{\theta}_n^{(r)})^{-1} \nabla Q_n(\hat{\theta}_n^{(r)}), \quad (1.5)$$

onde λ é o tamanho do passo da interação e $A(\hat{\theta}_n^{(r)})$ é uma matriz que pode ser uma função dos parâmetros θ . Por exemplo, no algoritmo Newton-Rapson, $A(\hat{\theta}_n^{(r)})$ é tomado

como a matriz hessiana. Um método bastante popular na literatura de redes neurais é denominado *descida íngreme (steepest descent)*, em que $A(\hat{\theta}_n^r)$ é igual a matriz identidade¹⁰. Este método é chamado de retropropagação de erros, que vem do fato de que os diferentes elementos do vetor gradiente $\nabla Q_n(\hat{\theta}_n^r)$ podem ser computados recursivamente. Por exemplo, a derivada parcial de $Q_n(\theta)$ com respeito a β_j é dada por

$$\frac{\partial Q_n(\theta)}{\partial \beta_j} = -2 \sum_{i=1}^n [y_i - F(x_i; \theta)] G(x_i' \gamma_j) \quad (1.6)$$

onde a derivada parcial de $Q_n(\theta)$ com respeito a $\gamma_{i,j}$ é iguala

$$\begin{aligned} \frac{\partial Q_n(\theta)}{\partial \gamma_{i,j}} &= -2 \sum_{i=1}^n [y_i - F(x_i; \theta)] \beta_j G(x_i' \gamma_j) [1 - G(x_i' \gamma_j)] x_{i,j} = \\ &= -2 \sum_{i=1}^n \frac{\partial q_i(\theta)}{\partial \beta_j} \beta_j [1 - G(x_i' \gamma_j)] x_{i,j} \end{aligned} \quad (1.7)$$

em que $q_i(\theta) = [y_i - F(x_i; \theta)]^2$ denota o resíduo quadrado para a i -ésima observação. O processo de estimação então é o que se segue. Dadas as estimativas $\hat{\theta}_n^{(r)}$, computa-se o valor das unidades escondidas $\hat{\beta}_j^{(r)} G(x_i' \hat{\gamma}_j^{(r)})$, o valor ajustado $F(x_i; \hat{\theta}_n^{(r)})$ e o resíduo $\hat{\varepsilon}_i^{(r)} = y_i - F(x_i; \hat{\theta}_n^{(r)})$

Em seguida, os resíduos são usados para calcular as derivadas parciais com respeito a $\hat{\theta}_n^{(r)}$ que depois são usadas para a obtenção das derivadas parciais acima. Finalmente as estimativas são atualizadas como $\hat{\theta}_n^{(r+1)} = \hat{\theta}_n^{(r)} - \lambda \nabla \theta_n(\hat{\theta}_n^{(r)})$. Assim cada

¹⁰ Haykin (1999) apresenta apenas esta caso.

iteração no algoritmo consiste em primeiro ir para frente através da rede e depois retropropagar novamente.

Para a avaliação e seleção do melhor modelo considere novamente o modelo RNA(p,q,1) para y_t

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \sum_{j=1}^q \beta_j G(\gamma_{0,j} + \gamma_{1,j} y_{t-1} + \gamma_{2,j} y_{t-2} + \dots + \gamma_{p,j} y_{t-p}) + \varepsilon_t \quad (1.8)$$

que contém $p+1+q(p+2)$ parâmetros a serem estimados. A implementação de um modelo RNA(p,q,1) requer que algumas decisões sejam tomadas:

- Escolha da função de ativação $G(\cdot)$;
- Escolha do algoritmo de aprendizado;
- Escolha do número de neurônios escondidos q ;
- Escolha do número de defasagens p que serão utilizadas como input.

1.3.2.Implementação do Procedimento

1.3.2.1.Descrição das Séries

As séries utilizadas foram fornecidas pela Economática e pela Bloomberg e compõem-se do logaritmo dos retornos diários em dólar dos índices de ações dos seguintes países: Brasil (IBovespa), Argentina (Merval) e México(Mexbol) , para o período de 03 de janeiro de 1993 a 24 de agosto de 2002.

Gráfico 1.1

Série do Índice Bovespa denominado em dólares – diário (3/01/93-24/07/02)



Gráfico 1.2

Série dos Retornos do Índice Bovespa denominado em dólares- diário (3/01/93-24/7/02)

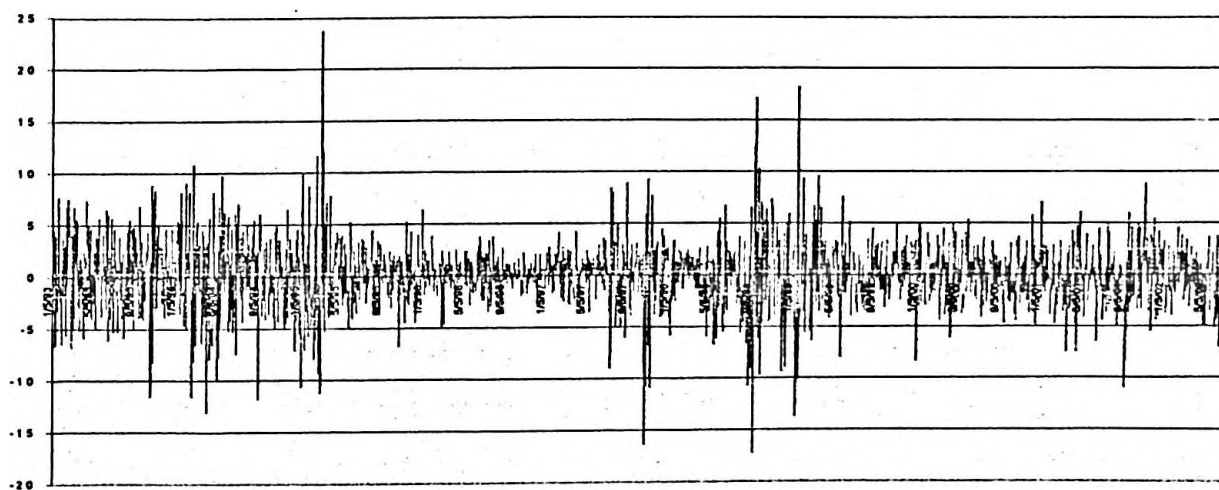


Gráfico 1.3

Série do Índice Merval denominado em dólares – diário (03.01/93-24/07/02)



Gráfico 1.4

Série dos Retornos do Índice Merval denominado em dólares – diário (03/01.93-24/07/02)

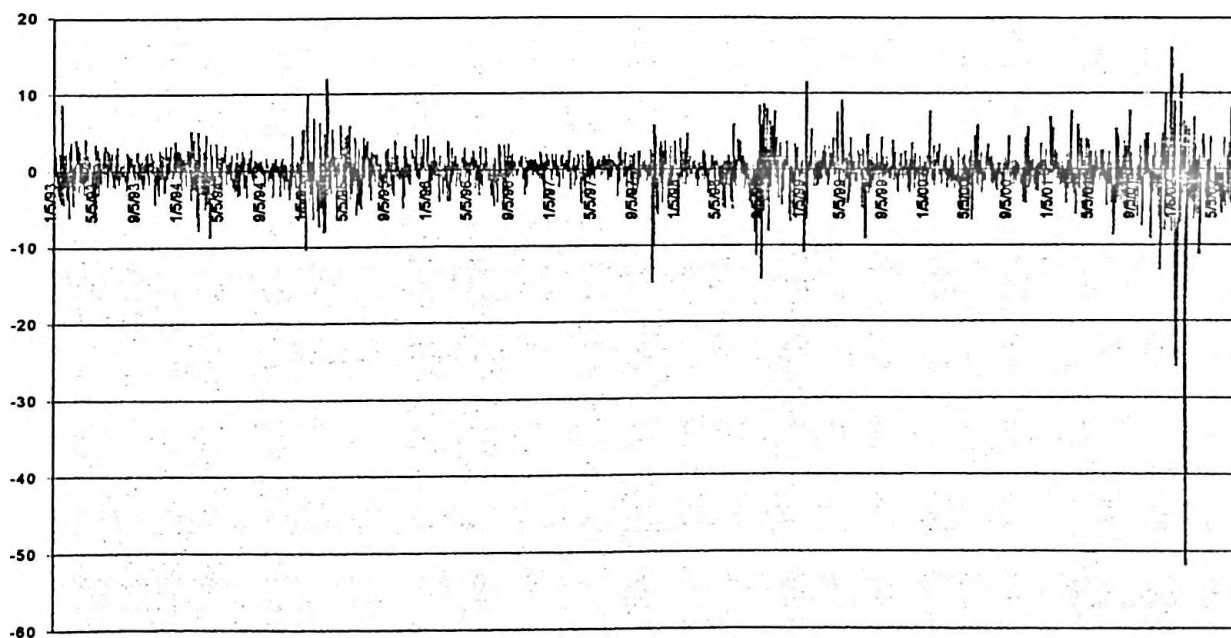


Gráfico 1.5

Série do Índice (MexBolsa) denominado em dólares – diário (03/01/93-24/07/02)



Gráfico 1.6

Série dos Retornos do Índice MexBolsa denominado em dólares– diário
(03/01.93-24/07/02)

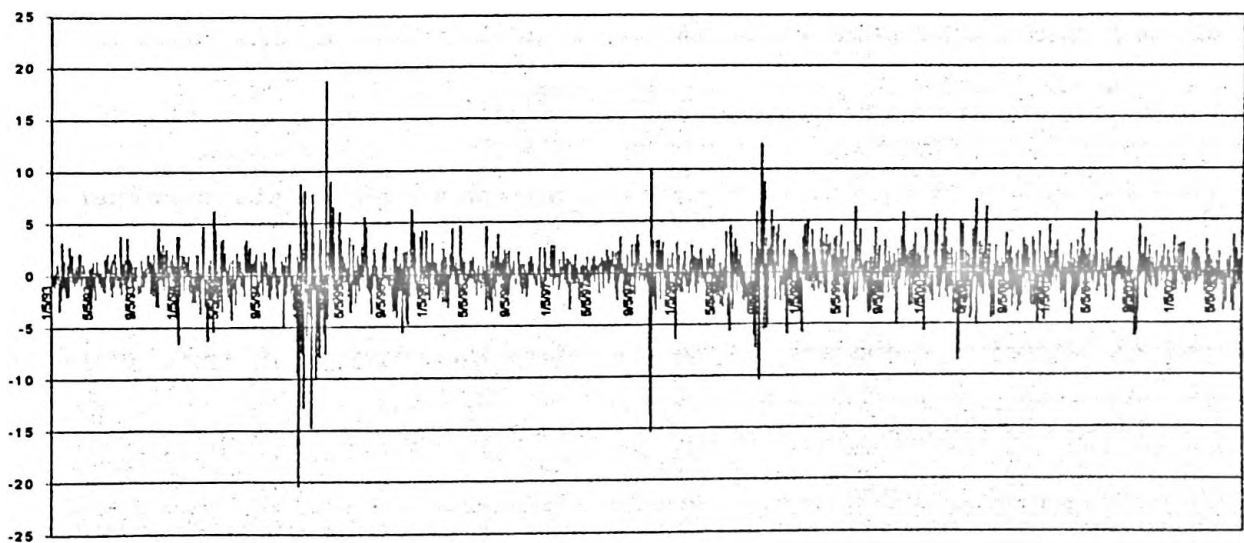


Tabela 1.1: Resumo das Estatísticas e dos Testes de Não-Linearidade

Variável	Média	DP	Min/Max	Assimetria	Curtose	Teste JB	LB(10)	BDS (M=2, e=sigma)	BDS (m=2 e=sigma/2)
Brasil	0.049612	3.057151	-17.2462 23.71758	0.018356	7.656815	2091.014	63.027	12.05024	11.14055
Argentina	-0.06101	2.726025	-51.8373 16.11652	-3.367869	64.48468	370618.2	39.52	12.9061	10.5055
México	0.013267	2.083918	-20.329 18.73566	-0.430867	14.49036	12983.89	71.92	11.59328	10.7179

Nota: (1) O termo “média” refere-se a média aritmética dos retornos em dólar
 (2) O termo DP refere-se ao desvio-padrão da série.
 (3) Teste JB refere-se ao teste de normalidade de Jarque-Bera.
 (4) LB(10) é a estatística Ljung-Box na defasagem 10
 (5) As duas últimas colunas referem-se a estatística BDS para testar a não linearidade da série.

Como fica evidente na tabela 1.1 as variáveis são assimetricamente distribuídas. No caso brasileiro essa assimetria não é grande,¹¹ porém todas as séries apresentam “caudas pesadas”, o que é uma característica de muitas séries financeiras. Dessa forma, a probabilidade da observação empírica de retornos extremos é maior do que a que se verificaria em uma distribuição normal. O excesso de curtose indica mais peso nas caudas da distribuição do que na distribuição normal. A normalidade de todas as séries é rejeitada pelo teste de Jarque-Bera. Todas as séries demonstram alto grau de não linearidade medida pelo teste BDS¹², que foi realizado considerando duas dimensões como parâmetros de vizinhança de ϵ e $\epsilon/2$.

O teste BDS é baseado no conceito de integral de correlação¹³. O procedimento toma uma série de vetores com m-dimensões de uma série de tempo, no períodos $t=1,2,\dots,T-m$, em que T é o tamanho da série. Iniciando no período $t=1$ e $s=t+1$, os pares

¹¹ Essa assimetria mostra-se não significativa dependendo da probabilidade do teste qui-quadrado.

¹² Brock (1986), Brock, Hsieh e Le Baron (1991), Brock e Dechert (1991) e Dechert (1996).

¹³ McNellis (2003p.45)

(x_t^m, x_s^m) são avaliados por uma “função indicadora” para checar se sua distância máxima, no horizonte m , é menor do que um valor especificado ε . A integração de correlação, mede a fração dos pares que estão na “região de tolerância” da “ dimensão” m . A estatística BDS testa a diferença entre a integral de correlação para a dimensão m , e a integral para a dimensão 1 elevada a m .

Sob a hipótese de independência, esta probabilidade será simplesmente o produto das probabilidades individuais para cada par. Dessa forma se as observações são independentes,

$$c_m(\varepsilon) = c_1^m(\varepsilon) \quad (1.9)$$

Quando se trabalha com dados amostrais, $c_1(\varepsilon)$ e $c_m(\varepsilon)$ não são observados diretamente. Só podem ser estimados da amostra.. Como resultado, não se espera que esta relação mantenha-se exatamente, mas apenas com algum erro. Quanto maior o erro, menor a probabilidade de que este erro seja causado por variações aleatórias da amostra. O teste BDS fornece uma base formal para o julgamento do tamanho do erro.

Para estimar a probabilidade de uma dimensão particular, simplesmente percorre-se todos os possíveis conjuntos daquela distancia que podem ser desenhados partindo da a mostra e contar-se o número de conjuntos que satisfazem a condição ε . A razão entre o número de conjuntos que satisfazem a condição dividida pelo número total de conjuntos fornece a estimativa da probabilidade. Dada uma a mostra de n observações de uma série X , pode-se colocar em notação matemática

$$c_{m,n}(\varepsilon) = \frac{2}{(n-m+1)(n-m)} \sum_{s=1}^{n-m+1} \sum_{t=s+1}^{n-m+1} \prod_{j=0}^{m-1} I_{\varepsilon}(X_{s+j}, X_{t+j}) \quad (1.10)$$

onde I_ε é a função indicadora

$$I_\varepsilon(x, y) = \begin{cases} 1 & \text{para } |x - y| \leq \varepsilon \\ 0 & \text{caso contrario} \end{cases} \quad (1.11)$$

Note que a estatística $c_{m,n}$ são chamadas de intergral de correlação.

Pode-se então utilizar estas probabilidades estimadas da amostra para construir um teste estatístico para independência.

$$b_{n,m}(\varepsilon) = c_{m,n}(\varepsilon) - c_{1,n-m+1}(\varepsilon)^m \quad (1.12)$$

onde o segundo termo descarta a última observação $m-1$ da amostra de forma que se baseia no mesmo número de termos da primeira estatística.

Sob a hipótese de independência, espera-se que esta estatística seja próxima a zero.

Em outras palavras, o propósito da estatística BDS é testar a hipótese nula de que a série é um ruído branco. Mills(1999) chama de um teste de diagnóstico, uma vez que a rejeição da hipótese nula é consistente com algum tipo de dependência na série, que pode ser resultado de um sistema estocástico linear, um sistema estocástico não linear ou um sistema determinista não linear. Os resultados confirmam a evidência de não linearidade nos retornos das séries a serem pesquisadas.

1.3.2.2. Escolha da Função da Ativação

O programa usado, Neuroshell 2¹⁴, permite que o usuário escolha entre diversos métodos que transformam as variáveis na camada de entrada bem como a escolha das funções de ativação na camada escondida e de saída. Dessa forma, optou-se por testar seis diferentes transformações na camada de entrada contra 4 diferentes funções de ativação na camada escondida, mantendo a função logística na camada de saída¹⁵.

Quando as variáveis são carregadas em uma rede neural, elas devem estar em uma escala de forma que a rede possa processar estes dados de maneira eficiente. Dessa forma, pode-se optar por transformações que mapeiem os dados no intervalo entre [0,1] ou [-1,1]. Alternativamente, optou-se por testar duas funções de escala não-lineares: logística e tangente. A função logística coloca os dados no intervalo de [0,1] de acordo com a seguinte fórmula:

$$f(v)=1/(1+\exp(-(v-m)/sd)) \quad (1.13)$$

onde m é a média de todos os valores da variável e sd é o desvio padrão.

A função tangente coloca em uma escala (-1, 1) de acordo com:

¹⁴ As referências bibliográficas que baseiam os algoritmos do NeuroShell são Hutton (1992), Lang e Whitbrock (1989), Nishikawa, Kita e Kawamura (1990), Rumelhart e McClelland (1986), Simpson (1990), Specht (1988,1990), Touretzky e Pormenteau (1989) e Wasserman (1989).

¹⁵ No caso específico do Ibovespa foram realizados testes em que se combinavam as 6 transformações na camada de entrada com as 8 funções de transformação na camada escondida, mantendo a função logística na camada de saída. No total, foram estimadas 2400 redes diferentes para o Ibovespa 1800 redes diferentes para o Merval Argentino e 1800 para a índice de ações da bolsa mexicana.

$$f(v)=\tanh((v-m)/sd) \quad (1.14)$$

onde \tanh é uma tangente hiperbólica.

Ambas as funções poderão ser efetivas na redução dos efeitos dos “outliers”.

Após a escolha das transformações na camada de entrada, escolhe-se a função de ativação na camada escondida. A maior parte dos trabalhos acadêmicos reporta resultados utilizando-se da função logística¹⁶. No entanto, outras escolhas são possíveis tais como \arctan , \tanh e \coseno .¹⁷ Neste ensaio testaremos não apenas esta, como também a função tangente, as funções gaussianas e gaussianas complementares. As fórmulas são as seguintes:

logística -- $f(x)=1/(1+\exp(-x))$

linear -- $f(x)=x$

\tanh -- $f(x)=\tanh(x)$

$\tanh 1.5$ -- $\tanh(1.5x)$

seno -- $\sin(x)$

logística simétrica -- $2/(1+\exp(-x))-1$

Gaussian -- $\exp(-x^2)$

Gaussian-complementar -- $1 - \exp(-x^2)$

¹⁶ Adya & Collopy (1998)

¹⁷ Hertz, Krog e Palmer (1991).

1.3.2.3. Escolha do Algoritmo de Aprendizado

O treinamento é um processo de ensino do que se quer que a RNA aprenda. Os dados podem ser apresentados à rede de forma aleatória ou seqüencial.

Como apresentado na seção teórica, o algoritmo de retropropagação pode ou não conter um termo de momentum. Foram testados três algoritmos possíveis. Primeiro, um sem o termo de momentum. Um segundo que inclui o termo de momentum, ou seja, um termo de aceleração do gradiente descendente. A taxa de aprendizado escolhida foi de 0.05 enquanto o termo de momentum escolhido foi de 0.5.¹⁸

Um algoritmo proprietário, chamado de Turbo Pro, também foi testado. Neste caso, o algoritmo não é sensível à determinação da taxa de aprendizagem e do momentum. O treinamento se dá através de uma época¹⁹ inteira antes que os pesos sejam atualizados. O algoritmo utiliza uma atualização de pesos independente para cada peso diferente, ao contrário dos métodos usuais que utilizam a mesma taxa de aprendizado e momentum para todos os pesos. Em adição, o tamanho do passo é ajustado adaptativamente com o progresso do aprendizado²⁰.

¹⁸ Segue Gately (1996).

¹⁹ Haykin (1999) define época como a apresentação completa do conjunto de treinamento que está sendo processado.

²⁰ Ward Systems (1996)

1.3.2.4. Escolha do número de neurônios escondidos e dos números de defasagens relevantes

Frances e Van Dijk(2000) sugerem duas estratégias para a escolha dos valores apropriados para os neurônios escondidos e para o número de defasagens. A primeira seria estimar todos os modelos possíveis e utilizar um critério de seleção, como o AIC, BIC ou Hannan-Quinn. Uma segunda alternativa seria determinar a ordem das defasagens através da escolha de um modelo AR(p) e depois determinar o número de neurônios escondidos através do critério de informação

Kaashoek e Van Dijk (1998, 2001) propõem um método de comparação de redes baseado no coeficiente de correlação dos valores estimados \hat{y}_t , definido por

$$R^2 = 1 - \frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{\sum_{t=1}^T (y_t - \bar{y})^2}, \quad (1.15)$$

onde

$$\bar{y} = (1/T) \sum_{t=1}^T y_t$$

A idéia é remover a i -ésima unidade escondida, $i=1, \dots, q$, reestimar o modelo e computar o valor percentual da diferença no R^2 , definido por

$$\Delta R^2 = \frac{R^2 - R_i^2}{R^2}, \quad (1.16)$$

onde R^2_i é o R^2 sem a i -ésima unidade escondida. Dessa forma se ΔR^2 é pequena ($<1\%$), remove-se a unidade do modelo. Depois de remover as unidades escondidas, o processo continua removendo-se cada uma das variáveis explicativas e comprando-se seus coeficientes de correlação. As variáveis irrelevantes são removidas do modelo utilizando-se o mesmo procedimento de redução de unidades escondidas, baseado na variação do R^2 . O principal problema desta estratégia é o tempo de estimação quando se tem um grande número de variáveis.

Para o número de defasagens optou-se então pela escolha de um “número grande de defasagens” de forma crescente. Para o número de neurônios escondidos utilizou-se uma regra baseada em Gately (1996) igualando ao número de variáveis explicativas.

Ward Systems (1996) apresenta uma formulação diferente para a escolha do número de neurônios na camada escondida²¹:

$$No\ de\ neuronios = 1/2(inputs + outputs) + \sqrt{No\ de\ padrões\ de\ treinamento} \quad (1.17)$$

Dessa forma, procedeu-se utilizando uma estratégia do tipo “step wise” nos moldes de Swanson e White (1995,1997a, 1997b) e Frances e Van Dick (2000). Primeiro estimou-se o modelo dentro da amostra contra a primeira defasagem e com apenas um neurônio na camada de entrada. Testou-se todas as combinações de seis funções na camada de entrada contra cada uma das quatro funções de ativação na camada escondida. Em seguida, repete-se o mesmo processo considerando as defasagens um e dois e

²¹ Esta regra piora os resultados em relação à regra mais simples que iguala o número de neurônios da camada escondida ao de unidades da camada de entrada.

igualando o número de neurônios na camada escondida ao número de inputs. O processo continua até que sejam testadas nove defasagens.

1.4. Resultados

O período da amostra vai de 3 de janeiro de 1993 até 24 de julho de 2002 e contém 2357 observações.

Primeiro, a amostra foi dividida em duas partes. A primeira de 3 de janeiro de 1993 a 22 de maio de 2002 e a outra, para os testes de previsão, vai de 23 de maio de 2002 a 24 de julho de 2002.

A primeira parte da amostra foi novamente dividida em duas. 80% configuram o conjunto de treinamento e 20% o conjunto de teste das diferentes arquiteturas.

O procedimento continua, escolhendo-se a melhor combinação de defasagens, neurônios e funções de transformação para a amostra que vai de 3 de janeiro de 1993 até 22 de maio de 2002. Escolhem-se os melhores modelos pelos diferentes critérios. Para o teste foram utilizados dois critérios. O primeiro de informação (AIC, BIC e HQ), em que

$$AIC = \ln(\hat{\sigma}^2) + \frac{2k}{T} \quad (1.18)$$

$$SBIC = \ln(\hat{\sigma}^2) + \frac{k}{T} \ln T \quad (1.19)$$

$$HQIC = \ln(\hat{\sigma}^2) + \frac{2k}{T} \ln(\ln(T)) \quad (1.20)$$

onde σ^2 é a variância residual (equivalente à soma dos resíduos ao quadrado dividida pelo número de graus de liberdade $(n-k)$ onde $k = p+1+q(p+2)$, p é o número de defasagens e q o número de neurônios da camada escondida. O segundo critério é do menor Erro Quadrático Médio (MSE).

Com isso pode-se gerar diferentes arquiteturas que serão testadas fora da amostra.

O procedimento fora da amostra é realizado da seguinte forma. A rede é modificada de forma a prever o log do retorno dos índices de ações em dólar em $t+1$, utilizando-se as defasagens anteriormente escolhidas. Prossegue-se estimando durante o período da amostra que se inicia em 23 de maio de 2002 reestimando-se para cada dia, quando a nova estimativa incorpora o dado efetivamente realizado no dia anterior. Repete-se o processo até o fim da amostra²².

Para a escolha da melhor arquitetura fora da amostra foram utilizadas cinco tradicionais medidas de ajustamento e previsão para os modelos alternativos:

$$1. RMSE = \sqrt{1/T \sum_{i=1}^T (y_i - \hat{y}_i)^2}$$

A raiz do erro quadrático médio entre os valores efetivos e os previstos.

$$2. MAE = 1/T \sum_{i=1}^T |y_i - \hat{y}_i|:$$

O erro absoluto médio entre os retornos efetivos e os previstos.

$$3. MAPE = 1/T \sum_{i=1}^T |(y_i - \hat{y}_i) / y_i|:$$

A percentagem absoluta de erro entre os retornos efetivos e os previstos.

$$4. Sinal = 1/T \sum z_i :$$

se \hat{y} e y tem o mesmo sinal então $z = 1$

senão $z = 0$

A proporção de sinais que é corretamente prevista.

Tabela 1.2 :Estatísticas dentro da amostra(3/01/93 a 22/5/02)

	R ²	r ²	MSE	AIC	BIC	HQ
Brasil						
RNA(1,1,1)	0.014	0.015	9.194	2.225	2.237	2.230
RNA(7,7,1)	0.048	0.058	8.862	2.272	2.448	2.336
RNA(6,6,1)(1)	0.034	0.037	8.996	2.261	2.398	2.311
Argentina						
RNA(1,1,1)	0.015	0.021	7.315	1.996	2.008	2.000
RNA(7,7,1)	0.040	0.054	7.135	2.054	2.230	2.118
México						
RNA(1,1,1)	0.031	0.032	4.181	1.434	1.449	1.441
RNA(6,6,1)	0.069	0.070	4.024	1.461	1.596	1.510

Notas:

1) RNA degenerada com defasagens 1,5,6,8,9,17

2) Apesar de serem representadas de forma idêntica as RNAs(1,1,1) para o Brasil, Argentina e México tem diferentes funções de ativação na camada escondida, bem como diferentes algoritmos de aprendizado. Para uma descrição completa das arquiteturas destas redes, ver o apêndice 1.

Para representar a estrutura da RNA utiliza-se a forma RNA(p,q,l) onde p é o número de unidades da camada de entrada, q é o número de unidades da camada escondida e l é o número de unidades da camada de saída. Esta forma de representação serve para uma estrutura de três camadas.

²² Este procedimento é chamado de validação cruzada recursiva. Ver Diebold (1998, p.117)

A tabela 1.2 reporta as melhores redes pelos diferentes critérios de escolha. Para o Brasil(Ibovespa), o menor MSE pertence à rede RNA(7,7,1)²³. Já a melhor rede escolhida utilizando-se um dos critérios de informação é a que tem apenas uma defasagem e um neurônio na camada escondida²⁴. Para a Argentina (Merval) e México (MexBolsa) a tabela reporta as duas melhores arquiteturas pelos diferentes critérios. Pelo critérios de informação (AIC,BIC e HQ) as melhores redes são as que contêm apenas uma defasagem como variável explicativa. Isso ocorre em função da penalização que estas estatísticas exercem sobre o número de variáveis explicativa e neurônios na camada escondida. Quando o critério usado é o MSE, mais defasagens são permitidas. No entanto, isto não significa um melhor desempenho fora da amostra.

Frances e VanDick(2000) sugerem que a escolha das defasagens seja primeiro feita estimando-se um modelo ARIMA convencional. Se dentro da amostra esta estratégia apresenta um desempenho satisfatório, fora da amostra a performance é decepcionante em relação às outras duas possibilidades (critérios de informação e menor MSE)²⁵, como pode ser visto na tabela 1.3. A RNA que se utiliza da metodologia de Frances e VanDick (2000) é a RNA(6,6,1) cujas as defasagens são 1,5,6,8,9,17. O resultado da estatística Sinal (SIGN) para as melhores redes está em linha com a literatura internacional.²⁶

²³ Para as características de cada uma das arquiteturas ver apêndice 1.

²⁴ Foram estimadas diferentes arquiteturas, funções diferentes para cada combinação de defasagens de 1 a 9.

²⁵ O apêndice 2 mostrar o desempenho da RNA(1,1,1) tanto no conjunto de treinamento quanto no conjunto de teste.

²⁶ Trippi e DeSieno (1992) estimando a performance de 6 redes alternativas para o índice TOPIX obtiveram para a melhor rede 52.8% de sinais certos.

Para o caso brasileiro, as 3 melhores redes apresentam uma performance positiva, já no caso argentino apenas a escolhida através do critério de informação conseg

ue manter um bom desempenho. Isso mostra a importância de algum critério diferente do Erro Quadrático Médio na escolha da melhor arquitetura. Bishop (1995) e McNellis(2003) chamam a atenção para a necessidade de um critério que penalize o número de parâmetros de forma a aumentar a generalização das RNAs. A simples minimização da função de erro é incapaz de determinar o número ótimo de parâmetros de uma RNA.

As redes estimadas para a série do MexBolsa têm um desempenho muito positivo quando é considerado o número de sinais corretos. Tanto a melhor rede escolhida pelos critérios de informação quanto a escolhida pelo menor MSE têm uma alta capacidade de geral sinais corretos. Algo semelhante ocorre na série do índice Bovespa. A série do MexBolsa também apresenta uma série de “quebras estruturais”, como a desvalorização do peso em 1994, as crise da Ásia, Rússia e Brasil, bem como, pela proximidade geográfica e econômica com os Estados Unidos, o estouro da bolha do mercado acionário americano em 2000.

Tabela 1.3 : Estatísticas fora da amostra (23/05/02 a 24/07/02)

Brasil	RMSE	MAE	MAPE	SIGN	CORR
RNA(1,1,1)	2.8615	2.2858	1.0472	0.5116	0.0761
RNA(7,7,1)	2.8989	2.2440	0.9886	0.5581	0.2080
RNA(6,6,1)	2.9366	2.3159	1.0748	0.4651	-0.0132
Argentina					
RNA(1,1,1)	3.6256	3.1159	1.0594	0.5581	0.1180
RNA(7,7,1)	3.6722	3.1565	1.0978	0.4186	-0.0236
México					
RNA(1,1,1)	1.6808	1.1927	1.2102	0.5238	0.0184
RNA(6,6,1)	1.6523	1.1838	1.1121	0.5476	0.1036

Nota: As redes apresentam características diferentes. O apêndice I lista as arquiteturas estimadas.

1.5. Lucratividade:

Se o investidor pode lucrar, quanto pode, e quanto de risco este pode tomar depende do tipo de estratégia de trading que cada um usa. Especialmente no caso de posições que são avaliadas diariamente, os lucros podem evaporar quando confrontados com custos de transação. Dessa forma, uma estratégia baseada em previsões recursivas pode não ter uma lucratividade superior a uma estratégia passiva.

A avaliação da rentabilidade fora da amostra desta estratégia segue Lo e Mackinlay(1999).

Para determinar a significância econômica da estratégia compara-se o retorno total da estratégia de um investimento passivo, em cada um dos índices, o que é conhecido na literatura com “buy-and-hold”, com o retorno total da estratégia de

alocação ativa determinada pelas redes neurais. Em particular, para cada um dos países calcula-se o seguinte.

$$W_T^{passiva} = \prod_{t=1}^T (1 + R_t^*), \quad (1.21)$$

$$W_T^{ativa} = \prod_{t=1}^T [\theta_t (1 + R_t^*) + (1 - \theta_t)(1 + R_{ft})] \quad (1.22)$$

em que

$$\theta_t = 1 \text{ se } \hat{y} > 0$$

e

$$\theta_t = 0 \text{ se } \hat{y} < 0$$

R_n é o retorno da libor de 1 semana e y é o retorno estimado.

Os gráficos 1.7 e 1.8 bem como as tabelas 1.4 e 1.5 mostram que a estratégia de alocação ativa usando redes neurais tem um desempenho muito melhor do que a estratégia passiva, tanto para o Brasil quanto para a Argentina.

Para determinar a importância dos custos de transação, a tabela 1.4 também reporta os custos de transação que anulariam os ganhos da estratégia passiva definindo-se como o custo percentual das compras ou vendas dos índices de ações que equalizariam o retorno total da estratégia passiva. Formalmente, se a estratégia ativa requer k mudanças nos 43 dias do período de investimento, então o break-even será definido por:

$$W_T^{passivo} = W_T^{ativo} * (1 - s)^k,$$

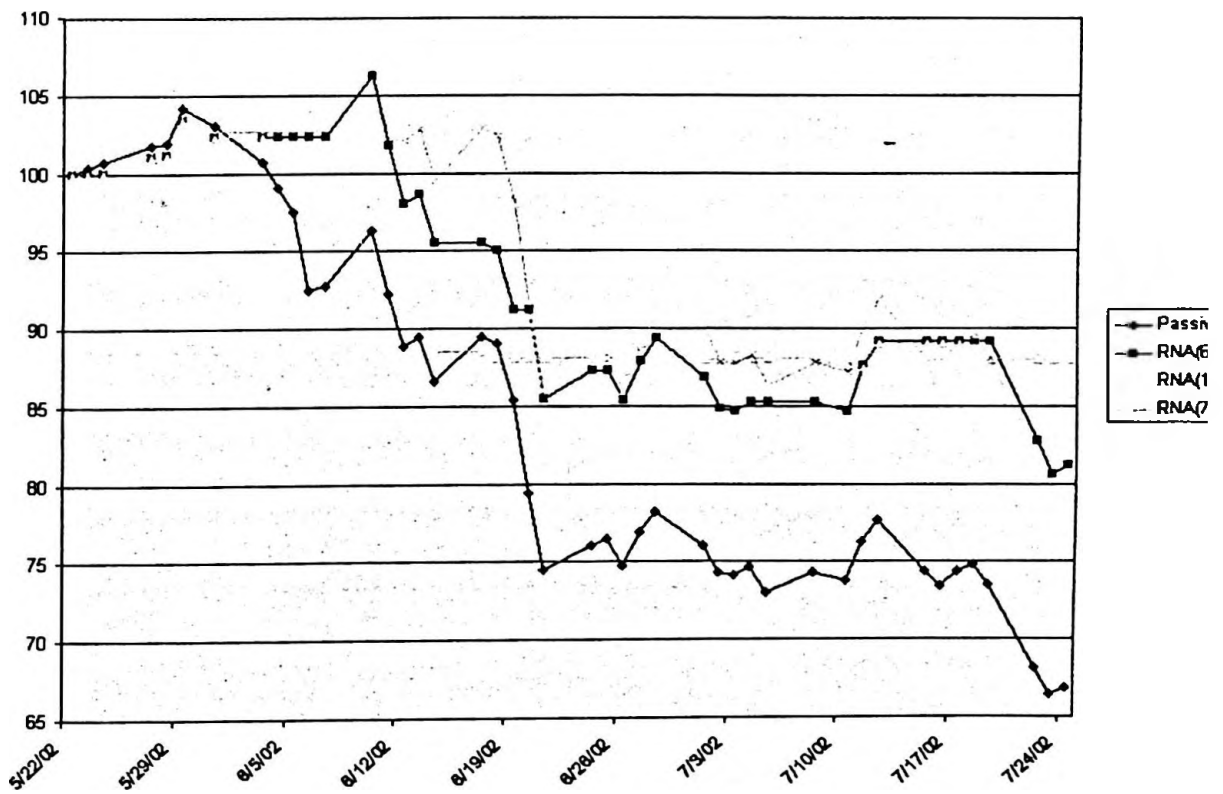
$$s = 1 - \left(\frac{W_T^{passivo}}{W_T^{ativo}} \right)^{1/k} \quad (1.23)$$

A tabela 1.4 mostra o número de mudanças de cada uma das estratégias. Para o caso do Ibovespa onde três modelos conseguem ter uma performance melhor do que o benchmark pode-se notar que apesar da arquitetura que apresenta apenas uma defasagem como input ser mais lucrativa, os custos de transação devem se manter pequenos em função do número de entradas e saídas.

Os resultados considerando-se os custos de transação como nulos são mostrados no gráfico 1.7.

Gráfico 1.7

Brasil- Resultado das Diferentes Estratégias para US\$ 100 Investidos



Uma avaliação do gráfico 1.7 acima mostra que para o caso brasileiro, a rede degenerada baseada na estimação do modelo ARIMA, RNA(6,6,1) e a RNA(7,7,1) apresentam desempenho superior tanto em relação à estratégia passiva quanto à rede RNA(1,1,1), quando se consideram as primeiras previsões. No entanto, com a passar do tempo, ambas perdem poder de previsão e a rede RNA(1,1,1) passa a ter um desempenho melhor. Nota-se que, apesar de quase todos os critérios de performance fora da amostra indicarem a escolha do RNA(7,7,1), a arquitetura RNA(1,1,1) foi capaz de evitar os maiores retornos negativos do período o que contribuiu para sua boa rentabilidade.

Tabela 1.4 : Brasil: Avaliação fora da amostra das previsões 1 passo à frente das arquiteturas de RNAs usando a comparação entre estratégias passivas e ativas de investimento

	Retorno Médio(1)	D.P.(%)(2)	Valor Final(\$)(3)	Mudanças (4)	Break-even(5)
Passiva	-16.00	11.70	66.94	-	-
RNA(1,1,1)	-7.98	5.43	88.09	20.00	1.36
RNA(7,7,1)	-5.95	7.80	79.56	8.00	2.14
RNA(6,6,1)	-7.25	7.45	81.27	13.00	1.48

Notas:

- 1) Média aritmética dos retornos
- 2) Desvio Padrão
- 3) Valor final da operação supondo um investimento inicial de 100 dólares
- 4) Número de trocas entre renda fixa e ações.
- 5) Custo de Transação que elimina a lucratividade da estratégia

A tabela 1.4 mostra a importância do custo de break-even das estratégias. Dessa forma, como é difícil identificar o melhor modelo, pode-se raciocinar que um grande investidor, ao explorar esta estratégia, pode se valer do RNA (1,1,1) para o Brasil uma

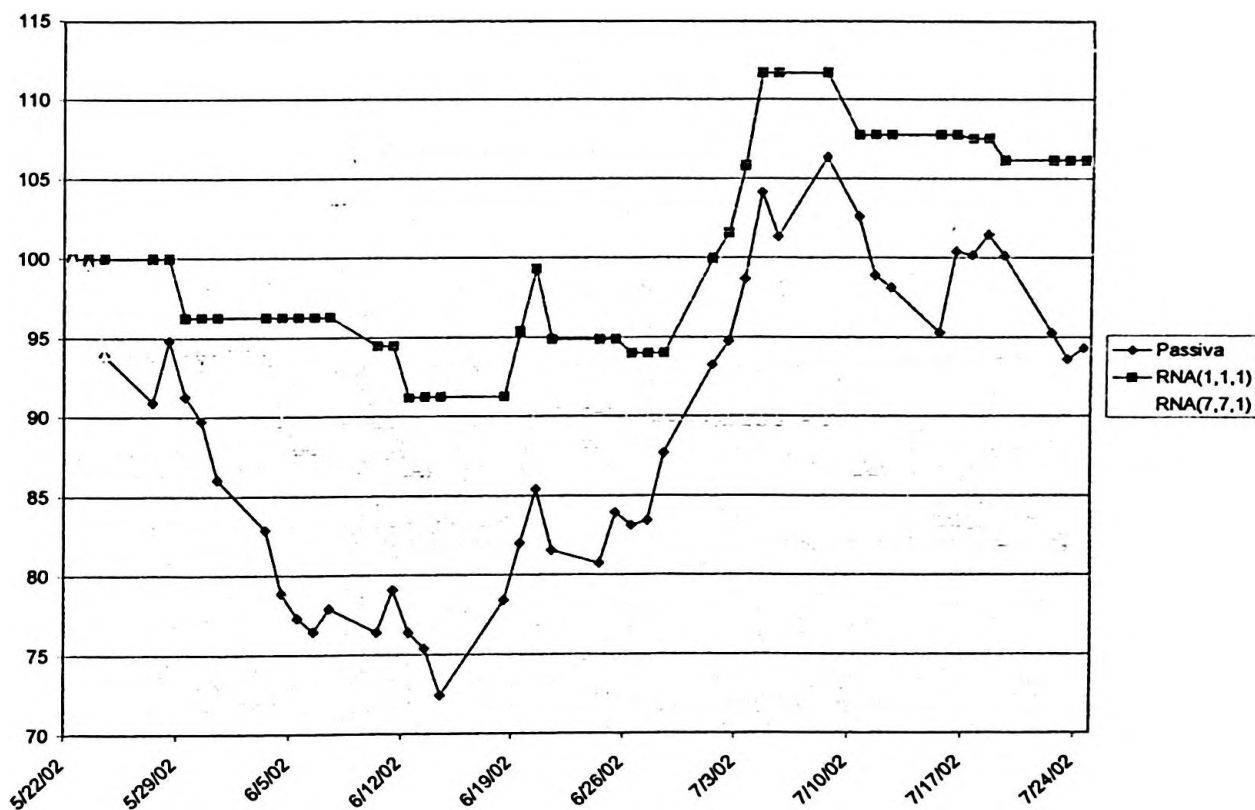
vez que é provável que este incorra em menores custos de transação. Especificamente, para o caso brasileiro os custos de transação gravitam na faixa de 0.3%.

Para evitar o viés da coluna valor final, reporta-se na tabela acima o retorno médio, mostrando o desempenho entre as diferentes arquiteturas.

O gráfico 1.8. mostra o retorno das redes neurais no caso argentino.

Gráfico 1.8

Argentina- Resultado das Diferentes Estratégias para US\$100 Investidos



No caso Argentino, o problema do “overfitting” fica mais claro. Isso porque a inclusão de mais variáveis (7 defasagens e 7 neurônios na camada escondida) reduz o

Erro Quadrático Médio, mas o desempenho fora da amostra poderia levar à conclusão de que uma RNA não pode ser usada para a alocação de recursos.

O índice Merval usado na estimação está denominado em dólares, o que implica conter pelo menos uma quebra estrutural, além das eleições presidenciais, crises da Ásia (1997), Rússia (1998) e Brasil (1999), ou seja, o calote argentino e a consequente desvalorização cambial em outubro de 2001. Como não havia mercado oficial de câmbio depois da revogação da Lei de Convertibilidade, a taxa de câmbio usada foi a praticada informalmente pelos bancos. O câmbio só se tornou livre depois de fevereiro de 2002 quando realmente a série apresenta uma máxima desvalorização do peso.

Tabela 1.5 : Argentina: Avaliação fora da amostra das previsões 1 passo à frente das arquiteturas de RNAs usando a comparação entre estratégias passivas e ativas de investimento

	Retorno Médio(1)	D.P.(%)(2)	Valor Final(\$)(3)	Mudanças (4)	Break-even(5)
Passiva	-10.32	9.618774	94.32167	-	-
RNA(1,1,1)	0.16	6.29371	106.1993	18	0.65
RNA(7,7,1)	-10.99	4.590357	82.79891	-	-

Notas:

- 1) Média aritmética dos retornos
- 2) Desvio Padrão
- 3) Valor final da operação supondo um investimento inicial de 100 dólares
- 4) Número de trocas entre renda fixa e ações.
- 5) Custo de Transação que elimina a lucratividade da estratégia

A tabela 1.5. mostra o desempenho quando são considerados os custos de transação. Apesar do resultado financeiro ser nominalmente positivo, os custos de implementação devem manter-se baixos, caso contrário esta estratégia perderá do buy and hold. O fato da RNA(1,1,1) no caso argentino apresentar um bom resultado, não

elimina as chances de que uma rede ainda melhor exista, já que foi utilizada uma regra para o número de neurônios da camada escondida.

No entanto, a metodologia usada permite que a rede “generalize” de forma satisfatória e reconheça as quebras estruturais diminuindo o risco de “overfitting”.

O gráfico 1.9. e a tabela 1.6 mostram o desempenho das estratégias de alocação entre bolsa e renda fixa no caso mexicano. Mesmo em uma economia considerada mais estável que a brasileira ou a argentina , é possível obter ganhos através de uma regra de alocação de carteiras baseado em RNAs. Na tabela 1.5 é possível notar que os ganhos financeiros da estratégia, mesmo considerando os custos de transação são extremamente altos, indicando que possivelmente, se um estudo semelhante fosse realizado para índices de ações de países mais desenvolvidos, e teoricamente mais eficientes, ainda assim, seria possível que um regra baseada em sinais gerados por uma RNA pudesse obter ganhos acima da estratégia passiva.

Gráfico 1.9

México- Resultado das Diferentes Estratégias para US\$100 Investidos

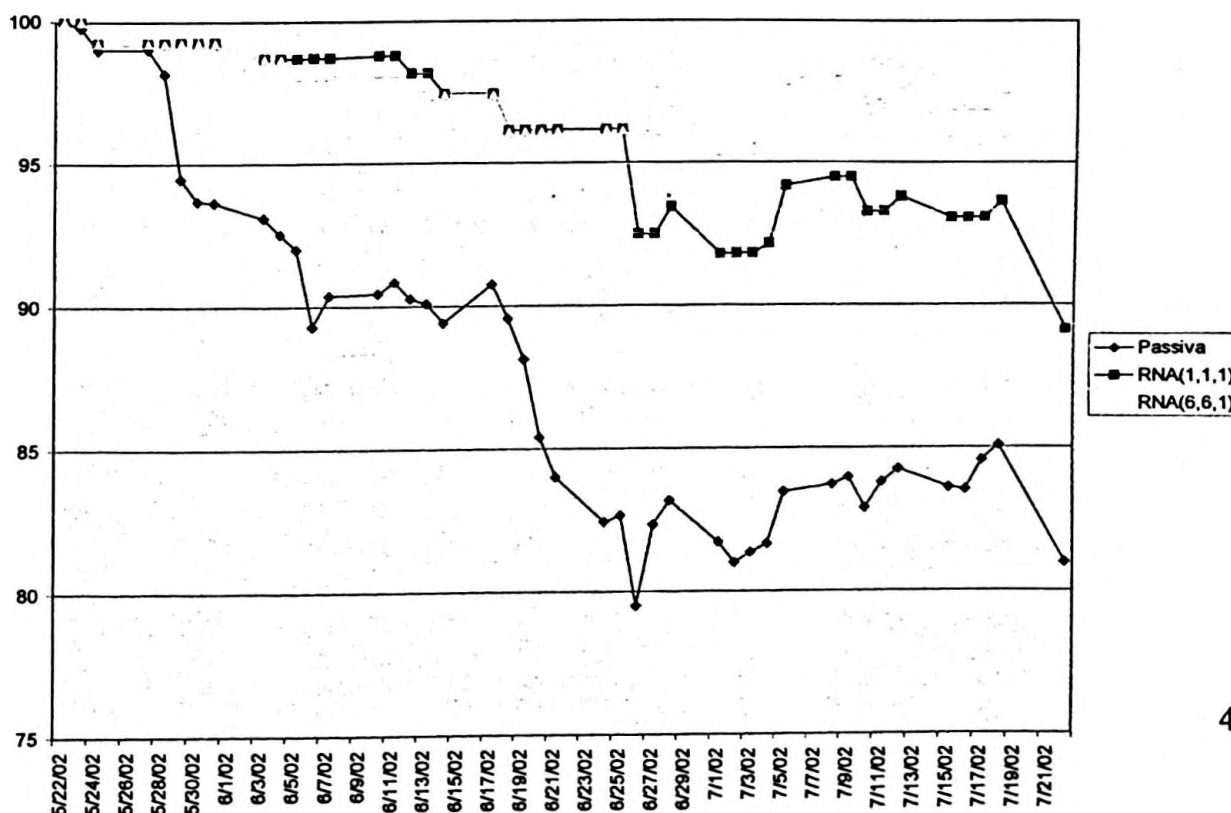


Tabela 1.6 : México: Avaliação fora da amostra das previsões 1 passo à frente das arquiteturas de RNAs usando a comparação entre estratégias passivas e ativas de investimento

	Retorno Médio(1)	D.P.(%)(2)	Valor Final(\$)(3)	Mudanç as(4)	Break- even(5)
Passiva	-12.09	5.86071	87.9169	-	
RNA(1,1,1)	-4.03	2.93939	95.9696	22	0.397
RNA(6,6,1)	-2.54	1.46134	97.4597	18	0.57

Notas:

- 1) Média aritmética dos retornos
- 2) Desvio Padrão
- 3) Valor final da operação supondo um investimento inicial de 100 dólares
- 4) Número de trocas entre renda fixa e ações.
- 5) Custo de Transação que elimina a lucratividade da estratégia

1.6. Conclusão:

Ao testar a poder de previsão de diferentes arquiteturas de redes neurais o ensaio procura adicionar uma contribuição ao estudo da eficiência dos mercado emergentes. Em termos gerais, pode-se notar que mais de uma rede cumpre esta função. Após a estimação de 2400 arquiteturas de RNAs para o Índice Bovespa, 1800 arquiteturas de RNAs para o Índice Merval argentino e 1800 para o índice MexBolsa mexicano, não se pode garantir que as melhores redes que utilizam o algoritmo de retropropagação tenham sido encontradas. Considerando os diferentes critérios de informação, uma rede com mais defasagens pode obter retornos superiores e com menos risco, se o número de neurônios na camada escondida forem reduzidos. Uma outra alternativa seria testar as defasagens

uma a uma, por exemplo, só a quinta defasagem do retorno como variável explicativa. Artigos mais recentes, como McNelis, Camba-Mendez e McAdam(2003) propõem que seja utilizada a média das previsões das diferentes redes, o que, se por um lado pode melhorar a capacidade de previsão desta ferramenta, por outro, levanta a dúvida sobre qual universo de redes utilizar para tomar esta decisão. A conclusão final deste ensaio é de que uma rede neural que utiliza o algoritmo de retropropagação pode ser uma ferramenta útil no processo de alocação de carteiras em diferentes países e em carteiras globais de ações.

Apêndice I: Características das Arquiteturas das Redes Neurais:

Brasil:

RNA(1,1,1)

Camada escondida: função gaussiana complementar

Algoritmo de aprendizagem: Momentum -Rotation

RNA(7,7,1)

Camada escondida: função gaussiana

Algoritmo de aprendizagem: Turbo Pro- Rotation

RNA(6,6,1)

Obs: rede gerada pelo AR(p) degenerado com defasagens 1,5,6,8,9,17

Camada escondida: função seno

Algoritmo de Aprendizagem: Turbo Pro – Rotation

Argentina

RNA(1,1,1)

Camada escondida: função gaussiana

Algoritmo de Aprendizagem: Momentum-Rotation

RNA(7,7,1)

Camada Escondida: função gaussiana complementar

Algoritmo de Aprendizagem: Momentum-Rotation

México:

RNA(1,1,1)

Camada Escondida: função gaussiana complementar

Algoritmo de aprendizagem: Vanila-Rotation

RNA(6,6,1)

Camada Escondida: função tangente

Algoritmo de aprendizagem: Momentum-Rotation

Apêndice 2: Desempenho da RNA(1,1,1) no conjunto de treinamento e no conjunto de teste

Gráfico 1.10

Desempenho da Rede RNA(1,1,1) no conjunto de treinamento

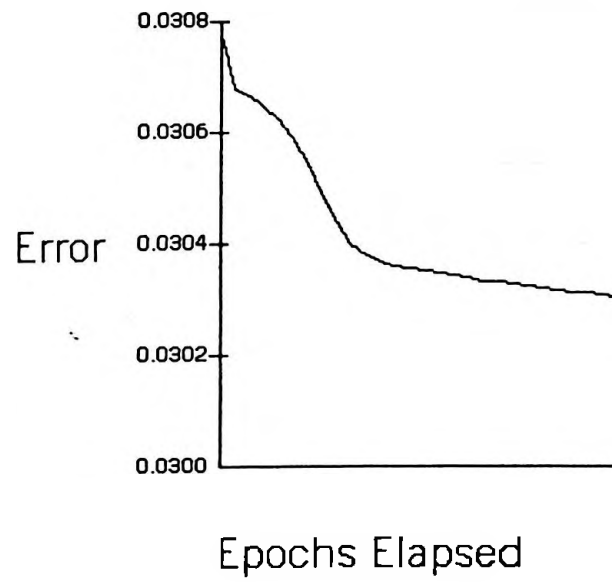
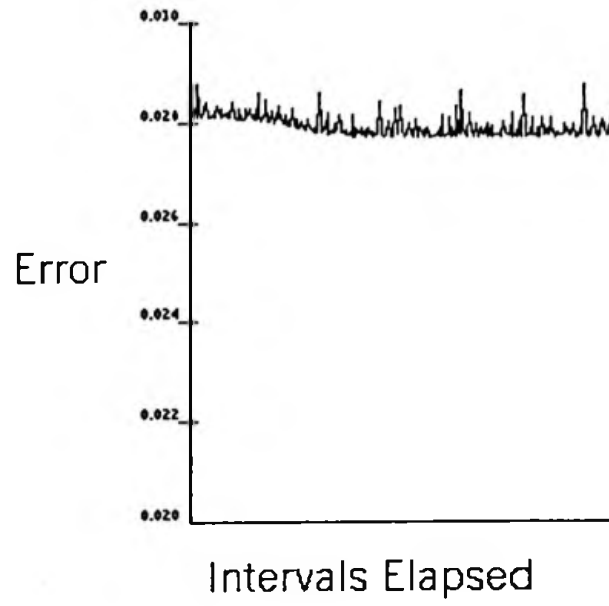


Gráfico 1.11

Desempenho da rede RNA(1,1,1) no conjunto de teste



Ensaio 2

Regras de Análise Técnica e a Previsibilidade dos Retornos com Redes Neurais

2.1. Introdução:

Este artigo é uma extensão do ensaio 1 e utiliza os retornos diários do Índice Bovespa em dólares no período de 3 de janeiro de 1993 até 24 de julho de 2002 para examinar a previsibilidade não linear de algumas regras de análise técnica. Algumas das regras são bem simples e populares, mas algumas regras mais complexas são testadas.

Edwards e Magee (1966)¹ definem assim a análise técnica de séries financeiras:

“Technical analysis is the science of recording, usually in graphic form, the actual history of trading (price changes, volume of transactions, etc.) in a certain stock or in “the averages” and then deducting from that pictured history the probable future trend”

Este artigo está dividido como segue. A próxima seção resenha a literatura que testa análise gráfica e a previsibilidade de retornos. esta seção estende a resenha iniciada no ensaio 1. A terceira parte introduz alguns estudos de análise técnica. A quarta seção apresenta a metodologia e a quinta discute os resultados. A última parte é dedicada às conclusões.

¹ Também citado em Campbell, Lo e Mackinlay (1997 pp.43)

2.2. Padrões e Indicadores Técnicos:

Os padrões e Indicadores Técnicos mais usados são:

Padrão 1: Ombro-Cabeça-Ombro (O-C-O): Lo et all(2000) definem uma formação de O-C-O e O-C-O invertido como uma sequência de cinco extremos locais consecutivos E_1, \dots, E_5 tal que

O-C-O	E_1 um máximo $E_3 > E_1, E_3 > E_5$ E_1 e E_5 dentro de 1.5% de sua média E_2 e E_4 dentro de 1.5% de sua média
O-C-O-I	E_1 é um $E_3 > E_1, E_3 < E_5$ E_1 e E_5 dentro de 1.5% de sua média E_2 e E_4 dentro de 1.5% de sua média

A figura 2.1 mostra o padrão do Ombro-Cabeça-Ombro enquanto a figura 2.2 apresenta o seu espelho. O O-C-O é considerado uma padrão de reversão de tendência².

² Ver Gourieroux e Jasiak(2001 p.146).

Figura 2.1: Formação Ombro-Cabeça-Ombro

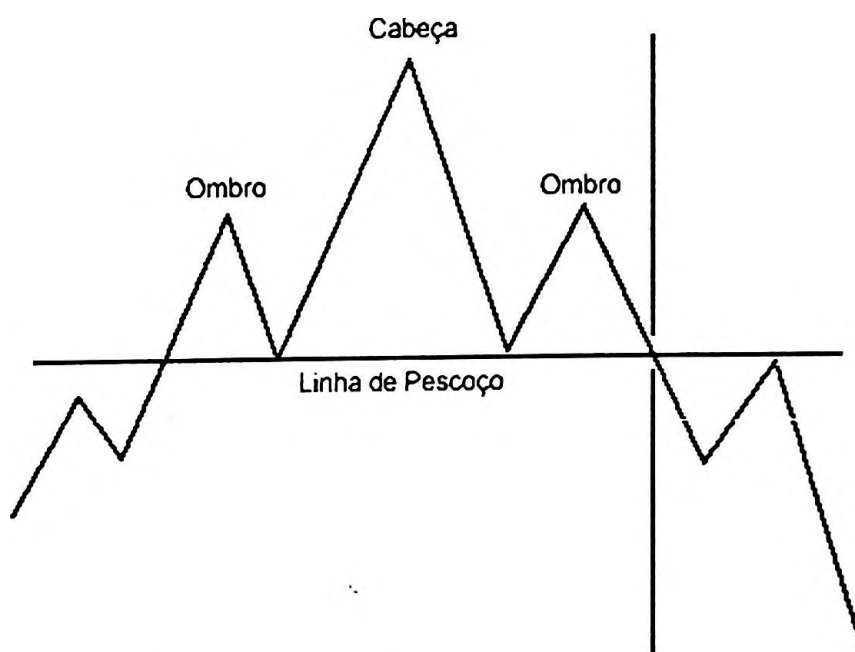
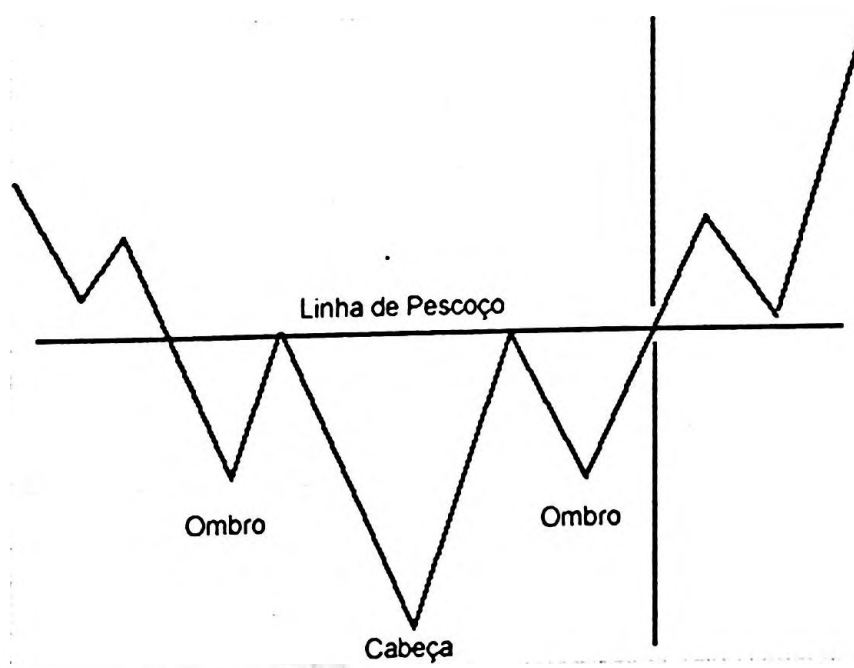


Figura 2.2: Formação Ombro-Cabeça-Ombro Invertido



Padrão 2: Topos Crescentes e Fundos Crescentes são caracterizados por uma seqüência de 5 extremos locais consecutivos E_1, \dots, E_5 , tal que

TOPO E_1 é um máximo

$$E_1 < E_3 < E_5$$

$$E_2 > E_4$$

FUNDO E_1 é um mínimo

$$E_1 > E_3 > E_4$$

$$E_2 < E_4$$

As figuras 2.3 e 2.4 ilustram estes padrões.

Figura 2.3: Topos Crescentes

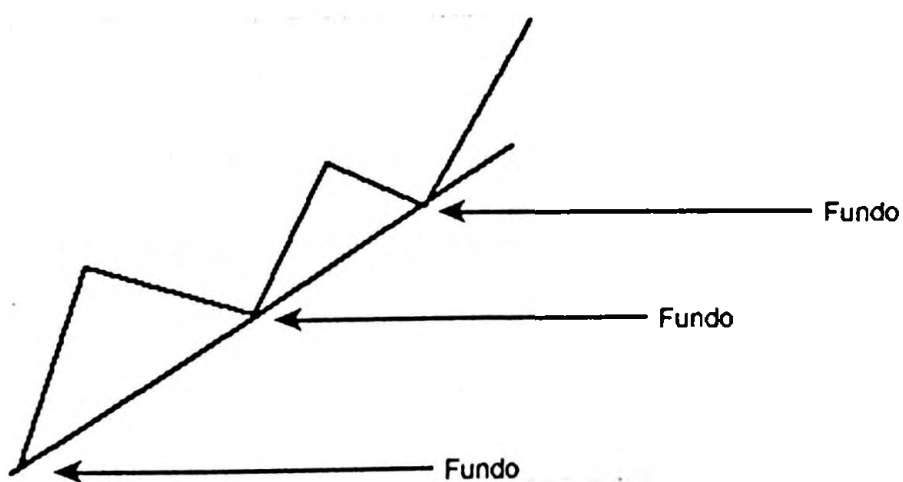
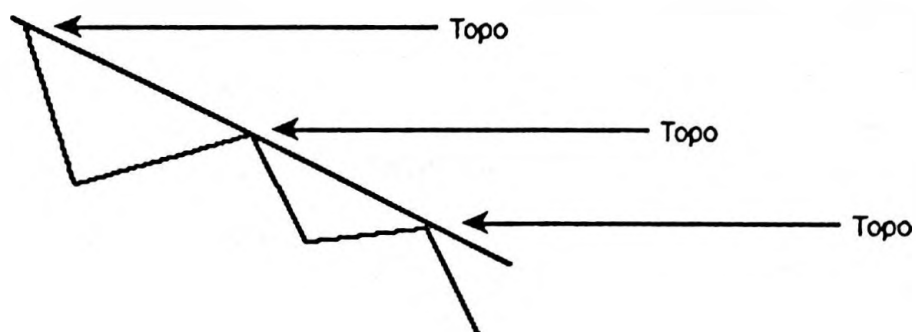


Figura 2.4: Fundos Decrescentes



Padrão 3: Triângulo: Este padrão é caracterizado por uma seqüência de cinco extremos locais consecutivos E_1, \dots, E_5 tal que:

TTOP E_1 é um mínimo

$$E_1 > E_3 > E_5$$

$$E_2 < E_4$$

TBOT E_1 é um máximo

$$E_1 < E_3 < E_5$$

$$E_2 > E_4$$

A formação de triângulo pode ser ilustrada pela figura 2.5.

Figura 2.5: Triângulos



Padrão 4(Retângulo): Retângulos podem ser formações de Topos ou Fundos e segundo Lo et al (2000) são caracterizados por uma seqüência de 5 extremos locais consecutivos $E_1.E_5$ tal que:

RTOP E_1 é um máximo

Topos entre 0.75% de sua média

Fundos entre 0.75% de sua média

Menor Topo > maior fundo

RBOT E_1 é um mínimo

Topos entre 0.75% de sua média

Fundos entre 0.75% de sua média

Menor Topo > maior fundo

Padrão 5: (Topos e Fundos Duplos): Topos Duplos e Fundos Duplos são caracterizados por um extremo local inicial E_1 e extremos locais subseqüente E_a e E_b tal que

$$E_a \equiv \sup \{P_{i_k}^* : t_k^* > t_1^*, k = 2, \dots, n\}$$

$$E_b \equiv \inf \{P_{i_k}^* : t_k^* > t_1^*, k = 2, \dots, n\}$$

e

DTOP E_1 é um máximo

E_1 e E_a entre 1.5% de sua média

$$T^*_a - t^*_1 > 22^3$$

DBOT E_1 é um mínimo

E_1 e E_b entre 1.5% de sua média

$$T^*_a - t^*_1 > 22$$

As figuras 2.6 e 2.7 ilustram a formação de fundos e topos duplos.

Figura 2.6: Topos Duplos

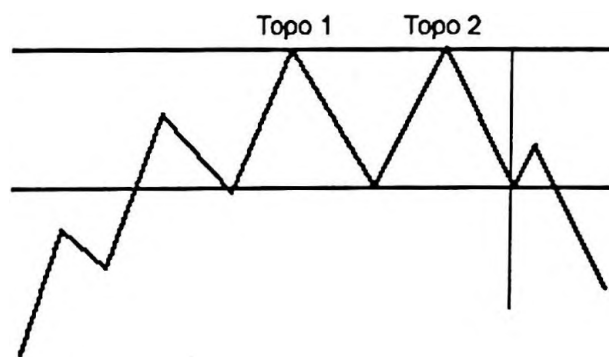
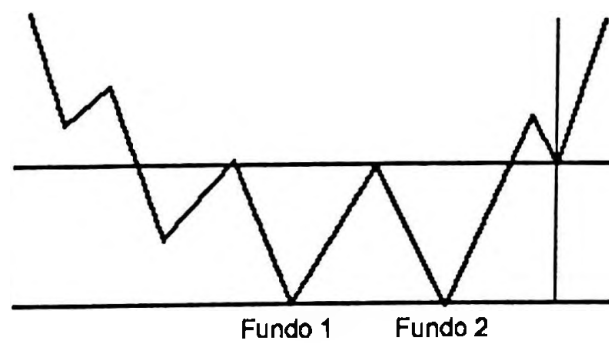


Figura 2.7: Fundos Duplos



³ Lo et al(2000) utilizam a definição de Edwards e Magee(1966) que requer que os dois topos ocorram em pelo menos 22 dias úteis.

Gráfico 2.1

Exemplo dos principais padrões utilizando a série diária do Índice Bovespa denominado em dólares entre janeiro de 1993 e agosto de 2002



Nota: Gráfico gerado utilizando o programa Metastock. Os padrões foram identificados pelo autor.

O gráfico 2.1 mostra alguns exemplos de padrões tendo como base a série diária do índice Bovespa denominado em dólares e que é utilizada nos 3 ensaios. Nota-se que o exame visual não é muito eficiente. O mesmo tipo de padrão pode surgir em gráfico de frequência mais alta como 15 e 60 minutos.

Indicador 1: Médias Móveis:

A regra pode ser o cruzamento do preço com sua média móvel (aritmética ou exponencial) em t períodos passados, ou o cruzamento de uma média móvel de “curto prazo”, em geral até 10 períodos, com uma de “longo prazo”, que pode ir de 11 a 200 períodos.

A média móvel aritmética é a soma do preços do ativo em n períodos dividida por n períodos.

A média móvel exponencial é calculada da seguinte forma. Considere o exemplo formulado em MetaStock (1997 p.459). Para o cálculo de uma média móvel exponencial de 9% segue-se a seguinte fórmula:

$$MA_t = (P_t \times 0.09) + (MA_{t-1} \times 0.91)$$

No entanto, convencionou-se transformar o número de períodos na ponderação percentual. A fórmula para a conversão de dias em porcentagem exponencial é a seguinte:

$$\text{porcentagem exponencial} = \frac{2}{\text{número de períodos} + 1}$$

Logo, para a mesma ponderação, pode-se dizer que está é uma média de 21 períodos. O gráfico 2 mostra um exemplo do tipo de sinal emitido pelo cruzamento do índice Bovespa com sua média móvel.

Gráfico 2.2

Exemplo dos sinais de compra e venda emitidos por uma regra de médias móveis aritméticas



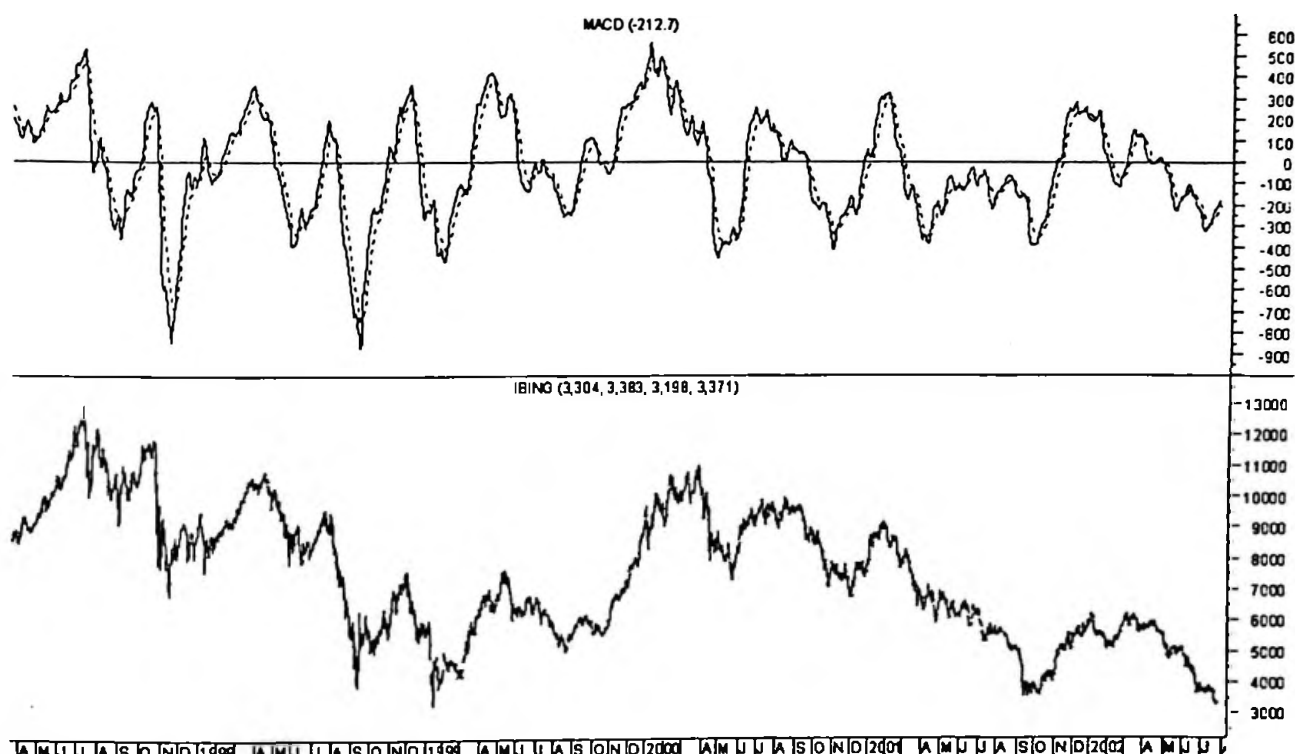
Indicador 2: MACD:

O indicador chamado de Moving Average Convergence/Divergence (MACD) é calculado pela subtração do valor de uma média móvel exponencial de 26 períodos com um média móvel exponencial de 12 períodos. Um sinal de compra ou venda é disparado quando esta medida é maior ou menor do que uma média móvel do MACD. No presente estudo serão testados os cruzamento do MACD com as médias móveis de 10 e 20 períodos. O gráfico 3 mostra um exemplo de como o MACD pode ser visualizado. Em geral, este indicador apresenta o mesmo problema que outras medidas que tentam prever a mudança de tendência do mercado, ou seja, a emissão de sinais de compra e venda errados quando o mercado uma determinada tendência se prolonga por fatores exógenos.

No entanto, como mostra o gráfico, o MACD teria sido muito útil para uma venda a descoberto ou no encerramento de uma posição “comprada” em um momento crítico como foi março de 2000, quando o mercado iniciou um movimento mais agressivo de queda.

Gráfico 2.3

Exemplo dos sinais de compra e venda emitidos pela regra de MACD



Nota: O exemplo foi gerado pelo programa Metastock.

Indicador 4: Estocástico

O cálculo inicial para o estocástico segue a seguinte fórmula:

$$\%K = \left[\frac{C(l) - L(n)}{H(n) - L(n)} \right] \times 100$$

em que

$C(1)$ = preço de fechamento “hoje”

$L(n)$ = menor preço mínimo em n períodos;

$H(n)$ = maior preço máximo em n períodos

Este indicador varia oscila de zero (quando o fechamento de “hoje” é igual ao menor preço em n períodos) a 100 (quando o fechamento de hoje é igual ao preço mais alto em n períodos). O período n pode se referir a semanas, dias ou minutos. Qualquer janela de tempo pode ser usada desde que exista uma máxima e uma mínima facilmente detectáveis.

O estocástico é considerado um indicador muito volátil uma vez que é derivado de preços de fechamento. Dessa forma é comum a aplicação de uma média móvel para filtrar esta volatilidade. A média móvel do oscilador é chamada de %D. No entanto, mesmo este filtro é considerado volátil, de forma que em muitos softwares de análise técnica é comum a utilização do estocástico “lento” que segue então o seguinte formato

%K rápido = estocástico;

%D rápido = média móvel de K rápido;

%K lento = %D rápido

%D lento = média móvel de K lento.

Existem 3 métodos populares para a interpretação do estocástico:

- 1) Comprar quando o oscilador cair abaixo de um nível específico (ex 20) e depois voltar a subir, e vender quando o oscilador subir até um determinado nível (ex 80). O problema desta estratégia está quando o

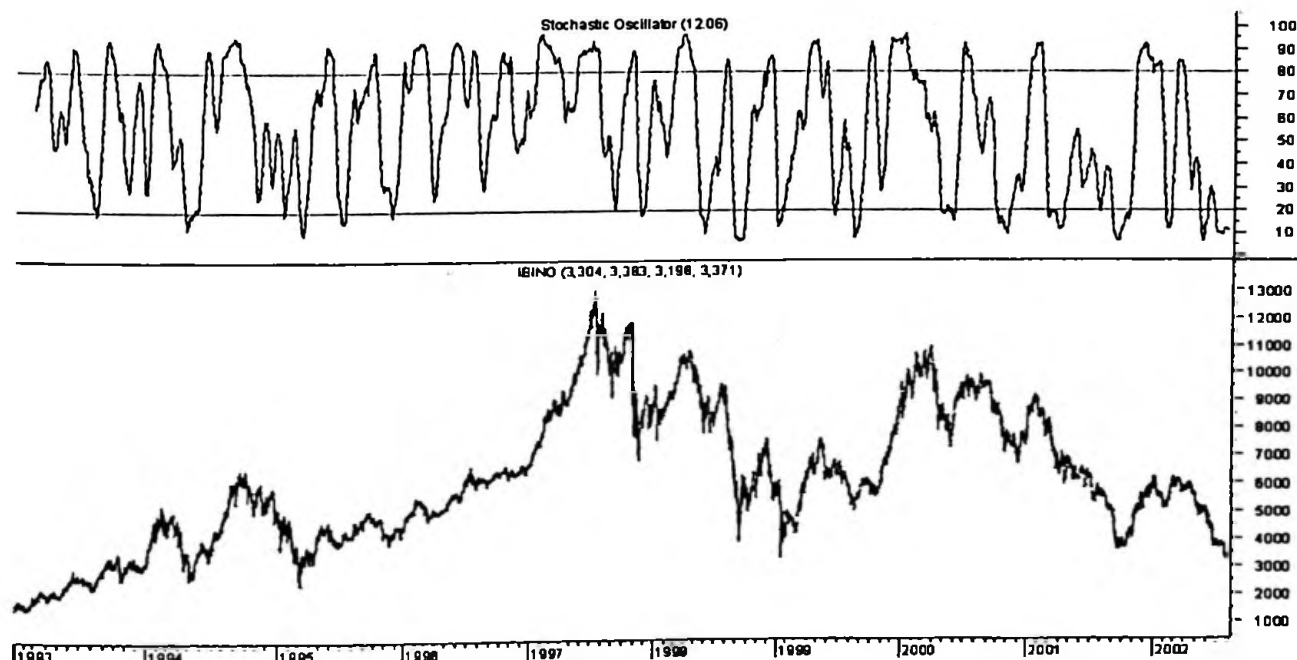
mercado apresenta tendência durante muito tempo, o que conduz a geração de sinais errados.

- 2) Comprar quando %K cruza para cima de %D e vender quando %K cruza abaixo.
- 3) Divergências entre o este indicador e a série, por exemplo, quando os preços estão fazendo novas máximas enquanto o estocástico não acompanha.

O gráfico 2.4 apresenta um exemplo dos sinais de compra e venda gerados pelo estocástico de 21 dias. Em geral, quando a série apresenta uma tendência de prazo longo, como por exemplo a queda iniciada em março de 2002, o estocástico gerou uma sequência de sinais errados.

Gráfico 2.4

Exemplo da geração de sinais de compra e venda do estocástico



Indicador 5: ADX

O ADX (Average Directional Index) é um indicador de tendência do mercado.

Primeiro determinam-se as seguintes variáveis:

+DI é igual ao preço máximo em t menos o preço máximo de t-1;

-DI é igual ao preço mínimo em t menos o preço mínimo em t-1

Em sua versão mais simples tem-se um sinal de compra quando $+DI > -DI$. A figura 2.8 mostra como são geradas as variáveis +DI e -DI.

O gráfico 2.5 mostra um exemplo deste estudo. Nota-se pela última observação, o investidor deveria estar vendido à descoberto ou investir 100% em renda fixa.. Neste caso o estudo emite um sinal igual a -1.

Figura 2.8: Determinação visual do Average Directional Index (ADX)

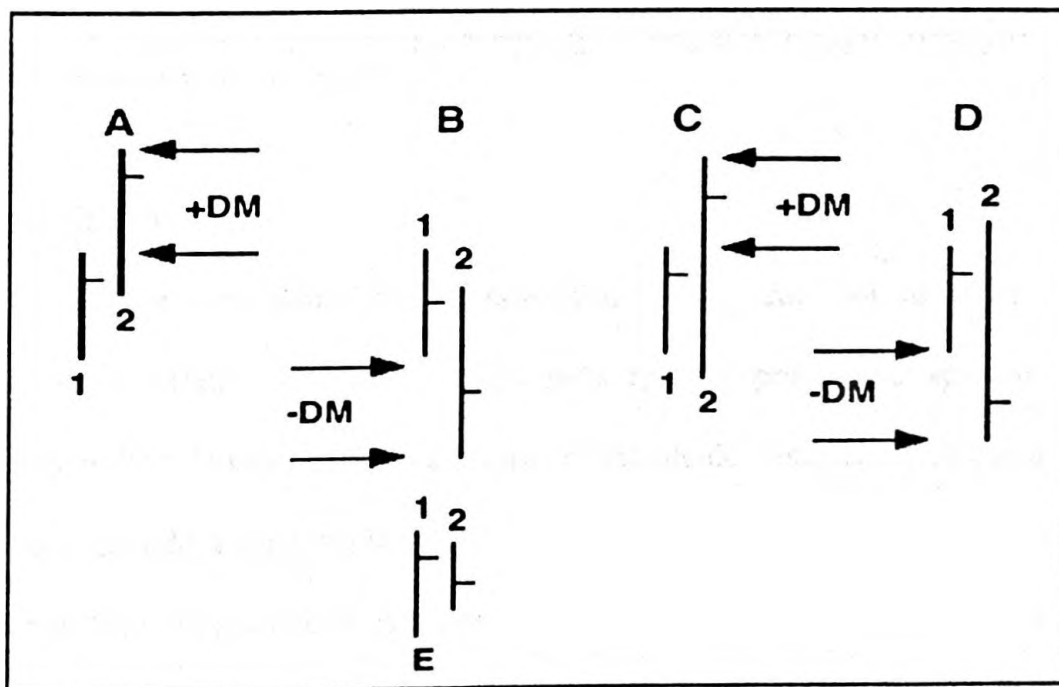
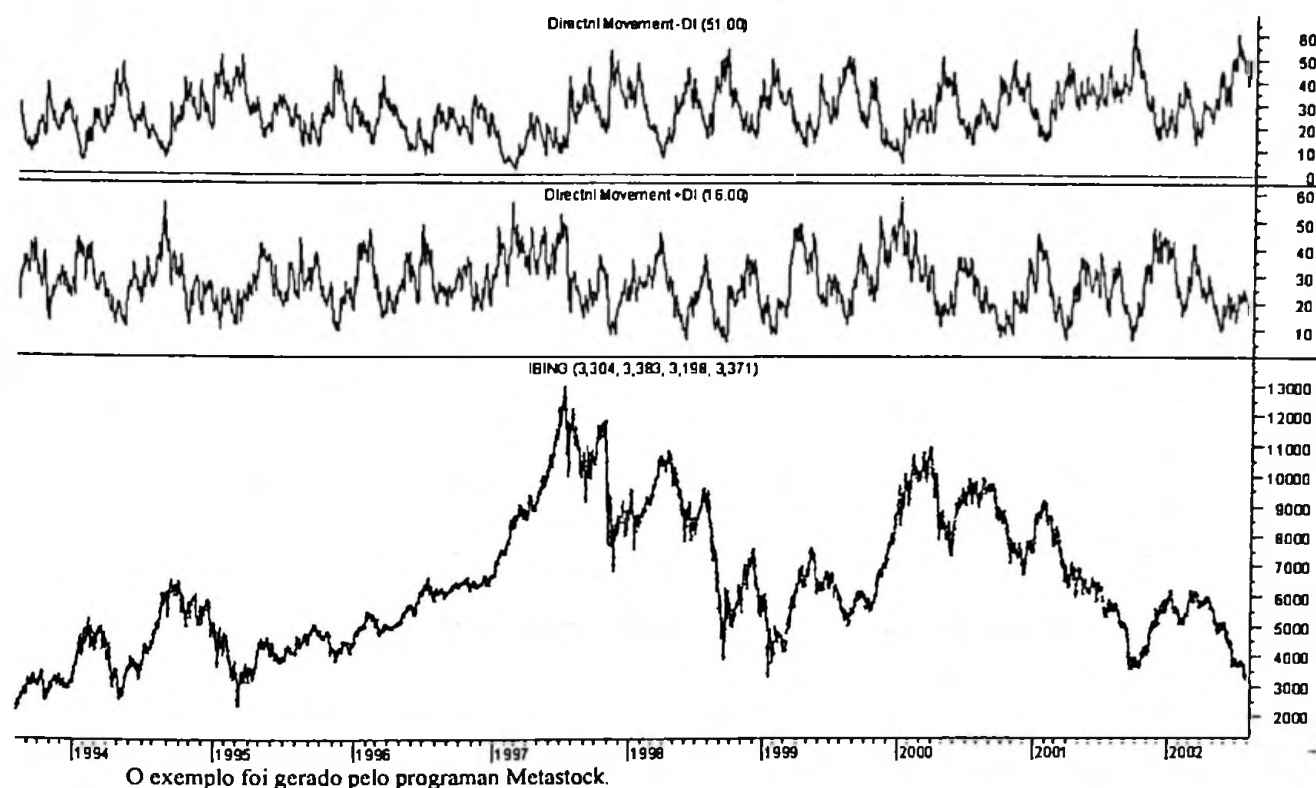


Gráfico 2.5

Exemplo dos sinais de compra e venda gerados pelo ADX no Ibovespa em dólares



Indicador 6: Trix

É uma técnica de alisamento exponencial triplo que foi descrita por Hudson(1983). Substituindo o log do preço pelo próprio preço, ele aplicou um alisamento exponencial três vezes usando a mesma constante de alisamento. Um sinal de compra foi gerado quando a série triplamente alisada subiu por dois dias consecutivos e um sinal de venda foi gerado quando caiu por dois dias seguidos. Kaufman (1998, p.97) mostra o efeito de um alisamento triplo sobre uma série. Um alisamento triplo em uma série de média móvel de três dias conduz ao seguinte resultado:

$$P = (P_1, P_2, \dots, P_n)$$

A média móvel simples no n-ésimo dia é

$$M_n = \frac{(P_{n-2} + P_{n-1} + P_n)}{3}$$

Alisando a nova série M temos

$$\begin{aligned} M'_n &= \frac{(M_{n-2} + M_{n-1} + M_n)}{3} = \\ &= \frac{(P_{n-4} + 2P_{n-3} + 3P_{n-2} + 2P_{n-1} + P_n)}{9} \end{aligned}$$

e o terceiro alisamento conduz

$$\begin{aligned} M''_n &= \frac{(M'_{n-2} + M'_{n-1} + M'_n)}{3} = \\ &= \frac{(P_{n-6} + 3P_{n-5} + 6P_{n-4} + 7P_{n-3} + 6P_{n-2} + 3P_{n-1} + P_n)}{27} \end{aligned}$$

O efeito das ponderações coloca mais peso nos valores do meio, reduzindo-se a significância dos extremos.

2.3 .Resenha da Literatura

Como parcialmente resenhado no primeiro ensaio, existem diferentes formas de demonstrar a não eficiência de mercados. Uma delas é através da utilização de alguma regra que permita “ganhos extraordinários”. Por exemplo, Lo e MacKinlay (1988,1999) mostram que os preços passados podem ser usados para prever a trajetória dos preços futuros o que é o cerne da análise técnica ou grafismo. Diferentes artigos tais como Pruitt e White (1988), Neftci (1991), Brock, Lakonishok e LeBaron (1992), Neely, Weber e

Ditmar (1997), Neely e Weller (1998), Chang e Osler (1999), Osler (2001) e Allen e Karjalainen (1998), têm dado suporte à análise técnica como forma de previsão.

Campbel, Lo e MacKinlay (1997, pp.43-44) provêm um exemplo da diferença de jargão entre acadêmicos e grafistas:

“Consider, for example, the following, which might be found in any recent academic finance journal:

The magnitudes and decay pattern of the first twelve autocorrelations and the statistical significance of the Box-Pierce Q-statistic suggest the presence of a high-frequency predictable component in stock returns.

Contrast this with the statement:

The presence of clearly identified support and resistance levels, coupled with a one-third retracement parameter when prices lie between them, suggests the presence of strong buying and selling opportunities in the near-term.”

Os analistas técnicos baseiam-se em padrões e regularidades dos preços e retornos passados para poderem fazer previsões obre o futuro. A abordagem heurística baseada apenas no exame visual pode não ser muito eficiente, no entanto, estes padrões podem servir como variáveis explicativas a serem utilizadas em modelos não lineares⁴.

Brock et al (1992) procuram testar dois tipos conhecidos de regras: cruzamento de médias móveis e rompimento de suportes e resistência usando o Índice Dow Jones de 1897 a 1986. Os autores concluem que sinais de compra e venda podem gerar retornos acima dos “normais”.

⁴ Gourioux e Jasiak(2001p.146)

Sullivan, Timmermann e White (1998) utilizando-se de uma metodologia de *bootstrap* desenvolvida em White (1997)⁵ avaliam as regras de análise gráfica de Brock et al (1992) bem como quantificam o viés de “data-snooping”. Campbell, Lo, Mackinlay (1999 p.523) definem este viés da seguinte forma.

“Data-snooping biases arise when we ignore the fact that many specification searches have been conducted to obtain the final specification of a model we are fitting to the data. Even if a model is in fact incorrect, by searching long enough over various data sets and/or parameter values, we are likely to find some combination that will fit the data. However, this fit is spurious and is merely a symptom of our extensive search procedure.”

Uma forma alternativa é ver o “data-snooping” como um viés de seleção em que se procura uma regra que melhor se ajusta aos dados e se gera resultados espúrios.

Sullivan, Timmermann e White (1998a) expandem o universo de regras de Brock et al (1992) e aplicam em 100 anos de dados do Dow Jones. Concluem que a maior parte dos resultados de Brock et al (1992) é robusta e não são uma ilusão estatística. No entanto, em Sullivan, Timmermann e White (1998b), que testam os efeitos calendário a conclusão é diferente.

Lo et al (2000) examinaram a presença de vários padrões de análise técnica que prevaleceram nos preços das ações americanas entre 1962-1996 e descobriram que estes padrões não são usualmente recorrentes. O estudo não prova que estes padrões sejam

⁵ Chamada de White Reality Check.

suficientemente previsíveis de forma a obter lucros que justifiquem o risco, mas os autores concluem que isso é provável.

Para as séries de taxas de câmbio, existe um considerável número de artigos que evidencia a boa performance de regras de análise técnica. (Levich e Thomas (1993), LeBaron (1996), Neely, Weller e Dittmar (1997), e Neely (1998).)

Neely & Weller(2000) usam programação genética para identificar regras de trading que utilizam informações passadas das séries de taxas de cambio e da atividade de intervenção dos bancos centrais. Allen e Karjalainen (1999) aplicam metodologia semelhante ao mercado acionário. Baestaens, Van DenBerg & Vandrey(1996) utilizam os cruzamentos de médias móveis de Brock et al.(1992) e comparam com o desempenho de uma rede neural simples para uma ação⁶. Estes autores sugerem que uma rede recorrente pode alcançar melhores resultados o que poderá ser evidenciado no ensaio 3 dedicado a este tipo de redes. Martin e Tan(1992) aplicam redes neurais na montagem de “trading systems” para a taxa de câmbio do dólar australiano. Os autores conseguem um desempenho superior a um modelo AR(1) com uma rede neural de 3 camadas.

Osler (1998) identifica que aqueles que especulam utilizando-se de padrões do tipo ombro-cabeça-ombro, trazem ruídos ao mercado. Dessa forma, o volume de trading é excepcionalmente alto quando existem “noise traders”, mas esta estratégia não é lucrativa e a utilização desta pode ser função de problemas relacionados à racionalidade limitada.

Gourieroux & Jasiak (2001, p.125) argumentam que modelos não lineares são uma evolução em relação à análise técnica em 3 aspectos:

⁶ Este modelo produziu 54% de sinais corretos.

1. Eles contêm um erro de previsão que pode depender da história dos retornos. Assim a previsão é suplementada com um intervalo de previsão.
2. A abordagem é mais rigorosa. A performance dos estimadores pode ser avaliada em intervalos de confiança e a estabilidade destes estimadores também pode ser avaliada.
3. Modelos autoregressivos não lineares podem ser ampliados de forma a incorporar outras variáveis, o que não acontece com a análise técnica de séries individuais.

Para o caso brasileiro, Ratner & Leal(1999) não encontram nenhuma forte evidência para a lucratividade de uma regra de médias móveis contra uma estratégia passiva. Os autores estudaram o retorno ajustado pela inflação do Ibovespa, bem como de outros índices de nove países emergentes, para o período de janeiro de 1982 a abril de 1995. Leal e Varanda (2000) avaliam o desempenho de estratégias grafistas, como a interseção de médias móveis utilizando-se de técnicas de simulação de “bootstrap”. Os autores concluem que o desempenho das regras é decepcionante. Ambos os trabalhos utilizaram as médias móveis sugeridas por Brock et al (1992).⁷

⁷ Tem sido comum entre os autores a utilização deste trabalho, como forma de evitar o viés de seleção na escolha das melhores regras.

2.4 Metodologia dos Testes:

A metodologia para implementação deste procedimento segue Gately (1996), Gençay (1998), Gençay e Stengos (1998) e Yao & Tan (2002).

Gately (1996), com uma versão mais antiga do NeuroShell, tenta estimar um modelo de previsão para o S&P500 utilizando estudos de análise técnica. O autor não usa regras do tipo “se-então” mas o próprio estudo. Por exemplo, o valor da média móvel de 10 períodos é utilizado para a previsão da série. Gately (1996) utiliza o step wise como metodologia. A escolha das variáveis mais importantes recai sobre uma função fornecida pelo software que permite ponderar os estudos mais importantes na previsão. No experimento deste ensaio optou-se pela utilização de regra “se-então” para todos os estudos escolhidos. Alternativamente, não foram colocados todos os estudos em uma mesma rede e sim estimadas redes diferentes para os diferentes estudos de forma a captar a contribuição marginal de cada um dos estudos em relação à rede benchmark estimada no ensaio 1.

Yao e Tan (2002) apresentam um modelo de previsão para o nível da taxa de câmbio utilizando como inputs as médias móveis de 5, 10, 20, 60 e 120 dias. Como medida de performance os autores utilizam o erro quadrático médio normalizado que é igual a $1-R^2$, o número de sinais certos previstos e o lucro real em termos financeiro se a estratégia tivesse sido implementada considerando-se os custos de transação. Os autores conseguiram um bom retorno com a estratégia. Primeiro, desenvolvem uma rede neural apenas com defasagens da taxa de câmbio. Os resultados em termos de sinais certos

foram um pouco acima de 50%, o que foi sensivelmente melhorado na presença de inputs de análise técnica.

Frances e Van Dijk (2000) não conseguem melhorar a geração de sinais corretos para uma rede neural que utilizava retornos semanais do yen e análise técnica. Em geral, os resultados gravitavam em torno de 55% de sinais corretos.

Gençay (1998) aplica sobre dados diários do índice Dow Jones de 1897 a 1988 regras de análise técnica em uma rede neural e compara o desempenho desta com modelos de mínimos quadrados e GARCH-M(1,1). A performance fora da amostra é medida através do MSPE (Mean Squared Prediction Error) e o horizonte de previsão é de 10 dias. O desempenho relativo da rede neural é determinado pela razão da MSPE entre o modelo linear, o GARCH-M e a rede neural com análise técnica.

No presente ensaio utiliza-se a RNA(1,1,1) estimada no capítulo 1 como referência, e em seguida testa-se se há uma melhora da capacidade preditiva das arquiteturas em relação a esta rede, ao se inserir uma nova variável. A nova variável é uma regra de análise técnica. Dessa forma, são testadas 15 diferentes regras de análise gráfica. Os sinais das regras de análise técnica assumem valor 1 caso haja indicação de compra e -1 caso venda. A escolha recaiu sobre a RNA(1,1,1) pois esta reúne duas características desejáveis: é a melhor rede quando os critérios de informação são utilizados e obteve um bom desempenho fora da amostra.

Gençay e Stengos (1998) mostram que uma rede neural com inputs de análise técnica (cruzamentos de médias móveis e volume) melhora a emissão de sinais corretos em relação a outros modelos. Uma rede que utiliza apenas o cruzamento de médias móveis gera em média 58% de sinais corretos enquanto que a inclusão do volume

transacionado eleva este resultado para 63%. Estes autores testam uma versão contínua da regra de médias móveis sob o argumento de que os resultados do artigo não eram sensíveis a escolha dos sinais⁸.

Esta versão pode ser escrita na forma:

$$s_t^{n1,n2} = m_t^{n1} - m_t^{n2}$$

em que n1 e n2 são as médias curtas e longas.

A rede a ser estimada no presente estudo tem a seguinte forma:

$$y_t = \phi + \sum_{j=1}^d \beta_j G(\alpha_{1j} + \alpha_{2j} y_{t-1} + \alpha_{3j} s_{t-1}^{n1n2})$$

Neste caso s assume os valores 1 para compra e -1 para venda para as regras abaixo:

Regra 1: Cruzamento do Preço de Fechamento com sua Média Móvel Aritmética

Sinal de Compra: Preço de fechamento > Média Móvel n períodos

Sinal de Venda: Preço de fechamento < Média Móvel n períodos

n= 10, 20, 30 períodos

⁸ Os autores utilizam a função logística na camada escondida por ser a "mais usada". No testes realizados com outros tipos de denominação para as regras como 1 caso compra e 0 caso venda as redes mostravam sinais diferentes. Isso é possível por causa da função escolhida tanto na primeira camada quanto na segunda. No experimento deste ensaio a camada escondida usa uma variação da função gaussiana.

Regra 2: Cruzamento da Média Móvel Aritmética de Curto Prazo com Média Móvel de Longo Prazo:

Sinal de Compra: Média Móvel de Curto Prazo n_1 períodos $>$ Média Móvel de Longo Prazo n_2 períodos

Sinal de Venda: Média Móvel de Curto Prazo n_1 períodos $<$ Média Móvel de Longo Prazo n_2 períodos

$n_1 = 3;$

$n_2 = 20, 30, 40$

Regra 3: Estocástico:

Sinal de Compra: Se Estocástico (%K,%D) $>$ 20

Sinal de Venda: Se Estocástico(%K,%D) $<$ 80

%K = 21

%D = 14

Regra 4: ADX

Sinal de Compra: $+DI(n) > -DI(n)$

Sinal de Venda: $+DI(n) < -DI(n)$

$n = 14$

Regra 5 :MACD

Sinal de Compra: $MACD() > \text{Média Móvel Exponencial do MACD em } n \text{ períodos}$

Sinal de Venda: $MACD() < \text{Média Móvel Exponencial do MACD em } n \text{ períodos}$

$n=10, 20$ períodos

Regra 6: TRIX

Sinal de Compra: Quando $Trix(n) > \text{Média Móvel Exponencial de } 3 \text{ períodos do Trix}$

Sinal de Venda: Quando $Trix(n) < \text{Média Móvel Exponencial de } 3 \text{ períodos do Trix}$

$n=30$ e 40 períodos

2.5. Resultados

Tabela 2.1 :Resultados dentro da Amostra considerando os sinais de compra e venda gerados pelas variáveis de Análise Técnica

Estudo	R2	r2	MSE	AIC	BIC	HQ
NN(1,1,1)	0.0144	0.0149	9.194	2.224999	2.237421	2.229527
macd(10)	0.0144	0.0156	9.134	2.211861	2.239189	2.221821
macd(20)	0.0148	0.0161	9.094	2.203137	2.230464	2.213097
ma(10)	0.0146	0.0153	9.18	2.227384	2.254712	2.237344
ma(20)	0.0133	0.0143	9.187	2.221217	2.248545	2.231177
ma(30)	0.014	0.0157	9.095	2.1957	2.223028	2.205661
ma(3-20)	0.0135	0.0143	9.17	2.222407	2.249735	2.232367
ma(3-30)	0.0144	0.0162	9.152	2.216057	2.243385	2.226018
ma(3-40)	0.015	0.0161	9.111	2.207205	2.234533	2.217166
ma*(10)	0.0144	0.0151	9.181	2.227089	2.254417	2.237049
ma*(20)	0.0126	0.0142	9.19	2.221966	2.249294	2.231927
ma*(30)	0.0154	0.0167	9.082	2.19428	2.221608	2.20424
stoch(21)	0.0174	0.0181	9.07	2.193909	2.221237	2.203869
Trix(30)	0.0189	0.0196	9.063	2.179238	2.206566	2.189198
Trix(40)	0.0184	0.0194	9.053	2.164608	2.191936	2.174569
ADX	0.0144	0.0152	9.165	2.221419	2.248747	2.23138

Os resultados descritos na tabela 2.1 parecem ser promissores para um teste fora da amostra. Utilizando-se os critérios de informação nota-se que, por exemplo o MACD pode ter um desempenho melhor do que a rede mais simples, o mesmo ocorrendo com outros indicadores, como o estocástico de 21 períodos e o TRIX. Dessa forma, a adição de sinais de compra e venda gerados por alguns estudos de análise técnica poderia melhorar a previsibilidade dos retornos da série estudada.

Tabela 2.2: Resultados Fora da Amostra considerando os sinais de compra e venda gerados pelas variáveis de Análise Técnica

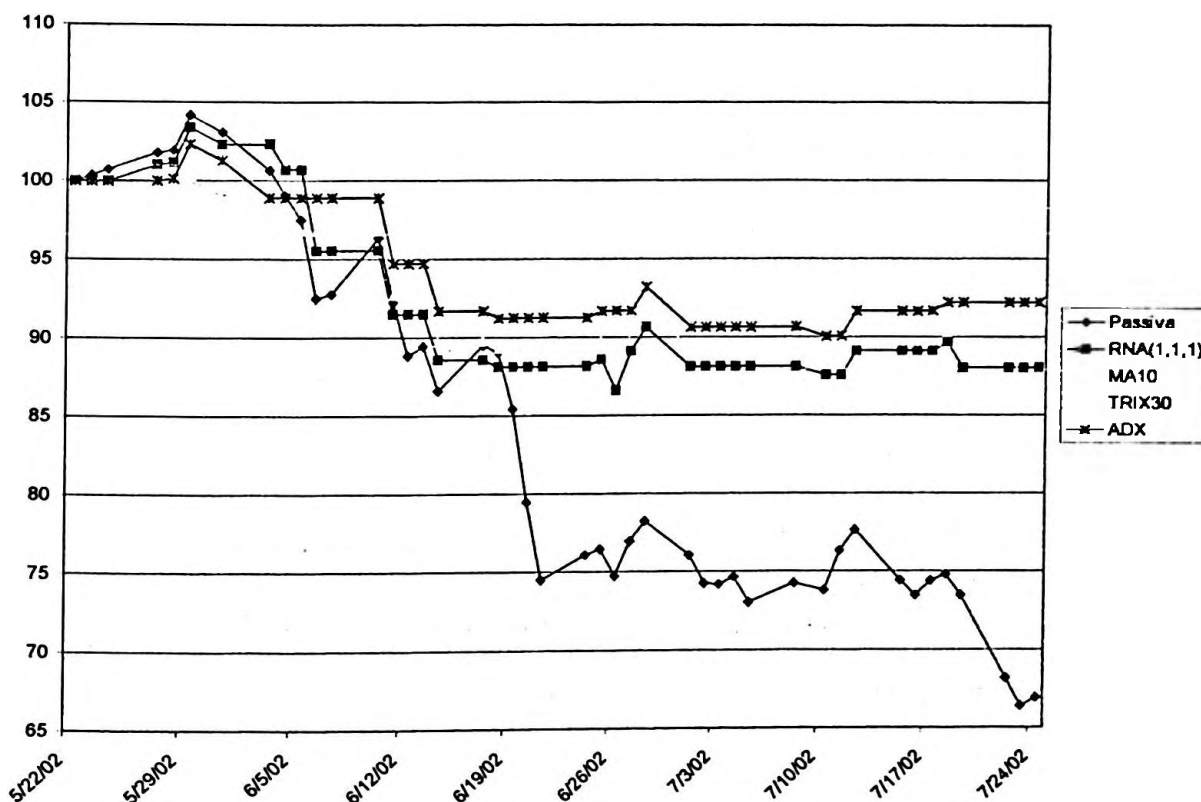
Estudo	RMSE	MAE	MAPE	SIGN	CORR
RNA(1,1,1)	2.861496	2.28576	1.047217	0.511628	0.076102
macd(10)	2.87592	2.302091	1.074847	0.418605	-0.00234
macd(20)	2.87918	2.31317	1.12165	0.418605	-0.03785
ma(10)	2.875922	2.314867	1.094557	0.44186	-0.01829
ma(20)	2.86588	2.30322	1.07917	0.465116	0.051161
ma(30)	2.865561	2.303882	1.081306	0.511628	0.019157
ma(3-20)	2.875653	2.307664	1.081504	0.395349	0.022481
ma(3-30)	2.850865	2.287801	1.092924	0.488372	0.064809
ma(3-40)	2.864315	2.300651	1.094851	0.488372	-0.00161
stoch(21)	2.847872	2.290682	1.091171	0.465116	0.146658
trix(30)	2.845414	2.298144	1.113221	0.465116	0.044372
trix(40)	2.844341	2.282952	1.076227	0.488372	0.077038
ADX	2.854066	2.291716	1.079968	0.511628	0.072018

Fica claro, pela tabela 2.2, que as estatísticas fora da amostra não apresentaram uma melhoria significativa. No entanto, o resultado financeiro apresentou um desempenho ligeiramente mais favorável no caso de alguns estudos. Utilizando a mesma estratégia do primeiro ensaio, ou seja, comprar o índice Bovespa em dólares quando a

RNA prever um sinal positivo para $t+1$ e manter-se em renda fixa (libor) caso contrário, pode-se chegar ao desempenho abaixo ilustrado pelo gráfico 2.6.

Gráfico 2. 6

Resultado de Alguns Estudos de Análise Técnica em relação à Estratégia Passiva



A tabela 2.3 apresenta os resultados completos para todas as estratégias, incluindo os custos de transação. De fato, os estudos de análise técnica não forneceram uma melhora acentuada em relação à rede estimada com apenas uma defasagem. Uma explicação plausível é que, uma vez escolhida a rede através dos critérios de informação, esta foi capaz de captar a maior parte da não linearidade da série de forma que o modelo de previsão já estaria especificado, não havendo necessidade de variáveis adicionais. A

outra possibilidade é de que os inputs não são bons o suficiente para significar uma melhora na qualidade da previsão.

No entanto, em alguns estudos, apesar do desempenho das estatísticas fora da amostra não ter mostrado uma melhoria significativa, houve uma melhora na relação risco/retorno, o que pode indicar a adequação a adição de uma variável baseada nos sinais de análise técnica.

Os custos não representam problema para estas estratégias.

Tabela 2.3: Avaliação das Previsões Um Passo à Frente das Arquiteturas de RNAs Comparando a Estratégia Passiva com as Diferentes Estratégias Ativas

	Retorno Médio(1)	D.P.(%)(2)	Valor Final\$(3)	Mudanças (4)	Break-even(5)
Passiva	-16	11.7	66.94	-	-
RNA(1,1,1)	-7.98	5.43	88.09		
MACD10	-12.28	6.98	80.14	21	0.85
MACD20	-7.10	5.39	85.41	19	1.27
MA10	-7.67	4.30	86.93	18	1.44
MA20	-7.73	4.89	87.69	24	1.11
MA30	-7.76	5.73	88.54	19	1.46
MA3-20	-12.00	7.87	79.16	24	0.69
MA3-30	-6.97	4.01	91.46	17	1.81
MA3-40	-7.79	5.13	89.12	17	1.66
stoch(21)	-7.11	5.32	88.56	19	1.46
TRIX30	-5.97	3.71	91.80	16	1.95
TRIX40	-6.71	4.40	90.40	18	1.65
ADX	-5.86	3.89	92.26	17	1.86

Notas:

- 1) Média aritmética dos retornos
- 2) Desvio Padrão
- 3) Valor final da operação supondo um investimento inicial de 100 dólares
- 4) Número de trocas entre renda fixa e ações.
- 5) Custo de Transação que elimina a lucratividade da estratégia

2.6. Conclusão

Este ensaio procurou verificar se inputs de análise técnica podem melhorar a capacidade de previsão de uma rede neural previamente estimada apenas com defasagens como variável explicativa. A maior parte das estatísticas fora-da-amostra não apresenta melhora significativa. O retorno total de algumas estratégias apresenta um desempenho um pouco mais favorável. No entanto, em alguns casos, a inclusão deste novo input melhora a relação entre risco e retorno indicada pelo retorno médio da operação e o desvio-padrão dos sinais gerados pela rede neural.

Isso não significa necessariamente que os inputs de análise técnica não possam gerar, através de arquiteturas de redes neurais, um desempenho superior quando comparado com uma formulação mais simples da RNA. O estudo sugere apenas que a pesquisa por uma arquitetura de rede que contenha apenas retornos defasados pode ser o suficiente para lidar com o processo não linear de uma série financeira.

Ensaio 3

Testando a Capacidade de Previsão de Diferentes Arquiteturas de Redes Neurais Artificiais : Uma Comparação entre diferentes extensões do algoritmo de retropropagação:

3.1.Introdução

O objetivo deste ensaio é comparar a RNA(1,1,1), gerada no capítulo 1, com diferentes arquiteturas que são extensões do algoritmo de retropropagação. Para a escolha desta rede foram utilizados os critérios de informação expostos no ensaio 1. Optou-se por esta estratégia uma vez que existem críticas ao uso do Erro Quadrático Médio como variável de escolha da arquitetura ótima das redes.

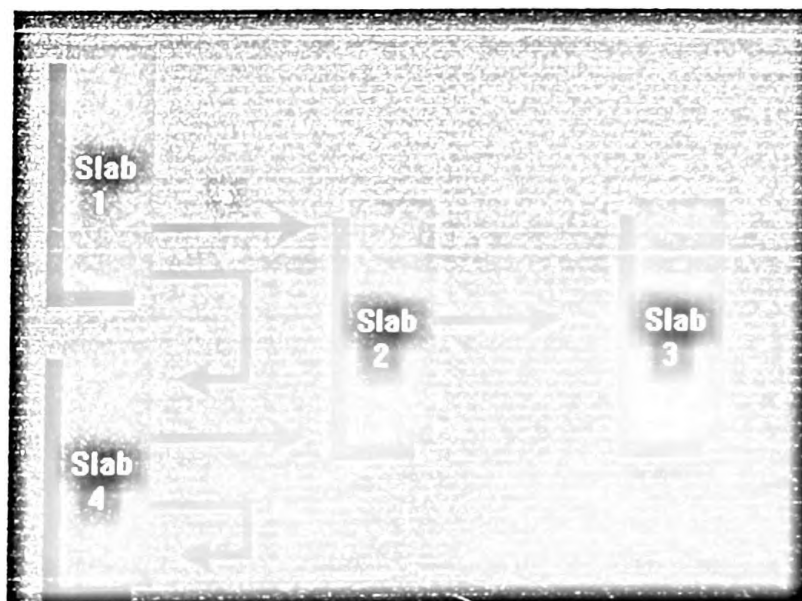
As extensões são as redes recorrentes e as redes com saltos de conexão. O ensaio está dividido em 4 partes, além desta introdução. A segunda parte trata da teoria das redes recorrentes e das redes com saltos. A terceira parte estima e discute os resultados das quatro redes. Esta parte está subdividida em dois subitens. O primeiro estima as redes usando apenas a defasagem de $t-1$ como variável explicativa. O segundo testa a capacidade de uma regra de análise técnica de melhorar as estimações. A quarta parte está dedicada às conclusões do ensaio.

Roman & Jamel(1996) propõem uma metodologia em que comparam a utilização de redes neurais com retropropagação e redes neurais recorrentes na escolha de um portfólio de ações global, obtendo resultados promissores em cinco mercados de ações: Canadá, Hong Kong, Japão, Inglaterra e Estados Unidos. O principal argumento para a

utilização de redes recorrentes é a sua estrutura temporal. Se as flutuações dos retornos são afetadas por seu histórico recente, as mudanças no preço dos ativos financeiros podem ser melhor aprendidas por uma rede que empregue algum mecanismo de feedback. Enquanto as redes que utilizam o algoritmo de retropropagação não levam em consideração a sequência dos dados, as redes recorrentes consideram esta sequência.

Estas redes podem ser visualizadas através dos diagramas abaixo.

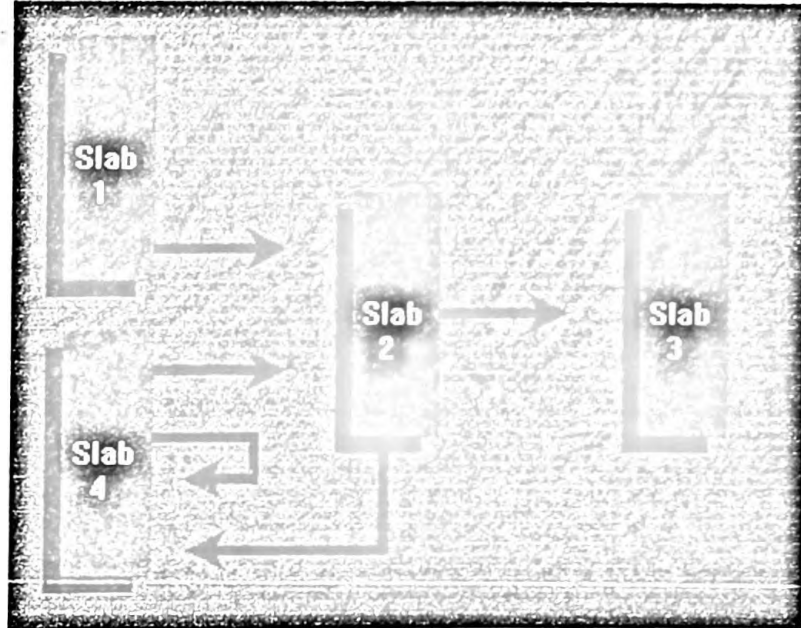
Figura 3.1: RR(1)



A figura 3.1 mostra uma rede em que a memória de longo prazo “lembra” dos novos inputs e usa estes dados quando um novo padrão é processado¹.

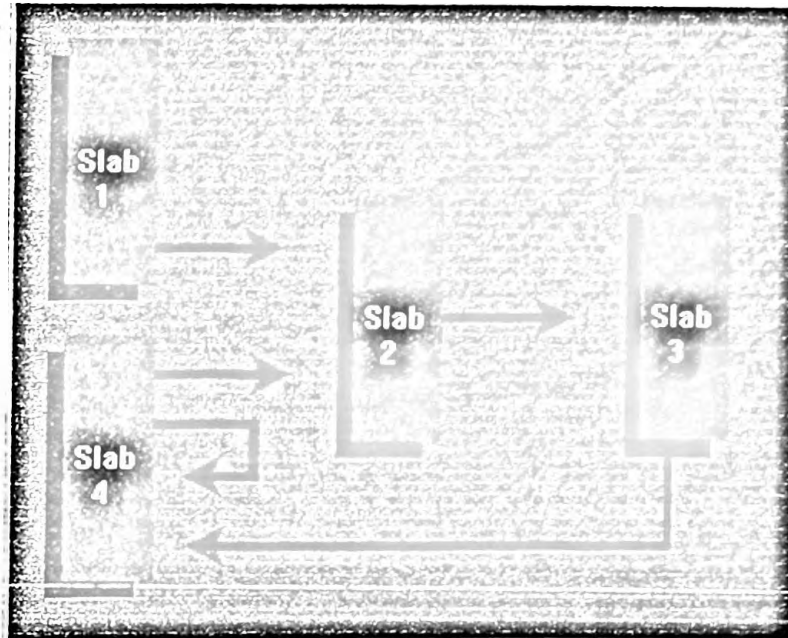
¹ Esta rede recorrente é um algoritmo proprietário da Ward Systems que detém os direitos sobre o NeuroShell

Figura 3.2: RR(2)



A figura 3.2 mostra uma rede em que a memória de longo prazo “relembra” as unidades escondidas, que já contêm as características dos dados dos padrões anteriores. Esta rede também é chamada de rede de Elman.

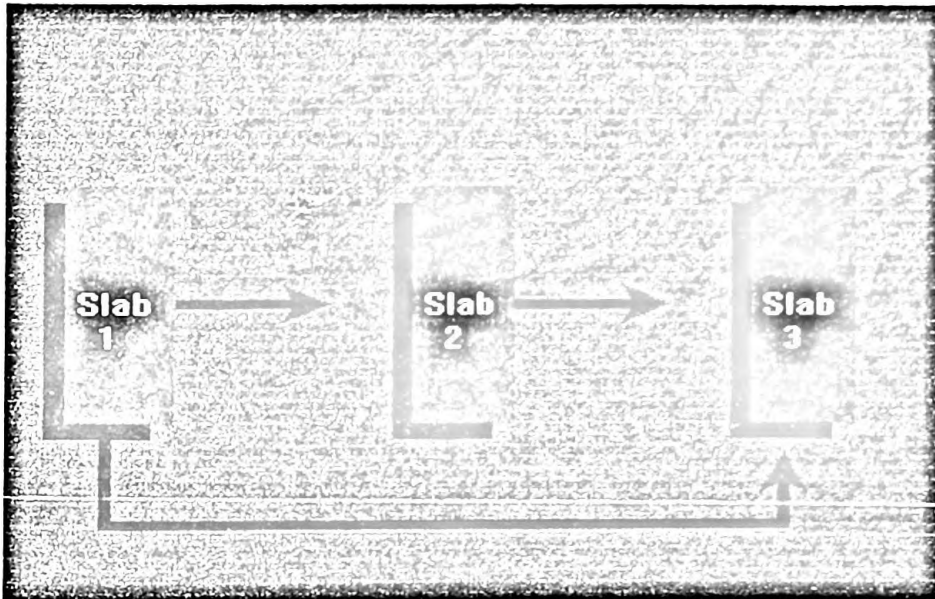
Figura 3.3: RR(3)



Na figura 3.3 a memória de longo prazo relembra os outputs e os reutiliza como inputs novamente. Esta rede também é chamada de Rede de Jordan.

Uma rede recorrente é treinada da mesma forma que as redes do primeiro ensaio. No entanto, os padrões devem ser apresentados em ordem não podendo haver aleatoriedade na seleção. Uma outra diferença é a presença de uma camada extra de inputs. Esta camada extra também é chamada de “memória de longo prazo”.

Figura 3.4: Rede com 3 camadas e saltos



A figura 3.4 mostra o diagrama de uma rede que utiliza o algoritmo de retropropagação do primeiro ensaio mas apresenta uma conexão direta entre a camada de entrada e a camada de saída.

3.2. Aspectos Teóricos das Diferentes Arquiteturas

Como visto no primeiro ensaio, os parâmetros das redes neurais são computados usando-se a técnica de retropropagação ou a regra delta generalizada. Frances e VanDick (2000) mostram que o aprendizado pode ser incorporado através da força das conexões, θ .

$$\theta_t = \theta_{t-1} + \Delta_t \quad (3.1)$$

onde θ_t são parâmetros usados no período t , θ_{t-1} são os pesos da rede em períodos anteriores e Δ_t reflete o processo de aprendizagem e é definido pela diferença entre o valor efetivo e o valor estimado do output.

A experiência de aprendizado da rede neural consiste na atualização de Δ_t de forma a reduzir o erro feito em cada período. A RNA aprende por experiência e Δ_t pode ser escrito como

$$\Delta_t = \nabla f(x_{t-1}, \theta_{t-1}) [y_t - f(x_{t-1}, \theta_{t-1})] \quad (3.2)$$

$$\theta_t = \theta_{t-1} + \lambda \nabla F(x_{t-1}, \theta_{t-1}) [y_t - F(x_{t-1}, \theta_{t-1})] \quad (3.3)$$

onde

$\nabla F(x_{t-1}, \theta_{t-1})$ é o gradiente de F com respeito ao peso θ , $y_t - F(x_t, \theta_{t-1})$ é o erro da rede, e λ é a taxa de aprendizado, independente do tempo.

No entanto, o termo de aprendizagem pode ser dependente do tempo de forma que

$$\theta_t = \theta_{t-1} + \lambda_t \nabla F(x_{t-1}, \theta_{t-1}) [y_t - F(x_{t-1}, \theta_{t-1})] \quad (3.4)$$

No caso das redes recorrentes podemos distinguir dois tipos básicos : Redes Recorrentes de Jordan (RRJ) e as Redes Recorrentes de Elman(RRE).

A estimação dos coeficientes é a mesma para estes dois tipos de redes.

Seja a seguinte equação:

$$\theta_{t+1} = \theta_t + \lambda_t \nabla F(x_t, \theta_t) [y_t - F(x_t, \theta_t)] \quad (3.5)$$

onde y é o valor real da variável a ser estimada e $F()$ é o valor do output computado. Se o termo de erro é $e_t(\theta) = y_t - F(x_t, \theta_t)$, então o gradiente necessário para o aprendizado é $\nabla e_t(\theta) = \nabla F(x_t, \theta_t)$

A equação pode ser reescrita de forma que

$$\theta_{t+1} = \theta_t + \lambda_t \nabla e_t(\theta) e_t(\theta) \quad (3.6)$$

Por hipótese a função da camada de saída é

$$Y_t = F(\alpha + H_t \beta q) \quad (3.7)$$

onde H é o vetor de variáveis escondidas, F é uma função identidade $F(x)=x$. A rede recorrente difere da apresentada no ensaio 1, com relação à forma funcional da camada escondida. A rede recorrente inclui as unidades da camada de saída ou das camadas escondidas como parte dos inputs.

Assim, a função da camada escondida pode ser escrita como segue:

$$H_t = G(\gamma x_t + \delta R_t) \quad (3.8)$$

onde

x_t : é o vetor de inputs

R_t : é o vetor de variáveis recorrentes

γ e δ são os pesos conectando inputs e unidades recorrentes das unidades inputs

G é a função de transformação (ou ativação) não-linear

Se é o produto (output) que “retorna” para a camada escondida, a rede será RRJ, então $R_t = Y_{t-1}$.

$$R_t = F\left(\alpha + \sum_{j=1}^J \beta_j G(\gamma x_{t-1} + \delta R_{t-1})\right) \quad (3.9)$$

Se a unidade escondida retorna para si mesma, tem-se uma RRE, $R_t = H_{t-1}$.

$$R_t = G(\gamma X_{t-1} + \delta R_{t-1}) \quad (3.10)$$

As equações acima indicam que a variável recorrente é uma função das variáveis defasadas X e R e dos parâmetros θ .

$$R_t = \rho(X_{t-1}, R_{t-1}, \theta) \quad (3.11)$$

Desde que R_t depende do valor efetivo da variável endógena, a equação acima torna-se $R_t = \rho(Z_{t-1}, R_{t-1}, \theta)$, onde $Z_t = (T_t, X_t)$ e $g(Z_t, \theta) = \nabla e_t(\theta) e_t(\theta)$. Como $e_t(\theta) = y_t - f(X_t, \theta)$, o termo de erro depende da unidade de saída efetiva, das unidades da camada de inputs, da camada recorrente e dos pesos. Assim tem-se

$$e_t = u(Z_t, R_t, \theta) \quad (3.12)$$

Khuan e White (1994) mostram que é computacionalmente complexa a estimação dos parâmetros θ em termos de ∇e_t e e_t diretamente, porque eles incluem a história inteira dos outputs efetivos e dos inputs. Pode-se derivar o algoritmo de aprendizado através de um método recursivo:

$$R_{t-1} = \rho(Z_{t-2}, R_{t-2}, \theta) \quad (3.13)$$

$$R_t = \rho(Z_{t-1}, l_t(Z_{t-2}, \dots, \theta), \theta) \equiv l_t(Z^{t-1}, \theta) \quad (3.14)$$

R_t é assim uma função da totalidade da história das variáveis da camada de saída e inputs Z^{t-1} , e dos coeficientes. A equação assim torna-se

$$e_t = u(Z_t, l_t(Z^{t-1}, \theta), \theta) \quad (3.15)$$

$$\nabla e_t = u_\theta(Z_t, l_t(Z^{t-1}, \theta), \theta) + \nabla l_t(Z^{t-1}, \theta) l_r(Z_t, l_t(Z^{t-1}, \theta), \theta) \quad (3.16)$$

em que

$$\nabla l_t(Z^{t-1}, \theta) = \rho_\theta(Z_{t-1}, R_{t-1}, \theta) + \nabla l_{t-2}(Z^{t-1}, \theta) \rho_r(Z_{t-1}, R_{t-1}, \theta) \quad (3.17)$$

por hipótese

$$\Delta_t = \nabla l_t(Z^{t-1}, \theta)$$

então,

$$\Delta_t = \rho_\theta(Z_{t-1}, R_{t-1}, \theta) + \Delta_{t-1} \rho_r(Z_{t-1}, R_{t-1}, \theta) \quad (3.18)$$

em que,

e_t : erro da RNA

∇e_t : gradiente de aprendizado

u_θ : derivada de u em relação a θ

- u_r : derivada de u em relação a R
- ρ_θ : derivada de ρ em relação a θ ,
- ρ_r : derivada de ρ em relação a R e
- Δ_i : matriz de gradiente em relação a θ

Se os valores dos \hat{R}_0 recorrentes, pesos e a matriz de gradientes são conhecidos, as estimativas dos coeficientes podem ser computadas usando um método de aprendizado recursivo.

$$\hat{e} = u(Z_t, \hat{R}_t, \hat{\theta}) \quad (3.19)$$

$$\nabla \hat{e}_t = u_\theta(Z_t, \hat{R}_t, \hat{\theta}) + \hat{\Delta}_t u_r(Z_t, \hat{R}_t, \hat{\theta}) \quad (3.20)$$

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \lambda_t \nabla \hat{e}_t \quad (3.21)$$

$$\hat{R}_{t+1} = \rho(Z_t, \hat{R}_t, \hat{\theta}) \quad (3.22)$$

$$\Delta_{t+1} = \rho_\theta(Z_t, \hat{R}_t, \hat{\theta}) + \Delta_t \rho_r(Z_t, \hat{R}_t, \hat{\theta}) \quad (3.23)$$

que é o algoritmo de retropropagação para redes recorrentes.

Outra variação do algoritmo de retropropagação permite “saltos” entre as conexões das diferentes camadas. O algoritmo pode ser resumido da seguinte forma:

$$n_{k,t} = \omega_{k,0} + \sum_{i=1}^{i^*} \omega_{k,i} x_{i,t} \quad (3.24)$$

$$N_{k,t} = \frac{1}{1 + e^{-n_{k,t}}} \quad (3.25)$$

$$y_t = \gamma_0 + \sum_{k=1}^{k^*} \gamma_k N_{k,t} + \sum_{i=1}^{i^*} \beta_i x_{i,t}$$

O output é uma combinação dos inputs processados na camada escondida e diretamente de forma linear dos inputs da primeira camada.

Segundo McNelis (2003, p.17) as redes com saltos aumentam o número de parâmetros, pois consideram o número de variáveis explicativas duas vezes, uma vez que existem conexões entre a camada de entrada e a camada escondida e diretamente entre a camada de inputs e a camada de output. Uma vantagem deste tipo de rede é a possibilidade de decompor a rede em um modelo linear e uma rede neural para a previsão. McNelis (2003) também afirma que uma outra aplicação prática para a rede com saltos é a possibilidade de testar não linearidades entre os inputs e outputs. Sugere então que primeiro se estime um modelo linear por mínimos quadrados e obtenha os resíduos. O passo seguinte é estimar uma rede neural em que os resíduos da regressão linear sejam as variáveis dependentes; os inputs são os mesmos da regressão linear. Se ainda existirem não linearidades, a rede deverá apresentar um poder explicativo significativo.

3.3.. Estimação e Resultados

A tabela 3.1 apresenta as estatísticas dentro da amostra das redes recorrentes estimadas. As redes foram estimadas usando a mesma base de dados do ensaio 1, ou seja, a série histórica do retorno do índice Bovespa denominado em dólares. A variável explicativa de todas as redes é a primeira defasagem do retorno do índice em dólares. Nota-se que não existe uma melhoria significativa no erro quadrático médio (MSE).

Tabela 3.1: Estatísticas dentro da amostra – Redes Recorrentes e com Salto

	R2	r2	MSE
RNA(1,1,1)	0.0144	0.149	9.194
RR(1)	0.0139	0.0146	9.198
RR(2)	0.0151	0.0157	9.187
RR(3)	0.0145	0.0151	9.193
RNA(jump)	0.0046	0.0049	9.285

Nota: A RNA(jump) se refere ao algoritmo de retropropagação com saltos.

Tabela 3.2: Estatísticas fora da amostra- Redes Recorrentes e com Saltos

	RMSE	MAE	MAPE	SIGN
RNA(1,1,1)	2.861496	2.28576	1.047217	0.511628
RR(1)	2.875172	2.288011	1.046769	0.581395
RR(2)	2.857225	2.282874	1.04575	0.488372
RR(3)	2.866768	2.28912	1.056047	0.488372
RNA(jump)	2.863987	2.286238	1.049057	0.465116

No entanto, a tabela 3.2 mostra que os resultados fora da amostra parecem promissores, já que uma das redes apresentou uma melhoria significativa na quantidade de sinais corretos. Esta melhoria também se refletiu no desempenho financeiro da estratégia como pode ser verificado no gráfico 3.1.

Gráfico 3.1

Resultado das Diferentes Redes para cada \$100 investidos

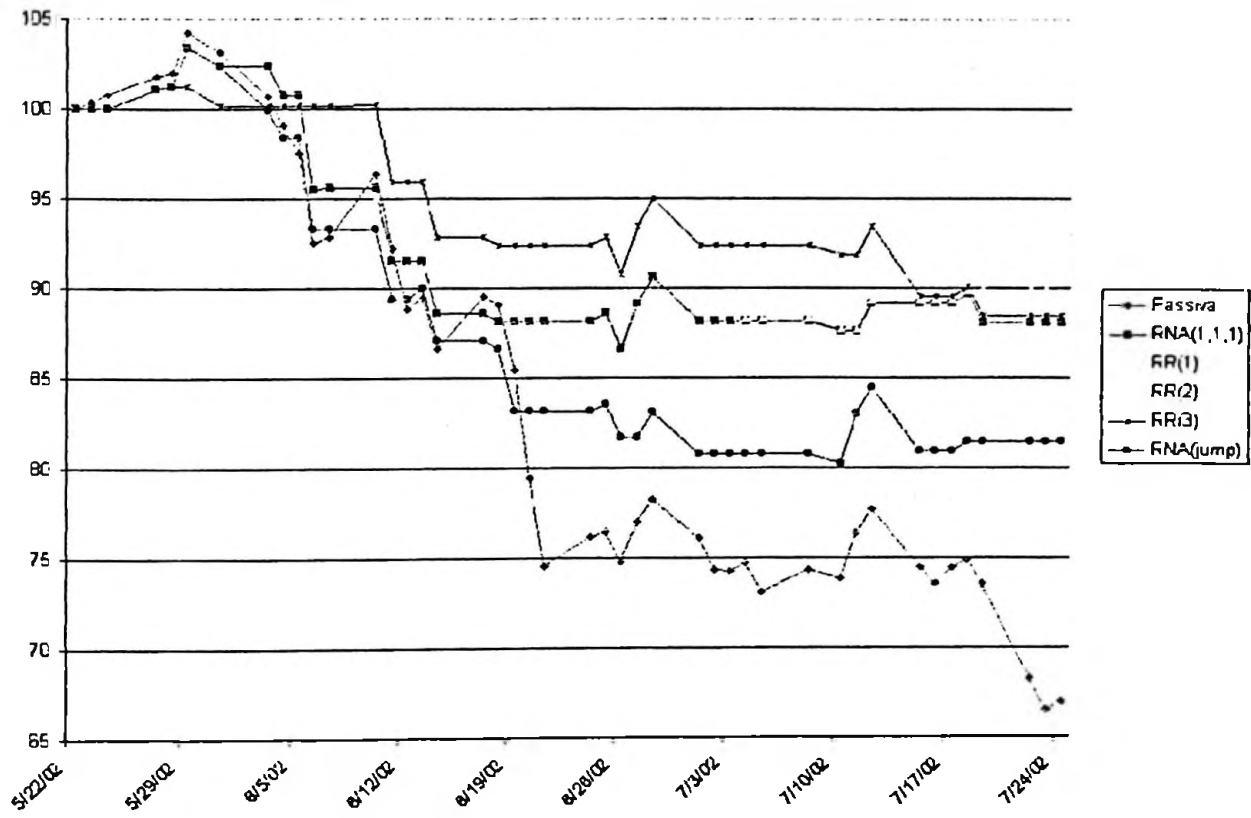


Tabela 3.3: Avaliação das Previsões Um Passo à Frente das Arquiteturas de RNAs Comparando a Estratégia Passiva com as Diferentes Estratégias Ativas

	Retorno Médio(1)	D.P.(%)(2)	Valor Final\$(3)	Mudanças (4)	Break-even(5)
Passiva	-16	11.7	66.94	-	-
RNA(1,1,1)	-7.98	5.43	88.09	20	1.36
RR(1)	-8.42	4.78	90.60	22	1.37
RR(2)	-6.60	4.46	88.66	19	1.47
RR(3)	-5.53	4.26	88.47	18	1.54
RNA(jump)	-12.29	7.87	81.44	18	1.08

Notas:

- 1) Média aritmética dos retornos
- 2) Desvio Padrão
- 3) Valor final da operação supondo um investimento inicial de 100 dólares
- 4) Número de trocas entre renda fixa e ações.
- 5) Custo de Transação que elimina a lucratividade da estratégia

A tabela 3.3 mostra que, apesar de gerar mais sinais corretos, a RR(1) tem um desempenho, em termos de retorno médio, pior do que a RNA(1,1,1) que utiliza a forma mais simples do algoritmo de retropropagação; mesmo assim todas as redes estimadas tem um desempenho superior à estratégia passiva. RR(2) e RR(3) apresentam uma performance mais rentável, reduzindo também o risco não apenas quando comparadas com a estratégia passiva mas também em relação à RNA(1,1,1). A rede neural com saltos nas conexões teve um desempenho relativamente medíocre quando comparada com as outras redes. Os custos de break-even permitem afirmar que estes modelos podem gerar ganhos reais quando implementados.

No ensaio 2, expandiu-se a RNA(1,1,1) testando a melhoria da previsão desta rede quando se acrescentava uma regra de análise gráfica. O resultado não foi tão expressivo em termos das medidas fora da amostra e levemente positivo em relação à

relação risco retorno. O mesmo procedimento será repetido para todas as 4 redes. A regra escolhida como variável explicativa além da primeira defasagem é o estudo chamado de ADX, que procura detectar tendências na série.

Gráfico 3.2

Resultados das Diferentes Redes com uma Regra de Análise Técnica para cada 100 dólares investidos

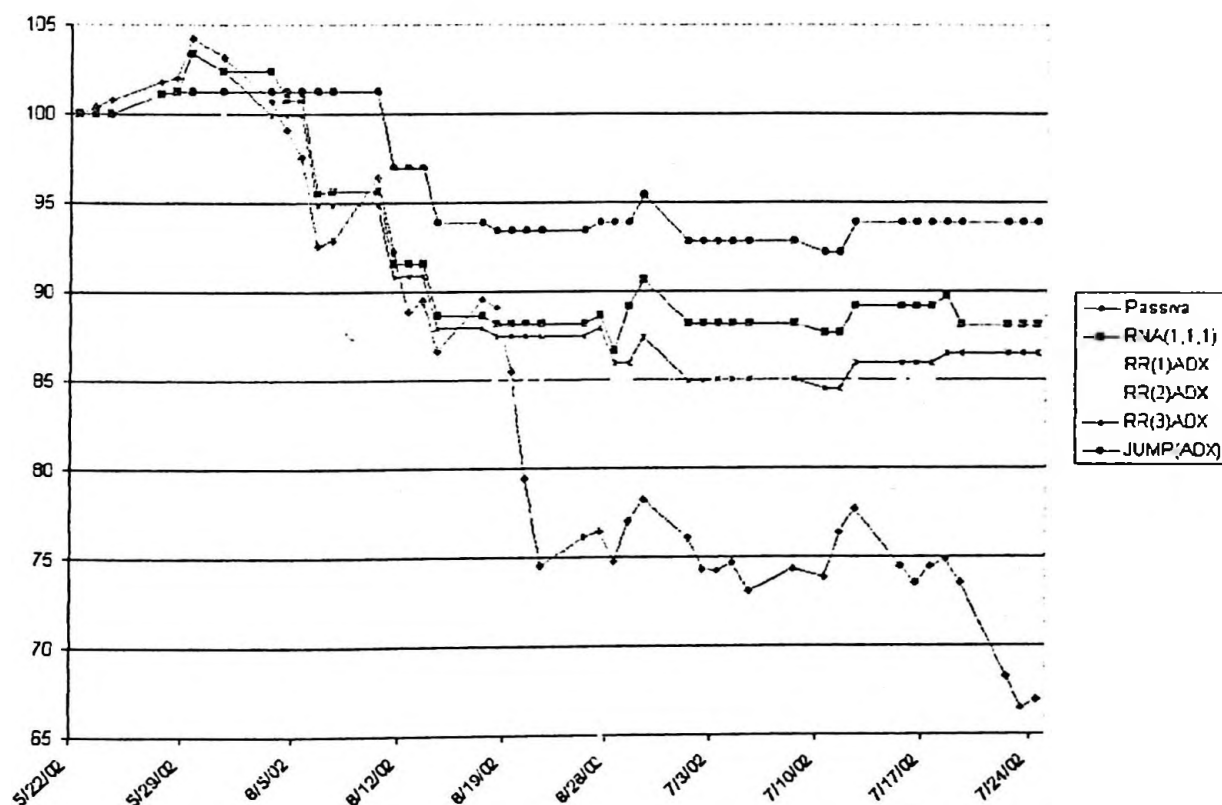


Tabela 3.4: Avaliação das Previsões Um Passo à Frente das Arquiteturas de RNAs Comparando a Estratégia Passiva com as Diferentes Estratégias Ativas considerando a geração de sinais da variável de análise técnica

	Retorno Médio(1)	D.P.(%)(2)	Valor Final\$(3)	Mudanças (4)	Break-even(5)
Passiva	-16	11.7	66.94	-	-
RNA(1,1,1)	-7.98	5.43	88.09	20	1.36
RR(1)ADX	-10.82	6.25	86.36	22	1.15
RR(2)ADX	-10.57	6.28	85.47	21	1.16
RR(3)ADX	-9.55	6.23	86.46	18	1.41
JUMP(ADX)	-4.09	3.45	93.86	15	2.23

Notas:

- 1) Média aritmética dos retornos
- 2) Desvio Padrão
- 3) Valor final da operação supondo um investimento inicial de 100 dólares
- 4) Número de trocas entre renda fixa e ações.
- 5) Custo de Transação que elimina a lucratividade da estratégia

Na tabela 3.4 pode-se verificar o resultado do uso do ADX nas diferentes redes. Nota-se que o desempenho de todas as redes recorrentes é pior do que quando estas são estimadas apenas com a desagem do retorno. No entanto, o resultado é diferente quando a RNA tem saltos entre as conexões; neste caso, o retorno médio e a relação risco/ retorno são superiores. A rentabilidade desta estratégia está garantida não apenas pelo alto retorno mas também pelo menor número de mudanças entre ações e renda fixa.

3.4. Conclusões

Os resultados obtidos neste ensaio indicam que as redes recorrentes podem capturar melhor as não linearidades da série estudada uma vez que estas consideram a sequência de dados e a ordem dos padrões. Os resultados das redes recorrentes foram superiores à rede estimada que utiliza a forma mais simples do algoritmo de retropropagação (RNA(1,1,1)). A inclusão de uma regra de análise técnica como input nas redes recorrentes resultou em danos ao poder de geração de sinais corretos destas redes, bem como ao seus resultados financeiros. No entanto, no caso das redes com salto o desempenho foi extremamente positivo.

Na estimação com apenas uma defasagem a rede com saltos foi incapaz de processar as não-linearidades da série de retornos do índice Bovespa. A introdução do ADX possibilitou a captação da não linearidade da série.

Conclusão Final

O principal objetivo do trabalho era, através de três ensaios, testar a eficácia da geração de sinais de compra e venda de ativos, das diferentes arquiteturas de redes neurais. O bom desempenho das RNAs simultaneamente testaria a eficiência relativa de alguns mercados de ações latino americanos. As variáveis explicativas, ou inputs, foram as defasagens dos retornos em dólares, e sinais de compra e venda gerados por indicadores de análise técnica.

A conclusão final é de que dependendo da arquitetura da rede, uma variável pode ou não melhorar a capacidade de previsão. O ADX no ensaio 2 foi capaz de melhorar o retorno médio da RNA(1,1,1) , piorou a performance no caso das redes recorrentes e melhorou sensivelmente o desempenho da rede com saltos. O mesmo poderia ter acontecido para outras regras e outras arquiteturas. Mesmo tendo estimado centenas de diferentes combinações, não se pode garantir que as redes apresentadas neste trabalho são as melhores.

Uma sugestão para a continuação desta linha de pesquisa seria a de estimar todas as combinações de arquiteturas de redes recorrentes, bem como testar o uso de algoritmos genéticos para a escolha das variáveis relevantes.

Bibliografia

- Adya M. & Collopy, F. (1998). "How effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation" *Journal of Forecasting* 17 p. 481-495.
- Allen, F. & Karjalainen R. (1998). "Using Genetic Algorithms to Find Technical Trading Rules" *Journal of Financial Economics* 51, p.245-271.
- Baestaens, D.J. Van Den Berg W.N. & Vaudrey.H.(1996). "Market Inefficiencies, Technical Trading and Neural Networks" In: Christian Dunis (Ed.). *Forecasting Financial Markets: Exchange Rates, Interest Rates and Asset Management*. p.245-259.
- Bekaert, G. E Harvey C. (1995). "Time -Varying World Market Integration" *Journal of Finance*, 1995 p.403-444.
- Bishop, C. (1995) *Neural Networks for Pattern Recognition* . Oxford University Press- Great Britain.
- Brock, W.A, Hisieh, D. and LeBaron, B.(1991) *A Test for Nonlinear Dynamics, Chaos and Instability*, Cambridge, Mass: MIT Press.
- Brock, W.A. e Dechert, W.D.(1991). "Non-Linear Dynamical Systems: Instability and Chaos in Economics" .In W. Hildenbrand e H.Sonnenschein(eds.), *Handbook of Mathematical Economics*, Amsterdam: North-Holland, p. 2209-35.
- Brock, W.A.(1986). "Distinguishing Randon and Deterministic Systems: Abridge Version", *Journal of Economic Theory*, 40, p.168-95.
- Brock,W, Lakonishok J & Lebaron B.(1992). "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns", *Journal of Finance*, 47(5), pp 1731-1764.

- Bueno, N.P.(2002). “Algumas evidências da Presença de Não-linearidades Compatíveis com Caos Determinístico no Ibovespa na Década de 1990”. *Economia Aplicada* Vol.6- no. 3 – Julho-Setembro 2002, p.555-574.
- Burrascano, P.(1991) “Learning Vector Quantization for the Probabilistic Neural Network”. *IEEE Trans. on Neural Networks*, July 1991, 2, 458-461.
- Campbell, J. Lo, W.& MacKinlay, C. (1997). *The Econometrics of Financial Markets*. Princeton University Press. NJ. USA.
- Ceretta P.S & Costa Júnior N.C.A.(2002). “Caminho Aleatório: revisão, síntese e novos resultados nos mercados de capitais da América Latina”. *Revista de Administração da USP*, Volume 37, número 3.- julho setembro de 2002, p.47-56.
- Chang, P.W. & Osler C.L.(1999). “Methodical Madness: Technical Analysis and Irrationality of Exchange-Rate Forecasts”, *Economic Journal* 109, p. 636-661.
- Chuah, K.L.(1993). “A Nonlinear Approach to Return Predictability in the Securities Markets Using Feedforward Neural Network Washington University mimeo.
- Cootner, P.(1974) *The Random Character of Stock Market Prices*, Cambridge, MA, MIT Press.
- Curcio, R. & Goodhart, C.(1992).”When Support/Resistance Levels are Broken, Can Profits be Made? Evidence from the Foreign Exchange Market” – *Discussion Paper No. 142. LSE Financial Markets Group*. – London School of Economics.
- Dechert, W.D(1996).”Testing Time Series for Nonlinearities: The BDS Approach”, In: W.A Barnett, A.P.Kirman e M.Salmon(eds), *Nonlinear Dynamics and Economics*, Cambridge University Press, p.191-200.

- Diebold, F. (1998). *Elements of Forecasting*. South-Western College Publishing, Cincinnati, Ohio, USA, 1998.
- Edwards R. & Magee J. (1966). *Technical Analysis of Stock Trends* (revised 5th edition), John Magee, Boston.
- Fama E.F (1965). "The Behavior of Stock Market Prices", *Journal of Business*, 38, p.34-105.
- Fama, E. & Blume, M.(1966) "Filter rules and stock market trading profits." *Journal of Business*, 39, p,226-241.
- Fausett L.(1994). *Fundamentals of Neural Networks- Architectures, algorithms, and applications*. New Jersey: Prentice Hall.
- Franses, P.H& Dijk, D. (2000). *Non-linear time series models in empirical finance*. Cambridge University Press..
- French, K.R. (1980) Stocks returns and the weekend effect. *Journal of Financial Economics*, 8, p.55-70
- Gately. E(1996) *Neural Networks for Financial Forecasting – Wiley Trader’s Advantage Series*. John Willey and Sons, Inc.
- Gençay R.& Stengos, T.(1998). "Moving Average Rules, Volume and the Predictability of Security Returns with Feedforward Networks" . *Journal of Forecasting* 17, p.401-414.
- Gençay R.(1998). "The predictability of security returns with simple technical trading rules" *Journal of Empirical Finance* 5 p.347-359.
- Golden, R.M. (1996). *Mathematical Methods for Neural Network Analysis and Design*. MIT Press USA.

- Gourieoux, O. & Jasiack J. (2001). *Financial Econometrics- Problems, Models and Methods* – Princeton Series in Finance – Princeton University Press.
- Grossman, S. & Stiglitz, J. (1980) “On the possibility of informationally efficient markets.” *American Economic Review*, 70, p.393-408.
- Grossman,S. (1976) “On the efficiency of competitive stock markets where trades have diverse information”. *Journal of Finance*, 31, p.573-585.
- Harvey, C.(1995) “Predictable Risk and Returns in Emerging Markets” *Review of Financial Studies*, 1995 p.773-816.
- Haykin, S.(1999). *Redes Neurais – Princípios e Práticas – 2ª edição*. Porto Alegre: Bookman
- Hebb, D.O (1949). *The Organization of Behavior*. New York :Wiley apud Golden R.M. *Mathematical Methods for Neural Network Analysis and Design*. Cabridge MIT Press.
- Hecht-Nielsen, R.(1988). “Applications of Counterpropagation Networks”, *Neural Networks*, vol 1, p. 131-140 apud Portugal M.S. & Fernandes L.G (1996) *Redes Neurais Artificiais e Previsão de Séries Econômicas*”.
- Hertz, J. , Krogh A. e Palmer, R.G (1991). *Introduction to the Theory of Neural Computation* (Addison-Wesley Publishing Company).
- Hudson, J.K.(1983). “Good TRIX”, *Technical Analysis of Stocks & Commodities* vol1.
- Hutton, L. “Using Statistics to Assess the Performance of Neural Network Classifiers” *Johns Hopkins Applied Physics Lab Technical Digest*, Vol 13, No. 1, 1992.
- Hwang, J.T. e Ding, A.A.(1997). “Prediction Intervals for Artificial Neural Networks,” *Journal of the American Statistical Association*, 92 p.748-757.

- Jennergreen, R. & Korsvold, P.(1975) "The non-random character of Norwegian and Swedish stock market prices. In *International Capital Markets* (eds. E.J. Elton e M.J. Gruber), Amsterdam, North Holland.
- Kaashoek, J.F. van Dick, H.(1998). "A simple strategy to prune neural networks with application to economic time series", *Report 9854/A*, Econometric Institute Erasmus University Rotterdam.
- Kaashoek, J.F & van Dick,H. (1999). "Neural Network analysis of varying trends in real exchange rates", *Report EI9915/A*. Econometric Institute. Erasmus University Rotterdam.
- Kaashoek, J.F. & van Dijk H. (2001) . "Neural Networks as econometric tool"- *Economic Institute Report EI 2001-5*.
- Kaufman, P.J.(1998). *Trading Systems and Methods – Third Edition – Wiley Trading Advantage*. John Willey & Sons.
- Kuan C.M. & White H. (1994). "Artificial neural networks: an econometric perspective". *Econometric Reviews*, 13, pp. 1-143.
- Lang, K. and Whitbrock, M. Learning to Tell Two Spirals Apart. *Proceedings of the 1988 Connectionist Models Summer School*. Touretzky, D., Hinton, G., and Sejnowski, T. (eds.), San Mateo, Ca.: Morgan Kaufmann, 1989, p.52-59.
- Leal, R.P.C. & Varanda, A.J.(2000). "Desempenho de Estratégias Grafistas de Interseção de Médias Móveis através de simulação de Bootstrap."In: *Mercado de Capitais – Análise Empírica no Brasil*. Newton Carneiro Affonso da Costa Jr., Ricardo Pereira Camara Leal e Eduardo Facó Lemgruber Org. Editora Atlas, 2000.

- LeBaron, B.(1996) "Technical Trading Rule profitability and foreign exchange intervention." *WP 5505* NBER , Cambridge, MA.
- Levich, R. & Thomas L.(1993) "The significance of technical trading rule profits in the foreign exchange market: a bootstrap approach" *Journal of International Money and Finance* 12.p.451-474.
- Lo, A. W.& MacKinlay A.C. (1988) "Stock Market do not follow random walks: evidence from a simple specification test." *The Review of Financial Studies*, 1 (1), p.41-66
- Lo, A.W. & MacKinlay A.C (1989) "The size and power of variance ratio test in finite samples: a Monte Carlo investigation." *Journal of Econometrics*, 40, p. 203-238.
- Lo, A.W.. & MacKinlay A.C. (1999). "Maximizing Predictability in the Stock and Bond Markets" In : Lo A, W, & Mackinlay A.C (Eds) *A Non-Random Walk Down Wall Street* . Princeton University Press, 1999.
- Lo, A. W., Mamaysky, H. & Wang J. (2000) "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation" *NBER Working Paper 7613*.
- Martin, V. & Tan C. (1997). "Artificial Neural Networks" In: Creedy, J. & Martin, V. (eds) *Nonlinear Economic Models – Cross-sectional Time Series and Neural Network Applications* .Edward Elgar Publishing Limited, Great Britain. p. 213-240.
- McCulloch, W.S. & Pitts, W.(1943). "A logical calculus of the ideas immanent in nervous activity". *Bulletin of Mathematical Biophysics* 5: 115-133 apud Golden

- R.M. *Mathematical Methods for Neural Network Analysis and Design*. Cambridge MIT Press.
- McNelis P.(1997). *Neural Networks and Genetic Algorithms: Tools for Forecasting and Risk Analysis in Financial Markets*- mimeo.
- McNelis, P. , Camba-Mendez, G. & McAdam, P (2003) “Forecasting Inflation with Thick Models and Neural Networks” – may 2003- mimeo.
- McNelis, P.D.(2003) . *Precision for Real-Time Prediction: Financial Engineering with Neural Networks*. mimeo
- Medsker,L. ,Turban E. & Trippi R.(1996). “Neural Network Fundamentals for Financial Analysts” In: Trippi R.& Turban E. (eds.) *Neural Networks in Finance and Investing – Using Artificial Intelligence to Improve Real-World Performance*. Irwing Professional Publishing .p.3-24.
- Metastock(1997). *User’s Manual*. Equis International
- Mills, T.(1999) . *The Econometric Modelling of Financial Time Series* –Second Edition. Cambridge: Cambridge University Press
- Neely, C & Weller P. (2000). “Technical Analysis and Central Bank Intervention” *Working Paper 97-002c* – Federal Reserve Bank of St. Louis
- Neely, C. Weller P(1998). “Technical Trading Rules in the European Monetary System” *Journal of International Money and Finance* 18(3) p. 429-58.
- Neely, C. Weller P.& Dittmar, R.(1997). “Is Technical Analysis in the Foreign Exchange Market Profitable? A Genetic Programming Approach” *Journal of Financial and Quantitative Analysis*, 32, p.405-26.

- Neftci, S.N.(1991). "Naïve trading rules in financial markets and Winer-Kolmogorov prediction theory: a study of "technical analysis"", *Journal of Business* 64, p.549-71.
- Nishikawa, Y., Kita, H., and Kawamura, A. (1990). NN/I: A Neural Network Which Divides and Learns Environments. Proceedings of the International Joint Conference on Neural Networks. January 1990, 1, I-684 to I-687.
- Osler,C.(1998) "Identifying Noise Traders: The Head-and-Shoulders Pattern in U.S. Equities." *Federal Reserve Bank of New York* . mimeo
- Osler,C (2001) "Currency Orders and Exchange-Rate Dynamics: Explaining the Success of Technical Analysis"- Federal Reserve Bank of New York, mimeo
- Pesaram,M. H. & Timmermann, A. (1995)."Predictability of Stock Returns: Robustness and Economic Significance", *Journal of Finance*, 50, p.1201-1228.
- Portugal M. S. & Fernandes L.G (1996). "Redes Neurais Artificiais e Previsão de Séries Econômicas: Uma Introdução" mimeo.
- Puitt, S.W. & White R.E(1988). "The CRISMA Trading System: Who says technical analysis can't beat the market" *Journal of Portfolio Management*, Spring 55-58
- Qi, M. (1999). "Nonlinear Predictability of Stock Returns Using Financial and Economic Variables". *Journal of Business & Economic Statistics* Oct 1999, vol 17, no 4. p. 419-429.
- Qi, M.(2001) "Predicting US Recessions with leading indicators via neural network models" *International Journal of Forecasting* 17 . p. 383-401.

- Ratner M.& Leal, R.P.C.(1999). "Tests of Technical Trading Strategies in the emerging equity markets of Latin America and Asia." *Journal of Banking & Finance* 23 (1999) p.1887-1905.
- Roman J. E Jameel, A.(1996). "Backpropagation and Recurrent Neural Networks in Financial Analysis of Multiple Stock Market Returns" Proceedings of the 29th Annual Hawaii International Conference on System Sciences – IEEE.
- Rosenblatt L.(1958). "The Perceptron: a Probabilistic Model for Information Storage and Organization in the Brain" *Psychological Review*, 65- p.386-408. Apud Fausett(1994). *Fundamentals of Neural Networks- Architectures, algorithms, and applications* .New Jersey: Prentice Hall.
- Rumelhart, D., and McClelland, J.(1986). *Parallel Distributed Processing*. Cambridge, Mass.: The MIT Press, 1986.
- Samuelson.P(1965) "Proof that properly anticipated prices fluctuate randomly. *Industrial Management REVIEW*, 6, p.41-49.
- Shaaf. M. (2000) "Predicting Recession using the yield curve: An Artificial Intelligence and Econometric Comparison" *Eastern Economic Journal*, Vol.26, No.2, Spring 2000 p. 171-190.
- Simpson, P.(1990) *Artificial Neural Systems*. New York, N.Y.: Pergamon Press, 1990.
- Specht, D.(1988) "Probabilistic Neural Networks for Classification, Mapping, or Associative Memory." *Proceedings of the IEEE International Conference on Neural Networks*, 1988, 1, p.525-532.
- Specht, D. (1991) A General Regression Neural Network. *IEEE Trans. on Neural Networks*, Nov. 1991, 2, 6, p.568-576.

- Specht, D. (1990) Probabilistic Neural Networks. *Neural Networks*, 1990, 3, p.109-118.
- Specht, D. and Shapiro, P. (1991) Generalization Accuracy of Probabilistic Neural Networks Compared with Back-Propagation Networks. *Proceedings of the International Joint Conference on Neural Networks*, July 8-12, 1991, 1, p.887-892.
- Stock J.H. & Watson M.W. (1998). "A Comparison of Linear and NonLinear Univariate Models for Forecasting Macroeconomic Time Series" *NBER Working Paper Series – WP 6607*.
- Sullivan R., Timmermann A. & White H. (1998a) "Data-Snooping, Technical Trading Rule Performance, and Bootstrap." Discussion Paper 303- London School of Economics.
- Sullivan R, Timmermann A. & White H.(1998b) "The Dangers of Data-Driven Inference: The Case of Calendar Effects in Stock Returns". Discussion Paper 304 – London School of Economics.
- Swanson, N.R. & White H. (1995). "A model selection approach to assessing the information in the term structure using linear models and artificial neural networks". *Journal of Business and Economic Statistics* 13, p. 265-75.
- Swanson N.R & White H. (1997a). "Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models". *International Journal of Forecasting* 13, p.439-61.
- Swanson N.R. & White H.(1997b). "A Model Selection Approach to real time macroeconomic forecasting using linear models and artificial networks". *Review of Economics and Statistics* 79, p.540-50.

- Taylor, J. (2002). "Neural Networks". In : *Neural Networks and the Financial Markets – Predicting, Combining and Portfolio Optimization*. Ed: Jimmy Shadbolt e John G Taylor – *Perspectives in Neural Computing*. Springer.
- Torres, R., Bonomo, M. & Fernandes, C. (2002). "A aleatoriedade do passeio na Bovespa: testando a eficiência do mercado acionário brasileiro". *Finanças Aplicadas ao Brasil*. Marco Bonomo Organizador _ FGV editora.
- Touretzky, D. and Pomerleau, D. (1989). What's Hidden in the Hidden Layers? *Byte*, 14 1989, 227-233.
- Trippi R. & De Sieno D. (1992). "Trading Equity Index Futures with a Neural Network". *The Journal of Portfolio Management*. fall 1992. p.27-33.
- Trippi R.R. & Turban, E. (1996). *Neural Networks in Finance and Investing – Using Artificial Intelligence to Improve Real-World Performance*. Irwin Professional Publishing.
- Ward Systems (1996). *NeuralShell User's Manual*. Ward Systems.
- Wasserman, P. (1989). *Neural Computing, Theory and Practice*. New York, N.Y.: Van Nostrand Reinhold, 1989.
- White, H. (1988). "Economic Prediction using Neural Networks: The Case of IBM Daily Stock Returns." In: Trippi R. & Turban E. (eds.) *Neural Networks in Finance and Investing – Using Artificial Intelligence to Improve Real-World Performance*. Irwing Professional Publishing p.469-482
- White, H. (1990) "Connectionist NonParametric Regression: Multilayer Feedforward Networks Can Learn Arbitrary Mappings." *Neural Networks*, 3(5): 535-549.

White H.(1997). "A Reality Check for Data Snooping" San Diego, NRD Technical Report 97-01.

White, H. Gallant, A.R.,Hornik,K., Stinchcombe, M. e Wooldridge, J. (1992). Artificial Neural Networks: Approximation and Learning Theory. Cambridge. M.A:Blackwell.

Yao, J. & Tan, C.L.(2002) "Neural Networks for Technical Forecasting of Foreign Exchange Rates" p.189-204 In Neural Networks in Business: Techniques and Applications eds: Kate Smith e Jatinder Gupta. Idea Group Publishing.