UNIVERSIDADE DE SÃO PAULO

FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DE RIBEIRAO PRETO DEPARTAMENTO DE CONTABILIDADE

MODELOS DE PREVISÃO DE SÉRIES TEMPORAIS FINANCEIRAS COM COMBINAÇÃO DE FILTROS DE KALMAN E WAVELETS

FABIANO GUASTI LIMA

RIBEIRÃO PRETO 2011

FABIANO GUASTI LIMA

MODELOS DE PREVISÃO DE SÉRIES TEMPORAIS FINANCEIRAS COM COMBINAÇÃO DE FILTROS DE KALMAN E WAVELETS

Tese apresentada ao Concurso de Livre-Docência no Departamento de Contabilidade da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo.

RIBEIRÃO PRETO 2011

Prof. Dr. João Grandino Rodas REITOR DA UNIVERSIDADE DE SÃO PAULO

Prof. Dr. Sigismundo Bialoskorski NetoDIRETOR DA FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E
CONTABILIDADE DE RIBEIRÃO PRETO

Profa. Dra. Adriana Maria Procópio de Araújo CHEFE DO DEPARTAMENTO DE CONTABILIDADE DA FEA-RP

FICHA CATALOGRÁFICA

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Catalogação na fonte

Lima, Fabiano Guasti

L698m

Modelos de previsão de séries temporais financeiras com combinação de filtros de Kalman e Wavelets / Fabiano Guasti Lima. – 2010 151 p.

Tese de Livre-Docência – Departamento de Contabilidade da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo

Concurso de Livre-Docência

1. Previsão. 2. Filtro de Kalman. 3. Filtro de Wavelets. 4. Séries temporais. I. Universidade de São Paulo. II. Título.

CDU - 657.4:330.1

Ao meu filho Pedro.

AGRADECIMENTOS

Primeiramente devo agradecer a **DEUS**, que me proporcionou a graça de viver, aprender e crescer. E a **NOSSA SENHORA APARECIDA**, por ter permitido que tantas coisas boas se realizassem em minha vida e por sempre ter mostrado aquilo que é melhor para mim;

À minha mãe, profa. **Zulmira**, que sempre me dedicou amor, educação e carinho. Pelos seus ensinamentos com os quais aprendi a ter discernimento, sonhar, acreditar e por isso me orgulho de ser seu filho;

À minha família, minha esposa profa. **Milena** e meu filho **Pedro** pela paciência, apoio e carinho dedicados em todas as etapas da vida;

Ao prof. **Alexandre Assaf Neto,** pela grande e positiva influência que efetivamente exerceu em minha vida pessoal e profissional. Pelo apoio incondicional, pela humildade de sempre estar disposto a ensinar e ajudar este vosso filho;

À profa. **Adriana Maria Procópio de Araújo**, pelo incentivo e apoio constantes dado em minha carreira. Sempre disposta e pronta a ajudar, sem nunca desistir;

Ao prof. **Antônio Carlos da Silva Filho**, pela imensa contribuição em me auxiliar nas discussões e no desenvolvimento da biblioteca de modelos em Matlab para previsão de séries temporais e uso combinado dos filtros. Esteve sempre presente e solícito desde meu mestrado;

Aos amigos do Departamento de Contabilidade da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto que me receberam muitíssimo bem nessa casa;

E a todas as outras pessoas que, das mais diferentes maneiras, participaram desta história.

RESUMO

LIMA, Fabiano Guasti. **Modelos de previsão de séries temporais financeiras com combinação de filtros de Kalman e Wavelets**. 2010, 151 f. Tese de Livre Docência. Departamento de Contabilidade. Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto. Universidade de São Paulo, Ribeirão Preto, 2010.

O estudo sobre o comportamento das séries temporais financeira, com objetivo de previsão futura de preços e retornos, é foco de discussão e embates, mediante distintas abordagens. Especialmente ao se tratar dos movimentos implícitos dos retornos de um ativo financeiro, analisando seu comportamento, as teorias sobre os modelos de ajustes a esse comportamento são divergentes na busca da compreensão desses fatos empíricos. Entre as diversas técnicas que desempenharam esse papel de prever um valor para o futuro, existem diversos modelos já verificados na literatura e, mais recentemente, os filtros de separação de séries surgiram como uma alternativa complementar às atuais técnicas de previsão. Encontrou-se suporte na literatura que já vinha apontando para a necessidade de melhoria nos filtros de volatilidade em conjunto com as técnicas já desenvolvidas e testadas. O background levantado apontou para a existência de dois principais filtros: wavelets e Kalman. Desta forma, o objetivo geral deste estudo é realizar uma análise comparativa do uso combinado de filtros de wavelets e kalman juntamente com modelos de previsão para séries temporais financeiras, a fim de verificar qual produz a melhor previsão futura para mensuração de ativos. Para se investigar o assunto, buscou-se uma metodologia quantitativa e descritiva dos modelos e das formas combinadas de uso dos filtros para previsão. Os resultados apontaram que realmente o uso das técnicas de filtragem consegue reduzir o erro das previsões. Testada a junção das técnicas para uma série com alta volatilidade como o IBOVESPA, o resultado aponta o uso do filtro de Kalman primeiro e em seguida o uso de wavelets com redes neurais recorrentes, com erro medido pelo MAPE de 0,72%. Já para a série de uma commoditie que, teoricamente, apresenta uma volatilidade menor, o uso combinado dos filtros não trouxe grande melhora na redução do erro, todavia, o erro foi menor quando do uso de wavelets com redes neurais recorrentes com MAPE de 0,49%. Testando as possíveis variações na forma da wavelet de filtragem para checar a possível interferência nos resultados das previsões, chegou-se ao resultado que o erro somente é inferior para as formas de onda primária, no caso para a wavelet de "Haar" e "daubesch 1". Dessa forma, o estudo contribui para a área contábil pois demonstra redução de erros de previsão futura e consequente melhor gestão de riscos em posições de investimentos no mercado financeiro.

Palavras-Chave: Previsão; Filtros; Wavelet; Filtro de Kalman; Gestão de Risco.

ABSTRACT

LIMA, Fabiano Guasti. **FINANCIAL TIME SERIES FORECASTING MODELS WITH A COMBINATION OF KALMAN AND WAVELETS FILTERS**. 2010, 151 f. Associate Professorship Thesis. Department of Accounting. Faculty of Economics, Business Administration and Accounting of Ribeirão Preto. University of São Paulo, Ribeirão Preto, 2010.

The study about the behavior of financial time series with the purpose of forecasting prices and returns is the constant focus of discussions and conflicts under different approaches. Specially when dealing with the implied movements of the returns from a financial asset, when analyzing its behavior, the theories about the adjusting models to such behavior are divergent in the search for understanding these empirical facts. Among the numerous techniques that played the role of forecasting a future value, there are several different previously verified models found in literature and, more recently, the time series separation filters appeared as a complementary alternative to the current forecasting techniques. Support was found in literature, which had already been indicating the need of improving the volatility filters, along with techniques previously developed and tested. The presented background indicated the existence of two main filters: wavelets and Kalman. Therefore, the general purpose of this study is to carry out a comparative analysis of the combined use of wavelets and Kalman filters along with forecasting models for financial time series in order to verify which of them produces the best forecast. To investigate the subject, it was based upon quantitative and descriptive methodology of the models and of the combined forms of use of forecasting filters. The results showed that the use of filtration techniques is indeed able to reduce the forecasting errors. After testing the junction of techniques for a high volatility time series, such as IBOVESPA, the results first indicate the use of the Kalman filter and next the use of wavelets with recurring neural networks, with error measured by MAPE of 0,72%. As for the commodity series that, theoretically, presents lower volatility, the combined use of the filters didn't bring great improvement in error reduction. However, the error was smaller when using wavelets with recurrent neural networks with MAPE of 0,49%. Testing the possible variations on the wavelet filtration to check the possible interference on the results of forecasting, the conclusion reached was that the error is only inferior to forms of primary waves, in this case, to the wavelet of "Haar" and "daubesch 1". Therefore, this paper is a contribution to the area by creating a way of reducing errors in forecasting and, consequently, developing better management of risks in investment positions in the financial market.

Keywords: Forecasting; Filters; *Wavelet*; Kalman Filter; Risk Management.

LISTA DE ILUSTRAÇÕES

Figura 1.1: Desenho do esquema proposto para a pesquisa	24
Figura 2.1: Modelo não linear de um neurônio	45
Figura 2.2: Esquema básico de um filtro	49
Figura 2.3: Modelo de um sistema dinâmico	49
Figura 2.4 : Árvore de decomposição em dois níveis de uma série temporal via decomposição por <i>wavelets</i>	50
Figura 2.5: Processo de filtragem por uma onda senoidal e por uma wavelet	52
Figura 2.6: Processo de análise de wavelets aplicado ao Ibovespa	56
Figura 3.1 : Linha do tempo dos principais contribuintes para análise de estimação de dados	72
Figura 3.2 : Modelo de Kalman e filtro de Kalman- <i>Wavelet</i> para previsão de XIE, ZHANG e YE(2007)	77
Figura 4.1: Fluxograma descritivo das etapas adotadas na pesquisa	85
Figura 4.2 : Fluxograma descritivo para previsão do redes neurais e filtro de wavelets	86
Figura 4.3: Fluxograma descritivo para previsão com filtro de Kalman	87
Figura 4.4 : Fluxograma descritivo para previsão com filtro de Kalman e wavelets	88
Figura 4.5: Fluxograma descritivo para previsão com filtro wavelets e Kalman	88
Figura 5.1: Série temporal nominal do IBOVESPA diário	95
Figura 5.2: Série temporal dos log retornos do IBOVESPA diário	96
Figura 5.3: Histograma dos log retornos diários do IBOVESPA diário e retornos diários contra distribuição normal	97
Figura 5.4: Histograma dos log retornos diários do IBOVESPA com teste de normalidade de Jarque Bera (1987)	98
Figura 5.5: Correlograma dos quadrados dos quadrados dos resíduos do IBOVESPA	101
Figura 5.6: Gráfico do IBOVESPA real e previsto 4 passos a frente com previsão estática para o modelo ARIMA-GARCH	104
Figura 5.7: Gráfico do IBOVESPA real e previsto 4 passos a frente com previsão estática para Redes Neurais Recorrentes	106
Figura 5.8: Gráfico do IBOVESPA real e previsto 4 passos a frente com previsão estática para Redes Neurais Recorrentes com filtro de	

	wavelets
Figura	5.9: Gráfico do IBOVESPA real e previsto 4 passos a frente com previsão estática para filtro de Kalman
Figura	5.10: Gráfico do IBOVESPA real e previsto 4 passos a frente com previsão estática pelo uso de redes neurais recorrentes com filtro de Kalman primeiro e <i>wavelets</i> depois
Figura	5.11: Gráfico do IBOVESPA real e previsto 4 passos a frente com previsão estática pelo uso de redes neurais recorrentes com filtro <i>wavelets</i> primeiro e Kalman depois
Figura :	5.12 Comportamento do preço do barril de petróleo (\$ US\$)
Figura :	5.13: Série temporal dos log retornos do petróleo cru diário
Figura	5.14: Histograma dos log retornos diários da série do petróleo cru diário e retornos diários contra distribuição normal
Figura	5.15: Histograma dos log retornos diários do petróleo cru com teste de normalidade de Jarque Bera (1987)
Figura :	5.16: Correlograma dos quadrados dos quadrados dos resíduos para o petróleo
Figura	5.17: Gráfico do Petróleo real e previsto 4 passos a frente com previsão estática para o modelo ARIMA-GARCH
Figura	5.18: Gráfico do Petróleo real e previsto 4 passos a frente com previsão via redes neurais recorrentes
Figura	5.19: Gráfico do Petróleo real e previsto 4 passos a frente com previsão para Redes Neurais Recorrentes com filtro de wavelets.
Figura	5.20: Gráfico do Petróleo real e previsto 4 passos a frente com previsão estática para filtro de Kalman
Figura	5.21: Gráfico do petróleo real e previsto 4 passos a frente com previsão estática pelo uso de redes neurais recorrentes com filtro de Kalman primeiro e <i>wavelets</i> depois
Figura	5.22: Gráfico do petróleo real e previsto 4 passos a frente com previsão estática pelo uso de redes neurais recorrentes com filtro <i>wavelets</i> primeiro e Kalman depois

LISTA DE TABELAS

Tabela 5.1 Teste ADF para os log retornos do IBOVESPA	99
Tabela 5.2 Valores obtidos pelo teste BDS nas respectivas dimensões para o IBOVESPA	100
Tabela 5.3 Valores obtidos pelo teste de McLeod-Li para o IBOVESPA	101
Tabela 5.4 Teste de Hsieh para o IBOVESPA	102
Tabela 5.5 Parâmetros do modelo AR(1)-GARCH(1,1) para os log retornos do IBOVESPA	103
Tabela 5.6 Estatísticas de acurácia do modelo AR(1)-GARCH(1,1) para o IBOVESPA	105
Tabela 5.7 Estatísticas de acurácia com uso de redes neurais recorrentes para o IBOVESPA	106
Tabela 5.8 Estatísticas de acurácia com uso de redes neurais recorrentes para o IBOVESPA com filtro de wavelets	107
Tabela 5.9 Estatísticas de acurácia com filtro de Kalman para o IBOVESPA	108
Tabela 5.10 Estatísticas de acurácia combinando redes neurais recorrentes com filtro de Kalman primeiro e <i>wavelets</i> depois para o IBOVESPA	109
Tabela 5.11 Estatísticas de acurácia combinando redes neurais recorrentes com filtro de <i>wavelets</i> primeiro e Kalman depois para o IBOVESPA	110
Tabela 5.12 Estatísticas de acurácia com uso de redes neurais recorrentes para o IBOVESPA com filtro de Kalman e filtro de wavelets	112
Tabela 5.13 Estatísticas de acurácia com uso de redes neurais recorrentes para o IBOVESPA com filtro de <i>wavelets</i> e filtro de Kalman.	113
Tabela 5.14 Resumo das estatísticas de previsão para o IBOVESPA	114
Tabela 5.15 Teste ADF para os log retornos do Petróleo	118
Tabela 5.16 Valores obtidos pelo teste BDS nas respectivas dimensões para o Petróleo	119
Tabela 5.17 Valores obtidos pelo teste de McLeod-Li para o Petróleo	119
Tabela 5.18 Teste de Hsieh para o Petróleo	120
Tabela 5.19 Parâmetros do modelo AR(1)-GARCH(1,1) para os log retornos do Petróleo	121
Tabela 5.20 Estatísticas de acurácia do modelo AR(1)-GARCH(1,1) para o	122

Petróleo	
Tabela 5.21 Estatísticas de acurácia com previsão feita por redes neurais recorrentes para o petróleo	123
Tabela 5.22 Estatísticas de acurácia com uso de redes neurais recorrentes para o petróleo com filtro de wavelets	124
Tabela 5.23 Estatísticas de acurácia com filtro de Kalman para o petróleo	125
Tabela 5.24 Estatísticas de acurácia combinando redes neurais recorrentes com filtro de Kalman primeiro e <i>wavelets</i> depois para o petróleo	125
Tabela 5.25 Estatísticas de acurácia combinando redes neurais recorrentes com filtro de <i>wavelets</i> primeiro e Kalman depois para o petróleo	126
Tabela 5.26 Resumo das estatísticas de previsão para o Petróleo	127

LISTA DE QUADROS

(Juadro 4 1	Ferramentas	utilizadas en	r cada etana da	nesa	uisa	80
ľ	Zuaui v 4.1	Terramentas	utilizauas eli	i Caua Etapa ua	i pesq	u18a	07

LISTA DE ABREVIATURAS E SIGLAS

\mathbf{y}_{t}	Série temporal financeira no nível
R_{t}	Retorno da série temporal, pode ser discreto ou contínuo
JВ	Estatística do teste de Jarque Bera
Â	Coeficiente de Assimetria
Ĉ	Coeficiente de Curtose
n	Tamanho total da série temporal financeira
\mathcal{E}_{t}	Termo do erro dos modelos ARIMA-GARCH
α	Coeficientes dos modelos ARIMA-GARCH
RU	Raízes unitárias
BDS	Estatística do teste BDS – Brock, Dechert e Scheinkmar
$\hat{ ho}$	Estatística do tesde de Hsieh
$\mu_{_{ m t}}$	Média condicionada da série temporal
h _t	Variância condicionada da série temporal
β	Coeficientes dos modelos ARIMA-GARCH
\mathbf{W}_{t}	Peso da rede neural
X_{i}	Variável de entrada da rede neural
y_i	Variável de saída da rede neural
$arphi_{ m i}$	Função de ativação da rede neural
F_{t}	Conjunto informacional de dados
FR _t	Notação para Filtro de uma série de retornos
Ay_{ta}	Sub-série filtrada por wavelets da parte aproximação
p,q	Índices do modelo GARCH
$\mathrm{Dy}_{\mathrm{td}}$	Sub-série decomposta por wavelets da parte detalhe
$\psi_{\rm t}$	Forma de wavelets
WT	Transformada de wavelet
A,B	Matrizes Jacobianas do filtro de Kalman
e(j)	Erro de medição da previsão
P	Matriz de covariância do filtro de Kalman

K Ganho de Kalman

WR Série de retornos filtrada por *wavelets*RKF Série de retornos filtrada por Kalman

WRKF Série de retornos filtrada primeiramente por Kalman e posteriormente por

wavelets

e^x Função exponencial

Z_t Série de retornos padronizadaMAPE Erro percentual absoluto médio

 ${f r}_{\!{}_{\!{}_{\!{}_{\!{}}}}}$ Coeficiente de correlação de Pearson TIC Coeficiente de desigualdade de Theil

 \hat{y}_t Série dos valores previstos

SUMÁRIO

RESUM	O	iv
ABSTR	ACT	v
LISTA I	DE ILUSTRAÇÕES	9
LISTA I	DE TABELAS	11
LISTA I	DE QUADROS	13
LISTA I	DE ABREVIATURAS E SIGLAS	14
1. IN	FRODUÇÃO	18
1.1	Problema de Pesquisa	23
1.2	Objetivos	25
1.2	.1 Objetivo Geral	25
1.2	.2 Objetivos Específicos	26
1.3	Metodologia	26
1.4	Estrutura do Trabalho	27
2. RE	VISÃO BIBLIOGRÁFICA	29
2.1	Séries Temporais	29
2.1	.1 Séries temporais financeiras	30
2.2	Análise de séries temporais financeiras	31
2.2.1	Estatísticas descritivas dos log retornos	33
2.2.2	Teste de normalidade	34
2.2.3	Teste de estacionariedade	34
2.2.4	Teste BDS – (Broch-Dechert-Scheinkman)	36
2.2.5	Teste de McLeodi-Li	38
2.2.6	Teste de Hsieh	39
2.3	Modelos para séries temporais financeiras	40
2.3	.1 Modelos ARCH	41
2.3	.2 Modelos GARCH	43
2.3	.3 Modelos de Redes Neurais	45
2.4	Filtros	48

2.4	.1 Filtro de wavelets	50
2.4		
2.4		
2.4	3 3	
2.4	.5 Redes Neurais com filtro de Kalman estendido	66
3. RE	VISÃO DA LITERATURA	69
3.1	Breve histórico sobre previsão de séries temporais	69
3.2	Previsão com filtro de Kalman	73
3.3	Previsão com filtro de Kalman e wavelets	76
3.4	Previsão com filtro de wavelets	77
4. MF	ETODOLOGIA	80
4.1	Natureza da Pesquisa	80
4.2	Operacionalização da pesquisa	81
4.2	2.1 Previsão com filtro de <i>wavelets</i> e redes neurais recorrente	es86
4.2	2 Previsão com filtro de Kalman	86
4.2	2.3 Previsão com filtro de Kalman e de wavelets	87
4.2	.4 Previsão com filtro de wavelets e de Kalman	88
4.2	.5 Ferramentas de análise	89
4.3	Estatísticas de erro e intervalo de confiança para as previsões .	89
4.4	Arquitetura da rede neural recorrente	92
5. RE	SULTADOS EXPERIMENTAIS	94
5.1	Análise da série do Ibovespa	94
5.2	Análise da série da <i>commoditie</i> do barril de petróleo	115
CONCL	LUSÃO	129
REFER	RÊNCIAS BIBLIOGRÁFICAS	133
	DICE A – Formas de wavelets	
	DICE B – Processo de Markov	
	DICE C – Programa para rodar a rede recorrente	

1. INTRODUÇÃO

As atuais crises financeiras advindas da instabilidade dos mercados financeiros deram o cenário para se aumentar à relevância acerca do gerenciamento de risco nas estratégias de investimentos. Destarte, o próprio mercado se tornou mais volátil e a preocupação dos investidores com surpresas desagradáveis exige que as informações sejam mais bem preparadas para lidar com as adversidades do mercado.

Os movimentos do mercado financeiro parecem ficar mais evidentes com relação às turbulências provocadas pelas informações divulgadas diariamente. Aliados a esses comportamentos súbitos do mercado, as empresas que operam junto nesses mercados estão buscando avanços em seus controles internos principalmente na mitigação do risco. A necessidade de projeções financeiras relativas a essa volatilidade dos mercados, principalmente as de curto prazo, são fundamentais para o gerenciamento de suas posições no mercado a vista e futuros.

Um dos principais pilares da gestão de riscos, que trata do uso dos modelos quantitativos, está sendo questionado. As recentes críticas ao uso de modelos matemáticos para mensuração de riscos aumentaram o coro entre os especialistas. Todavia, embora este seja um assunto delicado do ponto de vista técnico, há certo consenso de que se ruim com eles, pior sem eles. Isto porque em toda atividade financeira, como por exemplo, gestão de riscos, apreçamento de contratos derivativos, ou seleção de carteiras, há a necessidade de se prever volatilidade dos ativos em questão, justamente para se verificar a possibilidade de perda futura, seja de curto ou de longo prazo.

Mesmo incapazes de lidar com situações inéditas, os modelos são ainda uma forma de se precaver e ter um horizonte de valores estimados para se posicionar nos mercados. Com isso, a busca de modelos e sistemas integrados de previsão futura estão sendo buscados incessantemente e já em algumas décadas

Segundo Sabino e Bressan (2009, p. 2), a escolha de um modelo está sujeita ao objetivo do usuário e à capacidade de descrição do processo de volatilidade característico do mercado. Ainda segundo os autores, ao se avaliar o poder preditivo, deve-se levar em conta as vantagens de modelos mais complexos em relação aos custos de implementação dos mesmos para o tratamento de séries temporais financeiras.

O estudo sobre o comportamento das séries temporais financeiras com objetivo de previsão futura de preços e retornos é foco de discussão e embates, mediante distintas abordagens. Especialmente ao se tratar dos movimentos implícitos dos retornos de um ativo financeiro, analisando seu comportamento, as teorias sobre os modelos e ajustes a esse comportamento são divergentes na busca da compreensão destes fatos empíricos.

Bachelier (1900, p.23) já defendia a idéia de que a série dos retornos de um ativo financeiro segue um *random walk*, pelo fato de ser dependente da ocorrência de múltiplas variáveis tipicamente imprevisíveis. Suas idéias viriam posteriormente a ter desenvolvimentos múltiplos que deram origem a diversos modelos como o CAPM e à análise da gestão de portfólio de Markowitz.

Já Poincaré (1952) ponderava que, mesmo os sistemas relativamente simples de modelar, a previsão para o longo prazo era impossível devido à existência de um efeito que denominou *feedback*, que hoje entende-se por sensibilidade as condições iniciais que se conhece como característica da chamada "Teoria do Caos".

Analogamente, Fama (1970, p. 391) torna o assunto mais conciso ao aliar essa discussão da sensibilidade às condições iniciais com a teoria de avaliação de ativos financeiros, dizendo que tais alterações nos valores iniciais podem causar alteração no preço final do ativo, porém, a consequência imediata seria de que apresentariam um comportamento flutuante de acordo com um comportamento aleatório. O autor defende a hipótese da eficiência de mercado, afirmando que as decisões são tomadas sob a premissa de que todos os investidores possuem as mesmas informações e estão ao mesmo tempo no mercado. Dentro dessa teoria, os preços futuros não poderiam ser previstos nem modelados.

A partir dessas discussões, duas correntes distintas de análise surgiram: uma que defende que os mercados são eficientes, sendo os processos estocásticos, e outra que

advoga que os mercados são ineficientes, caracterizados por uma dependência nãolinear que deriva do fato dos mercados obedecerem a uma lei de comportamento caracterizável por um sistema caótico e, portanto, previsível a curto, mas não em longo prazo.

Ainda nesse sentido, Mandelbrot e Hudson (2004, p. 33) afirmam que, se for possível identificar padrões de comportamento em uma série de dados talvez seja possível haver previsibilidade. Ao atestar esse fato, mostrou que várias séries temporais financeiras possuem uma propriedade específica denominada "memória".

Na busca continua pelos estudos de memória em séries temporais financeiras, Mandelbrot e Hudson (2004, p. 78), em continuidade aos estudos de Hurst (1951), encontraram presença de memória no comportamento das cheias do rio Nilo e deixa a idéia de que seja possível aplicar as teorias das séries temporais não financeiras para as financeiras.

Corroborando essa teoria, Lo e Mackinlay (1999, p. 1301) encontrou evidências empíricas da não existência do comportamento de *random walk* no mercado norte americano. Os testes estatísticos foram feitos para o período de 1960 a 1980.

Segundo Cajueiro et al (2006, p. 4), a verificação do comportamento passado dos dados pode ser medida do ponto de vista da modelagem e tratamento da série temporal, visando encontrar os fatores que interagem entre si. Quando identificado tais fatores, o seu comportamento é processado separadamente para geração do modelo que rege seu processo no tempo. Mas como medir isso?

O grande dilema da análise de séries temporais é justamente esse: como medir determinados processos? Outra forma de se conseguir visualizar esse comportamento é analisar o próprio movimento em si e aplicar testes estatísticos para se identificar quais processos poderiam se ajustar ao movimento observado.

Todavia, para Costa e Vasconcelos (2003, p. 6), todo o conhecimento que se tem sobre um ativo está refletido em seu preço atual, sendo o seu passado de pouca ajuda para prever seus valores futuros. Essa teoria vai ao encontro da maioria das técnicas utilizadas hoje por alguns operadores do mercado financeiro. Técnicas como análise gráfica, que basicamente utiliza-se das cotações históricas dos ativos financeiros para

fazer suas predições, são amplamente divulgadas, publicadas em livros e utilizadas por profissionais do mercado e investidores.

Os trabalhos de Mantegna e Stanley (1999, p. 3) e Bouchaud e Potters (2000) mostraram que existem evidências de que o mercado não é tão eficiente quanto se imagina ser. Duas hipóteses justificaram esta afirmação pelos autores: a de que os dados podem violar as hipóteses de independência ou as de eficiência de mercado. E é nessa lacuna que se verifica o uso das técnicas grafistas para análise dos preços das ações no mercado onde se procura verificar o comportamento futuro nas séries temporais de ativos financeiros..

Dessa forma, dadas algumas observações passadas do comportamento de uma série de tempo, podem-se fazer previsões sobre o seu comportamento futuro e verificar quão precisa essas previsões podem ser. Tais processos já foram testados nos mais variados modelos e com resultados e metodologias variadas de previsão. Resultados satisfatórios foram encontrados por vários autores como pode ser encontrado em Lima (2004).

O que se identificou é que, em ambas as modelagens, os processos são feitos usando um período longo de dados e criando equações que são construídas com intuito de modelar os mecanismos responsáveis pela geração das séries temporais. O mesmo ocorre nos processos que são estudados via redes neurais que são treinadas para abstrair os processos gerados embutidos nas séries de tempo.

No entanto, em ambos os modelos são usadas séries temporais de longos períodos. E uma mesma série de tempo pode apresentar comportamentos distintos em diferentes momentos pelos mais variados motivos como crises, processos de correção de valores por ajustes nos preços e quebras estruturais.

Existem na literatura diversos métodos para se fazer previsão de séries temporais, desde os mais simples e de fácil compreensão até os mais complexos que envolvem diferentes parâmetros como os modelos ARIMA e os modelos da família GARCH. E o fato de se utilizar métodos estatísticos mais complexos não significa necessariamente uma melhora nos resultados da previsão.

O uso da decomposição de séries temporais via *wavelets*, visando à análise de séries temporais, surgiu como alternativa para redução de ruídos nas séries temporais.

Combinação dessa metodologia com os modelos tradicionais de previsão foram empregados por Granger (1992, p. 3), Tak (1995, p. 43), Ariño (1995), Ukil e Zivanovic (2001, p. 103), Ma, Wong e Sankar (2004, p. 5824) e Aminghafari (2007, p. 715).

No Brasil, destacam-se os trabalhos de previsão de séries temporais de Chiann (1997, p. 32), Homsy, Portugal e Araújo (2000, p. 10), Zandonade e Morettin (2003, p. 205), Lima (2004, p. 133), e Rocha (2008, p. 120).

Wavelets são funções que consistem em fracionar a série temporal original em duas subséries, uma relativa às altas frequências e a outra às baixas frequências com objetivo de reduzir os efeitos do ruído nas previsões (GENÇAY; SELÇUK; WHITCHER, 2002). O uso deste processo de filtragem da série trouxe melhora significativa nos modelos de previsão conforme pode ser visto nos trabalhos citados anteriormente.

Em contraposição à abordagem estática dos modelos ARIMA, surgiram os Modelos Lineares Dinâmicos (MLD), introduzidos por Kalman (1960), que são formulados com a característica de incorporar mudanças nos parâmetros, à medida que ocorrem evoluções na série temporal. O aumento no número de observações da série é interpretado, então, como informação adicional ao conjunto de informações atuais, fazendo com que os parâmetros apresentem uma evolução dinâmica, impedindo qualquer quantificação estática das relações subjacentes ao comportamento global da série.

A operacionalização do modelo MLD é obtida adotando-se o modelo de espaço de estados e utilizando-se o filtro de Kalman para a atualização sequencial dos componentes não observáveis. A representação em espaço de estados é feita por meio de um sistema de duas equações dinâmicas que descrevem a maneira pela qual as observações são geradas em função do vetor de estados e a evolução dinâmica desse vetor. O filtro de Kalman consiste basicamente de um algoritmo que fornece estimativas atualizadas do vetor de estados a cada instante de tempo.

Recentemente, novos trabalhos estão fazendo uso da metodologia do filtro de Kalman para realização de previsões como Aiube (2005, p. 108) e Corsini e Ribeiro (2008, p. 11) e com uso combinado de Kalman e *wavelets* como em Postalcioglu, Erikan e Bolat (2005, p. 951).

Fato é que, ambas as teorias e modelagens partem de suas premissas, adotam seus métodos de análise e auferem seus resultados. Isto reporta a dimensão de que ainda não há uma absoluta certeza em favor de uma ou de outra teoria. Tão pouco existe ainda uma forma de combinar esses modelos e filtros em um único esquema de previsão para tomada de decisão.

É nessa perspectiva que a modelagem das variações dos ativos financeiros possa colaborar com os usuários da informação advinda do mercado financeiro, e também da contabilidade e controladoria, possa não só obter a representação da evolução do comportamento dos dados financeiros, como também obter previsões confiáveis de valores para o futuro.

No mercado financeiro, têm-se hoje diversas modalidades de operações que necessitam da informação futura para tomada de decisão. As opções são um claro exemplo disso. Com vencimento em toda terceira segunda feira de cada mês, o investidor, que realiza uma operação de compra ou venda, precisa estar informado do comportamento previsto até o vencimento da opção para assumir uma posição comprada ou vendida no mercado. Dessa forma, a previsão é consistente com o comportamento que o preço do ativo produzirá uma medida útil do risco que se possa assumir no mercado.

As diversas aplicações que podem ser feitas com a junção das técnicas de filtragem com os modelos de previsão fazem com que esta área de estudo seja uma das mais dinâmicas no estudo das finanças.

É nesse contexto que se formula o problema de pesquisa que é objeto desta discussão.

1.1 Problema de Pesquisa

Uma previsão em finanças de um retorno de um ativo ou de um preço futuro é uma tentativa de se prognosticar o valor futuro por meio do conhecimento dos fatos memorizados no passado. Entre as diversas técnicas que desempenharam esse papel de prever um valor para o futuro, existem diversos modelos já verificados na literatura e, mais recentemente, os filtros de separação de séries surgiram como uma alternativa

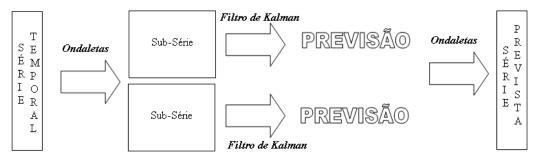
complementar às atuais técnicas de previsão, só que ainda pouco explorado em seu uso combinado com os modelos.

Dessa forma, justifica-se a continuidade dos estudos relativos ao tema. Além disso, por se tratar de um assunto relativamente recente, a exploração tende a trazer novos resultados.

Seguindo esta linha de estudo, surge, portanto, a seguinte questão de pesquisa: a combinação de filtros em séries temporais financeiras melhora a sua capacidade preditiva? Em outras palavras, se primeiro fizer a decomposição via *wavelets* e depois aplicar o filtro de Kalman melhora ou não a previsão futura da série?

Pode-se representar a pesquisa pela figura a seguir:

1ª Hipótese



2ª Hipótese

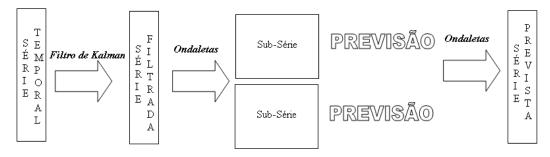


Figura 1.1: Desenho do esquema proposto para a pesquisa

A figura anterior mostra o esquema que foi empregado na pesquisa. Na primeira hipótese, dada uma série temporal financeira com objetivo de previsão de futura, será aplicado primeiramente um filtro de *wavelets* (ondaletas) para separação da série em alta e baixa frequência. Em seguida, aplica-se o filtro de Kalman para a etapa de

alisamento dos dados e, em seguida, realiza-se as previsões usando redes neurais recorrentes. Feitas as previsões, aplica-se a transformada inversa para obter a série no nível.

Na segunda hipótese, realiza-se a previsão ao contrário. Primeiro alisa-se a série pela aplicação do filtro de Kalman e, em seguida, a decomposição pelas *wavelets* (ondaletas), para posteriormente aplicar as redes neurais recorrentes para previsão. Aplicando-se a transformada inversa de *wavelets* obtém-se a série no nível novamente.

1.2 Objetivos

1.2.1 Objetivo Geral

Diante das diversas técnicas de previsão de séries temporais e uso de filtros para suavização de séries temporais, o objetivo geral deste estudo é realizar uma análise comparativa do uso combinado de filtros de *wavelets* e *kalman* juntamente com modelos de previsão para séries temporais financeiras para verificar qual produz a melhor previsão futura.

Dessa forma este trabalho tem por foco fazer uma análise do efeito da aplicação dupla de um filtro de espaço e estado aliado a um critério de decomposição de alta e baixa freqüência em ordens diferentes de ocorrência além de comparar a qualidade das previsões feitas em separado por cada hipótese.

Mais especificamente, a realização da pesquisa se deu em duas etapas. Primeiro tomou-se uma série temporal e aplicou-se sobre ela o filtro de Kalman e posteriormente decompor a série em alta e baixa frequência. Nessas sub-séries decompostas, efetuou-se as previsões e com a aplicação da ondaleta inversa reconstrói-se a série original. Numa segunda etapa, fez-se o processo reverso, ou seja, primeiro a decomposição da série em alta e baixa frequência, e sobre elas aplica-se o filtro de Kalman, e em seguida realiza-se a previsão.

Com efeito comparativo, serão realizadas ainda as previsões somente com filtro de Kalman, com redes neuras recorrentes e modelos ARIMA-GARCH para testar também o uso dos filtros como redutor de erro das previsões.

Definem-se ainda alguns objetivos específicos decorrentes dos processos preparatórios para obtenção do resultado do problema de pesquisa.

1.2.2 Objetivos Específicos

Como objetivos específicos têm-se:

- verificar o uso desta combinação de filtros para alisamento da série e
 consequente redução do ruído e melhor na qualidade da previsão, tem o
 mesmo efeito para séries de maior volatilidade como índices de bolsa e
 para séries menos voláteis como commodities;
- variar as wavelets para realização das decomposições para atestar seu efeito na qualidade final das previsões.

1.3 Metodologia

Para alcançar os objetivos desta pesquisa, alinhavou-se uma estrutura metodológica composta por três níveis. O primeiro nível foi uma revisão bibliográfica teórica e dos artigos de maior expressão na área com intuito de certificar o suporte acadêmico para a tese e atestar sua viabilidade.

Na descrição dessa revisão teórica chegou-se a discussão de que as séries temporais trazem em seu bojo uma série de efeitos incompatíveis com o movimento correto dos preços dos ativos financeiros. Tais movimentos necessitariam de separação do comportamento "normal" da série para a retirada destes termos indesejáveis presentes nas séries temporais financeiras que poderia ser feita por meio dos filtros de separação.

E foi com esta inspiração e motivação que se desenvolveu esta tese a partir de uma metodologia quantitativa de base econométrica para atingir seus objetivos. No segundo nível metodológico, ou seja, parte operacional da pesquisa, usaram-se as redes neurais recorrentes para previsão com uso dos filtros de *wavelets* e Kalman. As análises contaram com ferramentas e softwares específicos da área como Matlab e Eviews.

Vale lembrar que o duplo processo de filtragem dos dados pelos filtros de wavelets e Kalman, conforme descrito pela hipóteses anteriores, tiveram a finalidade de redução de ruídos para posterior execução de previsões com as redes neurais. Isso

porque, o objetivo das *wavelets* não é previsão por si mesma, e sim a separação de ruídos.

Já o filtro de Kalman possui duplo uso, um de capacidade preditiva, que viabiliza seu uso como modelo de previsão, e outro como ferramenta de alisamento de dados. Logo, usou-se também o filtro de Kalman para previsão em cadeia com as decomposições de *wavelets* e da mesma forma, usou-se a outra finalidade do filtro de *wavelets* que é suavização de dados para obter uma nova série em que foi realizada previsão com as redes neurais e com o próprio filtro de Kalman na vertente de previsão.

Dessa forma, são realizados previsões em seis estágios:

- I. redes neurais recorrentes pura e simplesmente;
- II. modelos ARIMA-GARCH;
- III. separação de wavelets e previsão com redes neurais recorrentes;
- IV. previsão com filtro de Kalman simples;
- v. uso do filtro de Kalman para filtragem e consequente previsão com as redes neurais recorrentes, e
- VI. uso do filtro *wavelets* e em seguida previsão com o filtro de Kalman e previsão com redes neurais recorrentes.

As etapas de I à IV já se encontram revistas na literatura e confirmadas por pesquisadores. Realizaram-se essas etapas novamente apenas para compor uma biblioteca de resultados com intenção de comparabilidade de variáveis de resultado. O diferencial da pesquisa encontra-se nos estágios V e VI conforme descritos anteriormente nas duas hipóteses junto ao objetivo da pesquisa.

No terceiro e último nível da pesquisa, checaram-se as estatísticas de erro com uso de planilhas eletrônicas para cada um dos seis estágios, para construção dos gráficos de previsão, intervalos de previsão e das estatísticas de acurácia dos modelos.

A seguir, descreve-se a estrutura elaborada para esta pesquisa.

1.4 Estrutura do Trabalho

O estudo está estruturado da seguinte forma: após essa introdução, no próximo capítulo dois é realizada a revisão bibliográfica sobre tratamento e modelagem de séries temporais financeiras para efeitos de previsão futura com objetivo de definição de conceitos que são utilizados e discutidos. O capítulo três contém a revisão da literatura atual, que se utilizou de alguma forma das teorias impostas no capítulo dois para tratamento de séries e uso combinado de filtros aplicáveis a séries temporais. No capítulo quatro, descreve-se a metodologia completa utilizada para a realização da pesquisa e os métodos utilizados para tratar os dados. No capítulo cinco, apresenta-se o desenvolvimento e os resultados obtidos. A conclusão é feita após as análises descritivas dos resultados e emprego das técnicas consideradas. No final, encontra-se ainda três apêndices onde se destacam as formas de *wavelets* discutidas na pesquisa, o processo de Markov usado na decomposição de uma série temporal em espaço e estado e o programa utilizado para redes neurais recorrente, respectivamente.

2. REVISÃO BIBLIOGRÁFICA

Nesta seção será apresentada uma síntese dos conceitos, preceitos e aspectos teóricos e práticos envolvidos na análise e previsão de séries temporais e uso combinado dos filtros aplicados aos modelos econométricos e de redes neurais voltados para previsão.

O capítulo é baseado em sua essência na bibliografia disponível nacional e internacional sobre séries temporais, modelos econométricos e redes neurais.

2.1 Séries Temporais

Uma série temporal é qualquer sequência de dados estocásticos ordenados obtidos de intervalos regulares no tempo por um período específico de observações. Sua representação matemática é dada por $\left\{y_t\right\}_{t=1}^n = \left\{y_1, y_2, ..., y_t, ..., y_n\right\}$.

Também chamada às vezes de série histórica, essa sequência de dados pode ser obtida de observações periódicas de tempo como diária, mensal, semestral, anual, ou até mesmo em períodos submúltiplos como cotações de minuto a minuto de uma ação ou índice de bolsa de valores. Este tipo de dados é chamado de 'alta frequência'.

Sua representação gráfica é comumente chamada de trajetória. Essa designação nada mais é do que a curva representativa do movimento obtido pelo processo estocástico. Morettin e Toloi (2004, p. 2) definem o que se chama de série temporal, é uma parte dessa trajetória, dentre muitas que poderiam ter sido observadas de um processo estocástico¹.

Já Enders (2004, p. 3) usa o termo 'sucessão cronológica' para uma série temporal de observações indexadas, devidamente equiespaçadas no tempo, cujo número é pelo menos 50 itens.

.

¹ Processos estocásticos são processos cuja evolução no tempo é regida por leis probabilísticas.

O conjunto de dados de uma série temporais pode ainda ser considerado discreto, quando as observações são obtidas de intervalos inteiros de posições, ou contínuo se geradas por intervalos racionais de tempo (Box, Jenkins e Reinsel, 1994, p. 12).

De acordo com a natureza dos dados da série temporal é também renomeada para atuar na respectiva área. Por exemplo, se forem dados médicos a série é chamada de 'série temporal clínica', se for de dados relativos à biologia é uma 'série temporal biológica e se for de dados financeiros tem-se a 'série temporal financeira'.

Como o foco deste trabalho são as séries temporais financeiras, as demais séries não serão tratadas aqui. A seguir, uma breve descrição das séries temporais obtidas junto a dados do mercado financeiro.

2.1.1 Séries temporais financeiras

As séries temporais financeiras se diferem das demais séries temporais por apresentar características próprias de seus elementos. Essas características, segundo Enders (2004, p. 10), é que essas séries são não serialmente correlacionadas mas dependentes.

Além disso, segundo o autor, tais séries apresentam em seus log retornos (esta expressão será definida no capítulo 2) a presença de *clusters* de volatilidade. Isso faz com que na análise dessas séries a qual se deseja modelar o fenômeno que as gera para posteriormente se fazer previsões, esses grupos de volatilidade podem ser definidos de diversas maneiras, mas não são claramente observáveis.

Nessa análise, a medida da variância de um log retorno para um determinado período de tempo, passa a depender dos log retornos passados além de outros elementos até então desconhecidos, de modo que sua variância condicional não coincide com a variância total da série (chamada variância incondicional).

Essa medida de variância é fundamental em finanças por trazer em seu bojo a medida do risco associado às variações dessa série temporal ao longo do tempo, usada comumente em diversos modelos da teoria de finanças como o modelo de precificação de ativos de capital, modelo de apreçamento de opções e outros.

Os modelos usados para esse fim usam frequentemente a medida de risco dada pelo desvio padrão dos log retornos de séries temporais financeiras. Usam-se os log retornos, pois segundo Morettin (2008, p.7), eles apresentam propriedades estatísticas como estacionariedade e ergodicidade (que se consegue medir e prever valores futuros por meio da estimação estatística da interpretação dos dados passados) e são também livres de escala.

Dessa forma, um dos objetivos das séries temporais financeiras é modelar esses log retornos. Diversas classes de modelos e técnicas poderão ser utilizadas para esse fim. A seguir, descrevem-se esses modelos e técnicas para análise destas séries.

2.2 Análise de séries temporais financeiras

A análise de séries temporais financeiras se inicia por meio do cálculo dos retornos. Um ativo financeiro é cotado no mercado pelo seu valor em moeda constante para o um dado período de tempo. O seu preço em um instante t qualquer é representado por P_t , e no instante anterior t-t por t-t-t a t, como sendo a medida relativa dada pelas seguintes expressões:

Retorno Discreto
$$R_t = \frac{P_t}{P_{t-1}} - 1$$
 (1)

Retorno Contínuo (log retorno)
$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$$
 (2)

Segundo Tsay (2005, p. 22), os preços dos ativos financeiros são considerados como variáveis aleatórias que admitem função de distribuição própria, devido à possibilidade teórica dos preços assumirem valores infinitos e não poderem apresentar valores negativos.

A diferença básica entre o uso do retorno discreto e do retorno contínuo está na interpretação dada pelo desvio padrão. O desvio padrão do retorno discreto é chamado de risco e o do retorno contínuo, ou log retorno, é chamada volatilidade. Para pequenos intervalos de tempo *t-1* e *t*, os valores dos retornos serão em geral valores próximos.

A expressão (2) é comumente mais utilizada por apresentar propriedades estatísticas como normalidade através do teorema do limite central. (Morettin, 2008, p. 8).

No caso de ativos financeiros de bolsa de valores como ações, os preços desses ativos devem estar ajustados para dividendos, bonificações e juros sobre o capital próprio.

Na inspeção visual da trajetória dos retornos de uma série temporal financeira, poderá ser verificada a presença de *clusters* de volatilidade, que são fortes oscilações constantes presente na série temporal financeira.

Para Sabino e Bressan (2009, p. 3), a modelagem da variabilidade dos ativos financeiros permite não só a representação da evolução do comportamento dos dados, mas também a previsão dos valores futuros que produzirá uma medida útil de risco que poderá ser assumido pelos investidores no mercado financeiro.

Isto ocorre devido às incertezas presentes no mercado, tanto nacional quanto internacional, seja por meio de crises econômicas, fortes oscilações de moeda e outros fatores.

Sabino e Bressan (2009, p. 3) citam ainda que, dentro da densa literatura sobre modelagem da previsão de volatilidade que ocorreu durante as últimas décadas contadas no artigo destes autores, foi sobretudo a partir de 1996, com o acordo da Basiléia que tais modelos estão cada vez mais presentes na gestão de riscos das corporações. Tal acordo criou os princípios que regem o gerenciamento de riscos financeiros. Estes fatos, segundo os autores, tornaram compulsório o exercício da previsão da volatilidade para as instituições financeiras.

Andersen et al (2006, p. 779) sugerem que a volatilidade tem sido uma área de pesquisa mais ativa e bem sucedida em econometria de séries temporais e previsão nas últimas décadas.

A partir de então, aplica-se testes estatísticos para análise do comportamento da série temporal. Seguindo o organograma desenvolvido por Lima (2004, p. 33), a seguir faz-se um resumo dos principais testes a serem aplicados.

2.2.1 Estatísticas descritivas dos log retornos

A primeira análise que se procede é a análise descritiva dos seus valores. Em geral, para a série dos log retornos, possuem média próxima de zero e variância constante. O desvio padrão incondicional da série dos log retornos representa as oscilações médias em torno da média. Quanto maior esse indicador, maior será o nível de risco da série.

Os valores de curtose e assimetria são importantes para o teste de normalidade que também identifica maior ou menor dispersão dos valores em relação à sua concentração em torno de zero.

As oscilações encontradas com os movimentos de "sobe" e "desce" no gráfico dos retornos são as indicações da variabilidade dos retornos em função do tempo, ou seja, a chamada volatilidade, e também poderá ser observado valores dos retornos atípicos, chamados de *outliers* se comparado com os demais retornos.

Os retornos raramente costumam apresentar tendências ou mesmo sazonalidades, salvo exceção eventual para os dados de alta frequência (intra-diários). Morettin (2008, p. 15) e Enders (2004, p. 12) ressaltam ainda que os retornos são em geral não auto-correlacionados. Já a série dos quadrados dos retornos são auto-correlacionados. Esses fatores observados em séries financeiras são chamados pelos autores de 'fatos estilizados'. Outros poderão ser observados visualmente por meio do gráfico da sua distribuição, como é o caso do histograma.

O histograma de uma série de retornos financeiros em geral possui a parte central mais alta do que a curva normal e com valores afastados da tendência central, aparentando comportamento leptocúrticos com as caudas mais pesadas que a normal.

Outro gráfico importante na análise é o gráfico *Quantis*× *Quantis*, onde um dos eixos mostra os quantis da normal padrão e no outro os quantis dos dados. Sua visualização mostra que se os dados serão aproximadamente o comportamento normal, os pontos estariam sobre uma reta diagonal.

O estudo da normalidade dos log retornos é um assunto extremamente importante por facilitar a modelagem por meio de modelos lineares. Dessa forma, o passo seguinte é o teste de normalidade.

2.2.2 Teste de normalidade

O teste de normalidade dos log retornos de Jarque Bera (1987) averigua se a série segue o comportamento de uma distribuição gaussiana (normal). As hipóteses para este teste são definidas como:

 H_0 : a série segue uma distribuição normal

 H_1 : a série não segue uma distribuição normal

A estatística de teste envolve o cálculo da assimetria e curtose por meio da expressão:

Jarque Bera (JB):
$$JB = n \left[\frac{\hat{A}^2}{6} + \frac{(\hat{C} - 3)^2}{24} \right]$$
 (3)

sendo que n é o tamanho da série temporal dos log retornos, \hat{A} é o coeficiente de assimetria dado por $\hat{A} = \frac{1}{n\hat{\sigma}^3} \sum_{t=1}^n (y_t - \overline{y})^3$ e \hat{C} representa a curtose dado por $\hat{C} = \frac{1}{n\hat{\sigma}^4} \sum_{t=1}^n (y_t - \overline{y})^4$. A comparação é com uma distribuição qui-quadrado com dois graus de liberdade ($\chi^2(2)$) e com o nível de significância adotado na análise.

A lineariedade e a estacionariedade são dois fatos relevantes para a análise de séries temporais por serem os primeiros indicadores para a escolha do melhor modelo estatístico. O teste de estacionariedade é descrito a seguir.

2.2.3 Teste de estacionariedade

O teste de estacionariedade busca identificar se o processo gerador dos dados da série temporal é ou não estacionário. A idéia de estacionariedade se dá por saber se os dados oscilam ao redor de uma média constante com sua variância também constante.

Identificou-se diversos tipos de testes na literatura. Os principais são o teste de Dickey e Fuller Aumentado (Dickey e Fuller, 1979), Phillips e Peron (1988) e Kwiatkowski-Phillips-Schmidt-Shin – KPSS (Kwiatkowski et al., 1992).

A utilização do teste de Dickey Fuller (DF) é válida quando o processo gerador da série temporal é um processo auto-regressivo de ordem um e seu termo de erro aleatório é um ruído branco. Se o for, considerando um modelo AR(1) da forma dada pela expressão abaixo:

$$R_t = \alpha R_{t-1} + \varepsilon_t$$
, com $\varepsilon_t \sim RBN(0, \sigma^2)$ (4)

Dessa forma, pode-se reescrever a equação acima da seguinte forma:

$$\Delta R_{t} = \alpha^* R_{t-1} + \varepsilon_{t} \tag{5}$$

com $\alpha^* = \alpha - 1$. Utilizando o método dos mínimos quadrados ordinários, tem-se a seguinte hipótese nula: $H_0: \alpha^* = 0$ contra a hipótese alternativa $H_1: \alpha^* < 0$.

O termo teste de raiz unitária (RU) vem da representação dada pela equação (5), uma vez que uma das raízes do polinômio autoregressivo é igual a um, todas as demais estarão fora do circulo de raio unitário para o processo ferador da série temporal.

Se o processo gerador da série temporal for um processo auto-regressivo de ordem superior, aplica-se então o teste de Dickey-Fuller Aumentado (ADF), que consiste em estimar a equação a seguir:

$$\Delta R_{t} = \alpha R_{t-1} + \sum_{i=1}^{p} \varphi_{i} R_{t-i} + \varepsilon_{t}$$
(6)

Optou-se por aplicar o teste de Dickey Fuller Aumentado (ADF) por ser o mais indicado e utilizado na literatura. Como visto, o teste verifica se a série já se encontra estacionária no nível ou se é necessário fazer diferenças entre ela para se tornar estacionária. As hipóteses do teste são:

 H_0 : 1RU: existe raiz unitária, isto é, a série é não estacionária

 $H_1: 0RU$: não existe raiz unitária, isto é, série estacionária

A estatística do teste ADF tem a mesma distribuição assintótica que a estatística de DF, de modo que podem ser usados os mesmos valores críticos.

Após verificado se existe ou não estacionariedade nos dados, é importante testar a presença de não-linearidade que será feito a seguir.

2.2.4 Teste BDS – (Broch-Dechert-Scheinkman)

Para averiguar a relação de dependência temporal, usa-se o teste de Broch, Dechert e Scheinkman (1986). Essa dependência significa investigar os desvios da série como não estacionariedade, não lineariedade e caos determinístico. Este teste verifica se a série dos log-retornos de uma série temporal são independentes e identicamente distribuídos, chamado i.i.d., isto é, se todos os termos tem a mesma distribuição de probabilidade. O objetivo deste teste é identificar lineariedade ou não lineariedade determinista ou estocástica nos dados dos log retornos de uma série temporal financeira.

Para isso, aplica-se o teste desenvolvido por Brock, Dechert e Scheinkman (1986) que é melhor ajustado a séries temporais financeiras cujas hipóteses são:

 H_0 : a série é estocasticamente independente

 H_1 : a série é estocasticamente dependente

Segundo os autores, inicialmente organiza-se os dados em vetores n-dimensionais como $\{R_t\}_{t=1}^n=\{R_{t-n+1},\ldots,R_t\}$, sendo que o parâmetro n é chamado dimensão de imersão. Por questão de simplificação de notação, neste tópico, T será o tamanho total da amostra. A estatística do teste BDS é baseada na integral de correlação entre dois vetores de sucessões cronológicas, definidas por meio das seguintes quantidades:

$$C_n(k) \equiv \lim_{T \to \infty} C_{n,T}(k) \tag{7}$$

$$C_{n,T}(k) = \frac{\sum_{s=1}^{T} \sum_{t=s}^{T} k_{st}}{T(T-1)/2}$$
(8)

$$k_{st} = \begin{cases} 1, \text{ se max } \left| R_{s-i} - R_{t-1} \right| < k \\ 0, \text{ caso contrário} \end{cases}$$
 (9)

sendo que é o tamanho da sucessão cronológica e então:

$$C_n(k) = C_1^n(k) \tag{10}$$

Para entender melhor este resultado, a razão:

$$\begin{split} &\frac{C_{n+1}(k)}{C_n(k)} = \Pr\bigg(\max \left|X_{s-i} - X_{t-1}\right| < k \left|\max \left|R_{s-i} - R_{t-1}\right| < k\right.\bigg) = \Pr\bigg(\left|R_s - R_t\right| < k \left|\max \left|R_{s-i} - R_{t-1}\right| < k\right.\bigg) \\ &\text{pode ser interpretada como uma probabilidade condicionada. Fixando} \\ &C_{n+1}(k)/C_n(k) = C_1(k) \text{ para todo } n \text{ positivo, obtém-se a equação (10).} \end{split}$$

Os autores propuseram a seguinte estatística para o teste BDS:

$$BDS = J_{n,T}(k) = \sqrt{T} \frac{C_{n,T}(k) - C_{1,T}^{n}(k)}{\hat{\sigma}_{n,T}(k)}$$
(11)

sendo que:

$$\hat{\sigma}_{n,T}(k) = \left(4\left[w^n + 2\sum_{j=1}^{n-1} w^{n-j}c^{2j} + (n-1)^2c^{2n} - n^2wc^{2n-2}\right]\right)^{\frac{1}{2}}$$
(12)

em que

$$c = C_{1,n}(k)$$

$$w = \frac{6}{n(n-1)(n-2)} \sum_{t=1}^{n} \sum_{s=t+1}^{n} \sum_{r=s+1}^{n} h_k(R_t, R_s, R_r)$$
(13)

$$h_k(i,j,k) = \frac{1}{3} \left[k_{i,j} k_{j,w} + k_{i,w} k_{w,j} + k_{j,i} k_{i,w} \right]$$
 (14)

e ainda $C_{n,T}(k)$ e $C_{1,T}(k)$ são as integrais de correlação amostral definidas em (10) e $\hat{\sigma}_{n,T}(k)$ é o estimador do desvio-padrão assintótico de $C_{n,T}(k) - C_{1,T}^{n}(k)$.

O teste BDS deve ser executado da segunda até a sexta dimensão, portanto, uma determinada série temporal financeira de log-retornos rejeitará a hipótese nula se, e somente se, pelo menos um dos cinco níveis descritivos do teste (p-valor) forem menores do que o nível de significância adotado no teste com uma distribuição normal padrão.

No caso da não aceitação da hipótese nula, ou seja, verificado a existência de dependência nos dados dos log retornos, caberá ainda investigar por meio de testes mais

38

específicos, quais seriam as origens dessa rejeição. A resposta a esta inspeção está na possibilidade da presença de não lineariedade na média ou na variância. Esses testes são descritos a seguir.

2.2.5 Teste de McLeodi-Li

Em sequência, a análise de séries temporais é feita para o fato dos dados da série temporal não serem identicamente distribuídos refere-se que eles possuem intervalos de tempos com diferentes distribuições de probabilidade para os log retornos. Essa característica mostra a presença de não lineariedade dos log retornos.

Para verificar a presença de não linearidade, pode-se aplicar o teste de McLeod-Li (1983). É a partir deste teste que ficará evidente que a rejeição da hipótese da distribuição independente e idêntica estará mais fortemente ligada à existência de não lineariedade nas séries dos retornos, em consonância com os chamados fatos estilizados observados em séries de retornos financeiros.

O teste é baseado na autocorrelação amostral dos quadrados dos resíduos de uma sucessão cronológica estacionária gaussiana. As hipóteses do teste são:

 H_0 : a série temporal é linear

 H_1 : a série temporal é não-linear

O teste é feito após ajustar um modelo autoregressivo, AR(1) no caso, que remova a dependência linear da série dos log retornos, e faz-se a análise dos resíduos dessa regressão.

A aceitação da hipótese nula também não exclui a possibilidade da dependência temporal não linear ser de ordem superior a quadrática. Faz-se também o teste para a série dos quadrados dos retornos. Se a série dos log retornos ao quadrado apresentar forte autocorrelação, haverá indícios de presença de um comportamento heterocedástico.

Um fato estilizado e bem conhecido das séries temporais de retornos financeiras é o comportamento heterocedástico, ou seja, se a variância dos dados não é constante para diferentes intervalos de tempo. Resumidamente, este efeito é caracterizado pelas

fortes oscilações do mercado que ocorrem quando este se comporta de forma inesperada e inconstante.

Uma vez estabelecido que exista algum tipo de não-linearidade em uma série temporal, é preciso identificar o tipo de não-linearidade existente, se na média ou na variância condicionadas. Tal teste será definido a seguir.

2.2.6 Teste de Hsieh

Uma vez estabelecido que fosse verificada a existência de não-linearidade em uma série temporal financeira, é preciso filtrar o tipo de não-linearidade existente. Hsieh (1989) propôs um teste para detectar os dois tipos diferentes de não-linearidade: aditiva (não-linearidade na média) ou multiplicativa (não-linearidade na variância).

As hipóteses do teste são:

 H_0 : dependência não linear na variância condicionada

 H_1 : dependência não linear na média

A dependência multiplicativa implica que a esperança condicionada dos resíduos dada as defasagens anteriores da variável dos log retornos R_i e R_{i-1} é zero:

$$E[R_{t}|R_{t-1},...,R_{t-k}] = 0 (15)$$

Já a dependência aditiva implica que a mesma esperança condicionada seja diferente de zero:

$$E[R_t|R_{t-1},...,R_{t-k}] \neq 0$$
 (16)

A estatística do teste de Hsieh para a hipótese nula é:

$$\hat{\rho}(i,j) = \frac{\left(\frac{1}{n}\right)\sum_{t} E\left[R_{t}\right] E\left[R_{t-i}\right] E\left[R_{t-j}\right]}{\left[\left(\frac{1}{n}\right)\sum_{t} \left(E\left[R_{t}\right]\right)^{2}\right]^{\frac{3}{2}}}$$
(17)

sendo i, j > 0 representam as defasagens do processo $\{R_i\}$ que segue assintoticamente uma distribuição normal padrão N(0,1).

A partir de agora, pode-se aplicar os devidos modelos na análise da série temporal, uma vez que todo o processo de identificação da modelagem correta se fecha.

Nos itens seguintes, tem-se a descrição sucinta dos modelos aplicáveis a séries temporais financeiras.

2.3 Modelos para séries temporais financeiras

Nesta seção, serão abordados os principais modelos adotados na literatura científica nacional e internacional acerca do ajuste e das previsões de séries temporais financeiras. Os modelos aqui apresentados mereceram extensos debates e diversos aperfeiçoamentos ao longo dos anos. Destaca-se um resumo dos modelos com suas principais características matemáticas voltados para séries temporais financeiras.

Os modelos para séries temporais já revisados pela literatura mostram que para séries financeiras os melhores são os não-lineares. Para Enders (2004, p. 32), os modelos da família ARCH e GARCH são os mais comuns a serem aplicados na modelagem da variância condicionada, embora outros modelos possam ser ajustados.

Para Tsay (2005, p. 154), existem ainda outras classes de modelos não-lineares como os modelos de volatilidade estocástica que modelam a variância através de um processo não observado, tentando captar a notícia que chega ao mercado. São os modelos TAR (Limiar Auto-regressivo) que incluem não-linearidades na esperança e os bilineares.

A seguir, serão destacados os principais modelos citados acima com suas equações e características peculiares a cada um deles.

2.3.1 Modelos ARCH

Como visto, há uma variedade muito rica em modelos não lineares disponíveis na literatura de séries temporais. Todavia, o presente trabalho concentra-se mais especificamente na classe dos modelos ARCH (*Autoregressive Conditional Heterocedasticity*), desenvolvidos originalmente por Engle (1982, p. 991) e suas extensões.

O objetivo dessa classe de modelos é modelar a medida que se chama de volatilidade, que nada mais é do que a variância condicional da série dos log retornos. A volatilidade manifesta-se numa série de log retornos em grupos de maior ou menor variações no seu comportamento, evoluindo continuamente no tempo, chegando até a ser considerada estacionária (MORETTIN e TOLOI, 2004, p. 313).

Com o intuito de homogeneizar as notações, a partir da equação (2) que também pode ser escrita na forma abaixo.

$$R_{t} = \ln\left(\frac{P_{t}}{P_{t-1}}\right) = \ln\left(\frac{y_{t}}{y_{t-1}}\right) = \ln\left(y_{t}\right) - \ln\left(y_{t-1}\right)$$

$$\tag{18}$$

Considere então os seguintes valores estatísticos calculados a partir da série exposta pela equação (18) que se refere à média e à variância condicionada em que \mathcal{F}_{t-1} é a informação dada pela série até o instante t-1 que considera-se ser $\mathcal{F}_{t-1} = \{R_{t-1}, \dots, R_1\}$:

$$\mu_{t} = E[R_{t} | \mathcal{F}_{t-1}] = 0$$
 (18)

Média Condicionada

$$h_{t} = Var\left(R_{t} \middle| \mathcal{F}_{t-1}\right) \tag{19}$$

Variância Condicionada

Assume-se, por conseguinte que
$$\mu_t = 0$$
 e que $h_t = E(R_t^2 | \mathcal{F}_{t-1})$.

A idéia básica do modelo é assumir que os retornos são não correlacionados serialmente, mas a variância condicional depende dos retornos passados por meio de uma função quadrática que pode ser escrito pela seguinte equação para o modelo ARCH (r) (MORETTIN e TOLOI, 2004, p. 315):

$$R_{t} = \sqrt{h_{t}} \varepsilon_{t} \tag{20}$$

$$h_{t} = \alpha_{0} + \alpha_{1}R_{t}^{2} + \ldots + \alpha_{r}R_{t-r}^{2}$$
(21)

onde ε_i é uma sequência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d) com média zero e variância um, $\alpha_0 > 0, \alpha_i \geq 0, i > 0$.

Com maior rigor estatístico Campbell, Lo e Mackinlay (1997, p. 210) afirmam que os valores de ε_t são, em geral, supostos independentes e identicamente distribuídos (i.i.D), mas buscam ser escritos por uma função não-linear relacionando à série dos log retornos R_t com os chamados choques estocásticos, cuja representação geral é dada pela expressão a seguir:

$$X_{t} = f\left(\left\{\varepsilon_{s}\right\}_{s=t}^{-\infty}\right) = f\left(\varepsilon_{t}, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\right)$$
(22)

onde se assume que tais choques possuem média zero e variância um. Os autores definem ε_t como um choque estocástico contemporâneo e os passados $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$ como os choques estocásticos passados, isto é $\{\varepsilon_s\}_{s=t}^{-\infty} \sim \text{RBI}$. Maiores detalhes matemáticos pode ser encontrado nos autores acima citados.

A estimação deste modelo passa inicialmente pelo ajuste de um modelo ARIMA cujo objetivo é remover a correlação serial da série, caso exista, obtendo dessa forma o seu resíduo através da aplicação deste modelo. Se for o caso tem-se:

$$\phi(B)R_{\bullet} = \theta_0 + \theta(B)a_{\bullet} \tag{23}$$

em que a_r segue um modelo ARCH(r).

Tais parâmetros são obtidos pelo método de máxima verossimilhança condicional, que é dada pela equação abaixo, na hipótese dos ε_t serem normais:

$$L(\alpha | R_1, ..., R_n) = f(R_n | \mathcal{F}_{n-1}) f(R_{n-1} | \mathcal{F}_{n-2}) \cdots f(R_{n+1} | \mathcal{F}_t) f(R_1, ..., R_t | \alpha)$$
(24)

A maximização desta função pode ser obtida por algoritmos numéricos como Neuwton Raphson e outros. As previsões para a volatilidade pelo modelo ARCH (r) são obtidas recursivamente pela equação:

$$\hat{h}_{t} = \alpha_{0} + \alpha_{1} R_{t}^{2} + \dots + \alpha_{r} R_{t-r}^{2}$$
(25)

$$\hat{R}_{t} = \sqrt{\hat{h}_{t}} \varepsilon_{t} \tag{26}$$

Segundo Tsay (2005, p. 167), os modelos da série ARCH dão o mesmo tratamento para valores positivos e negativos para os log retornos. Isto ocorre uma vez que os quadrados dos retornos entram na forma da volatilidade, todavia, na prática, sabe-se que a volatilidade reage de forma diferenciada a retornos positivos e negativos.

Outra característica destes modelos é que ao trabalhar com retornos ao quadrado, alguns valores grandes e isolados podem levar a previsões errôneas, além de usar um número relativamente grande de parâmetros para redução do erro no processo de geração do modelo.

Uma tentativa de redução dos parâmetros é encontrada nos modelos GARCH, que serão descritos a seguir.

2.3.2 Modelos GARCH

Os modelos da família GARCH (*Generalized* ARCH) representam uma generalização dos modelos ARCH (*Autoregressive Conditional Heteroskedasticity*) isto é, de heterocedasticidade autoregressiva condicionada, foi desenvolvido por Bollerslev (1986, p. 308) para descrever a volatilidade com um número menor de parâmetros do que os usados num modelo ARCH.

Dessa forma, um modelo GARCH (p,q), pode ser definido a partir da seguintes expressões:

$$R_{t} = \sqrt{h_{t}} \varepsilon_{t} \tag{27}$$

$$h_{t} = \alpha_{0} + \sum_{i=1}^{p} \alpha_{i} R_{t-i}^{2} + \ldots + \sum_{i=1}^{q} \beta_{j} h_{t-j}$$
(28)

onde ε_t é uma sequência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d) com média zero e variância um, $\alpha_0 > 0, \alpha_i \ge 0, \beta_j \ge 0$ e ainda

$$\sum_{i=1}^{s} (\alpha_i + \beta_i) < 1 \text{ e } s = \max(p, q).$$

Segundo Morettin (2008, p. 131), a identificação da ordem do modelo GARCH a ser ajustado para uma série temporal financeira é usualmente trabalhosa. Por isso, o autor recomenda o uso de modelos de ordem baixa. Um modelo bastante utilizado na prática é o GARCH (1,1), para o qual a volatilidade é expressa como:

$$h_{t} = \alpha_{0} + \alpha_{1} R_{t-1}^{2} + \beta_{1} h_{t-1}$$
 (29)

As previsões de volatilidade utilizando um modelo GARCH, são feitas da mesma maneira que para o modelo ARCH descritas anteriormente.

Deve-se ficar atento para muitas situações práticas em que, no modelo gerado pela equação (29), pode-se ter $\alpha_1 + \beta_1$ muito próximo de um. Nesse caso, define-se o modelo gerado como IGARCH (*Integrated GARCH*), e a variância incondicional de R_t não será definida, o que revela baixa capacidade preditiva do modelo.

Pode-se tentar ajustar esse fato dos parâmetros do modelo por indicarem baixa capacidade preditiva e pelo fato de que o mercado está exposto a uma série de informações discrepantes advindos dos mais diversos meios de comunicação e não captados pelo modelo.

Esses fatores exógenos presentes no dia a dia das operações dos negócios no mercado financeiro acabam impactando diretamente nos preços dos ativos. Tais informações não comuns e podem também acarretar interpretações dúbias sobre o comportamento futuro dos mesmos ativos.

Uma saída para esses fatos na modelagem das séries temporais financeiras é usar modelos que tentam captar essas informações e aprender com elas para diagnosticar eventos futuros. A esses modelos com aprendizado recorrente através dos próprios elementos são tratados no mercado como modelos de redes neurais que serão abordados a seguir.

2.3.3 Modelos de Redes Neurais

Os modelos de redes neurais, ou mais precisamente redes neurais artificiais, são modelos de processamento paralelos distribuídos, formados por unidades de ajustes simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para o uso (HAYKIN, 2001, p. 28).

Ainda segundo o mesmo autor, o conhecimento experimental nas redes neurais é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem e guardados em variáveis chamadas neurônios. A força de conexão entre os neurônios é chamada de pesos sinápticos.

O procedimento utilizado para realizar o processo de aprendizagem é chamado de algoritmo de aprendizagem, cuja função é alterar os chamados pesos sinápticos de forma ordenada. A modificação desses pesos é o método tradicional nas redes neurais para conseguir o aprendizado.

A unidade básica da rede neural é o neurônio que é a peça fundamental para a operacionalização da rede. A figura a seguir, apresenta o modelo estrutural de um neurônio:

Figura 2.1: Modelo não linear de um neurônio

Fonte: HAYKIN (2001, p. 36)

Nas redes neurais, um conjunto de sinapses, também chamados elos de conexão, é caracterizado por um peso ou força própria. Cada entrada da rede conectada a um neurônio é multiplicado pelo peso sináptico. Vale ressaltar como estão escritos os

índices do peso sináptico, uma vez que o primeiro índice se refere ao neurônio em questão e o segundo ao terminal de entrada da sinapse à qual o peso se articula.

Os sinais de entrada são então somados ponderados pelas respectivas sinapses do neurônio constituindo uma combinação linear. A função de ativação é também referida como função restritiva já que limita o intervalo permissível de amplitude do sinal de saída a um valor finito.

Usualmente, a função de transferência mais simples é a função linear, a qual iguala a saída à entrada. A possibilidade de utilização de funções de transferência não-lineares é uma das características principais da rede neural, que podem ser observadas pelos elementos da figura acima.

Na figura 2.1, os x_1 , x_2 , ..., x_m são os sinais de entrada da rede; w_{k1} , w_{k2} , ..., w_{km} são os pesos sinápticos do neurônio k; u_k é a saída do combinador linear devido aos sinais de entrada; b_k é o viés; $\varphi(\bullet)$ é a função de ativação e y_k é o sinal de saída do neurônio. O uso do viés b_k tem o efeito de aplicar uma transformação afim à saída u_k do combinador linear.

Ligando os elementos acima do ponto de vista matemático, pode-se escrever o processo de funcionamento da rede através das seguintes equações:

$$u_k = \sum_{j=1}^{m} w_{kj} x_j$$
 (29)

$$v_k = u_k + b_k \tag{30}$$

$$y_k = \varphi(u_k + b_k) = \varphi(v_k) \tag{31}$$

As equações acima são descritas a partir da maneira pela qual os neurônios de uma rede neural estão estruturados, as quais ficam diretamente ligadas com o algoritmo de aprendizagem usado para treinar a rede para o aprendizado.

A propriedade de aprendizagem é crucial para as redes neurais, pois define sua habilidade de aprender e de melhorar o seu desempenho por meio dessa aprendizagem. Essa melhora se deve a uma medida preestabelecida de erro do sistema.

Haykin (2001, p. 46) afirma que, em geral, podem-se identificar três classes de arquiteturas de redes neurais: redes alimentadas adiante com camada única, redes alimentadas diretamente com múltiplas camadas e as redes recorrentes.

A rede neural em camadas é aquela que os neurônios ficam organizados em camadas e não há laços de realimentação, tanto na camada simples como nas multicamadas. Já as redes recorrentes consideram como ponto chave do processamento: os laços de realimentação, o que acaba melhorando a capacidade de aprendizagem e desempenho preditivo.

Dessa forma, as redes neurais vêm sendo uma melhor alternativa aos modelos tradicionais lineares e não lineares e a algumas abordagens não-paramétricas para modelagem de séries temporais financeiras.

Segundo Oliveira (2003, p. 89) a razão para o uso de uma rede neural é bem simples e direta, já que se pode encontrar uma abordagem para modelagem que aperfeiçoe as previsões para séries temporais financeiras que são dados altamente não lineares com uma quantidade baixa de parâmetros com facilidade de estimação dos mesmos.

Nas redes neurais o que fornece o seu poder de predição é justamente o seu processamento paralelo, ou seja, o processamento sequencial no qual apenas uma entrada observada da série é fornecida e com uma única saída a qual é obtida pela ponderação dos neurônios de entrada e na camada oculta, compõe o processo de forma paralela de tal forma a melhorar o desempenho da rede (ZHANG; PATUWO; HU, 1998, p. 47).

A idéia do uso das redes neurais está baseada em como projetar um modelo de múltiplas camadas e tratar esse modelo como um sistema dinâmico. De acordo com esse modelo, quer-se estimar um valor futuro com base no conhecimento de um conjunto informacional de valores passados que é assumida por conter informação suficiente para se prever a evolução futura do sistema.

Por um conjunto informacional de ordem p para um dado instante de tempo t define-se como sendo um conjunto de variáveis estocásticas p dimensionais indexadas e

equiespaçadas (HIRSCH, 1989, p. 340). Por exemplo, $\mathcal{F}_t^p = \left\{ \Phi_i \atop (p \times 1) \right\}_{i=-\infty}^{\infty}$ é conjunto informacional de ordem p=1 até o instante t inclusive.

O algoritmo inicia a partir de uma configuração arbitrária para os pesos sinápticos dos neurônios. Em respostas às variações estatísticas os pesos são ajustados de forma contínua no tempo. Os cálculos desses ajustes são completados dentro de um intervalo de convergência de tempo o qual é o período de amostragem. Esse processo é conhecido como filtro adaptativo (HAYKIN, 2001, p. 145).

Esse processo de filtragem é que irá constituir o laço de realimentação em torno do neurônio, constituindo um elemento de fundamental importância no processo de previsão de séries temporais por meio de redes neurais.

A seguir, serão descritos os métodos de filtragem e suas características operacionais que constituem um importante ponto de abordagem deste trabalho.

2.4 Filtros

A denominação 'filtro' advém da área de engenharia de comunicações e significa usar um mecanismo que possibilita a passagem de componentes com frequências em uma dada faixa de frequência (MORETTIN e TOLOI, 2004, p. 441).

Esta pesquisa entende que um filtro funciona na verdade como um processo de transmissão de dados que passaram por um processo de "limpeza". Esse processo ocorre mediante transformações matemáticas nas séries temporais que possibilitam essa purificação dos elementos da série.

Ilustrativamente, pode-se admitir a seguinte estrutura para o processo em que de um lado entra no filtro uma série temporal financeira de retornos R_t , e do outro lado sai uma série devidamente filtrada pelo processo FR_t , em que F é o filtro aplicado. A notação para esse procedimento seria $FR_t = F(R_t)$.

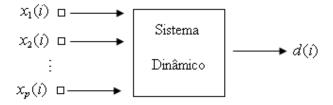
Figura 2.2: Esquema básico de um filtro



Fonte: Adaptado de MORETTIN e TOLOI (2004, p. 441)

Mais precisamente, HAYKIN (2001, p. 146) afirma que o processo de filtragem envolve o cálculo de dois tipos de sinais: uma saída, representada por y(i), gerada em resposta aos p elementos do vetor de estímulo x(i), e um sinal de erro, representado por e(i), que é obtido comparando-se a saída y(i) com a saída correspondente d(i), produzida pelo sistema agindo como uma resposta desejada ou sinal alvo. Esse processo pode ser visto na figura abaixo em que define especificamente o processo de filtragem como um sistema dinâmico:

Figura 2.3: Modelo de um sistema dinâmico



Fonte: HAYKIN (2001, p. 145)

Morettin e Toloi (2004, p. 415) atribuem o uso deste tipo de processo ao campo da análise espectral que de uma maneira geral decompõe a série temporal estacionária em seus componentes senoidais com coeficientes aleatórios não correlacionados.

Ainda segundo os autores, a principal razão pelo uso deste tipo de análise no tratamento de séries temporais é pelo fato de que o espectro fornece uma descrição bastante simples do efeito de uma transformação linear de um processo estacionário.

Existem diversos tipos de filtros que podem ser usados no tratamento de séries temporais. Neste trabalho serão usados dois filtros: de *wavelets* e de Kalman. Estes serão descritos a seguir.

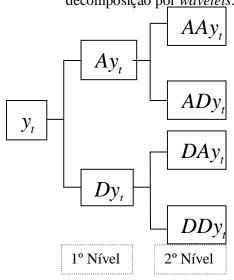
2.4.1 Filtro de wavelets

O processo de filtragem por *wavelets* ou ondaletas tem por finalidade fazer a separação dos dados da série original em duas outras sub-séries por seus componentes de frequência. As *wavelets* são funções matemáticas que ampliam intervalos de dados, possibilitando que cada componente seja alocado em sua respectiva escala (MISITI et al., 1997, p. 4-3).

Polikar (1999, p. 7) afirma que este processo de filtragem para alocar cada componente da série temporal na sua devida escala refere-se a identificação dos coeficientes correspondentes a cada escala, se alta ou baixa frequencia, formando suas novas sub-séries. A filtragem inversa, também chamada transformada inversa, consiste em aplicar os filtros inversos no sinal decomposto e tem o poder de reconstruir o sinal original, juntando novamente as duas bandas de frequencia.

O autor cita ainda, que o processo pode ser iterativo, ou seja, usar uma banda de filtros que aplicadas sucesivamente irão decompor o sinal sempre em duas novas faixas de frequencia, uma de alta frequencia (chamada detalhe ou diferença) e outra de baixa frequencia (chamada aproximação), formando uma árvore de decomposição com 2^n caminhos diferentes para a decomposição da série, como poder ser visto na figura a seguir:

Figura 2.4: Árvore de decomposição em dois níveis de uma série temporal via decomposição por *wavelets*.



Fonte: MISITI et al., 1997.

O uso das *wavelets* é segundo Gençay, Selçuk e Whitcher (2002, p. 10), é válido graças a sua capacidade de decompor uma série temporal em escalas referentes ao domínio da frequência quanto ao domínio do tempo. É esse poder de separação que a diferencia, segundo os autores, da análise de Fourier, onde somente se conhece as frequencias já que usam funções para separação da forma de senos e cossenos que são periódicas.

Para Percival e Walden (2006, p. 32), as *wavelets* dizem ser que são dependentes de ambos os domínios da freqüência e do tempo sendo que, da freqüência é via dilação da sua forma de onda e do tempo é via translação da onda, o que pode ser uma vantagem em diversos casos de análise de séries.

Ainda segundo os autores, as bases das funções de Fourier não apresentam bom desempenho para o tratamento local de dados, pois são séries de comportamento infinitas e não se ajustam à análise de dados descontínuos como é o caso das séries temporais financeiras, daí o surgimento de funções que se prestam a esses efeitos temporais que são as *wavelets*.

Interpretando as informações destes autores e aplicando a pesquisa em questão, pode-se realizar o processo de filtragem comparando as formas das funções que se prestam a esse processo. Na figura a seguir, considerou-se a título de ilustração uma série de log retornos de uma série de um ativo financeiro e aplicou-se o processo de filtragem por uma onda senoidal e por uma *wavelet* (no caso foi escolhido a forma de *wavelet* de Morlet).

Pode-se observar que onde as ondas senoidais são lisas e suaves, as ondaletas tendem a ser irregulares e assimétricas, captando assim o melhor movimento dos retornos dos ativos.

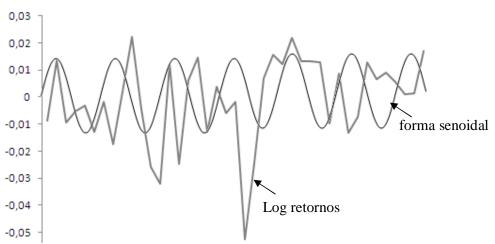
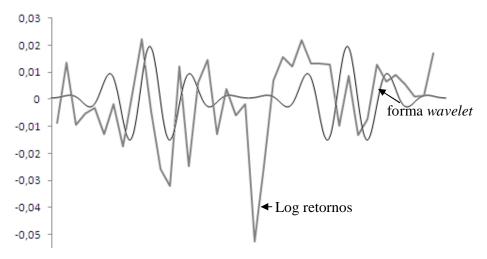


Figura 2.5: Processo de filtragem por uma onda senoidal e por uma wavelet

(a) Processo de filtragem por função senoidal



(b) Processo de filtragem por função de wavelet de Morlet

Como visto na figura acima, as *wavelets* podem alterar sua forma e comprimento de onda. Tais mudanças ocorrem através de suas formulações matemáticas, pois cada uma delas são descritas por um conjunto de pontos que se expressam através de dilatações ou compressões e translações a partir de uma forma de onda original, chamada de ondaleta mãe (PERCIVAL e WALDEN (2006, p. 44)).

No tratamento de análise de séries temporais, o uso das *wavelets* não rompe nenhum valor da série original. Possui ainda a vantagem de reconhecer os pontos onde ocorrem picos alternados de maior ou menor oscilação. O que se procura então é que a partir de uma série temporal financeira possa-se obter suas sub-séries representativas de alta e baixa frequência numa tentativa de suavizar os efeitos das variações no mercado financeiro através da aplicação de um filtro por uma função de *wavelet*.

Segundo Misiti (2007, p. 21), para que uma função Ψ_t possa ser considerada uma *wavelet*, essa função tem que atender as seguintes propriedades:

a)
$$\int_{-\infty}^{\infty} \Psi(t)dt = 0$$
 (32)

b)
$$\int_{-\infty}^{\infty} \left| \Psi(t) \right|^2 dt < \infty \tag{33}$$

As expressões acima equivalem a dizer que uma wavelet é uma função que pertence ao conjunto das funções quadrado integrável, representadas por $L^2(IR)$. Em outras palavras, isto quer dizer que a função tende a oscilar acima e abaixo do eixo t.

Morettin (1999, p. 161) define o produto interno de duas funções f(t) e g(t) em $L^2[a,b]^2$ como sendo:

$$\langle f(t), g(t) \rangle = \int_a^b f(t).g^*(t)dt$$
 (34)

Define ainda a transformada de *wavelet* (WT) como o produto interno do sinal de teste com as funções base $\Psi(a,b)(t)$:

$$WT_{x}^{\psi}(a,b) = \Psi_{x}^{\psi}(a,b) = \int y(t).\psi_{a,b}^{*}(t)dt$$
 (35)

onde $\psi_{a,b} = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right)$ e y(t) é a série temporal financeira em estudo, que pode ser uma série de log retornos.

Ainda segundo o autor, o objetivo é estender essa função para o campo das funções chamada $L^2(IR)$, isto é, gerar um espaço, a partir de uma função Ψ , que é obtida por meio de dilatações através do parâmetro "a" da sentença matemática ou compressões e translações definidos pelo parâmetro "b" de Ψ , dada por:

_

² Denota o conjunto de funções quadráticas integráveis no intervalo [*a,b*].

$$\Psi_{a,b}(x) = |a|^{-\frac{1}{2}} \Psi\left(\frac{x-b}{a}\right), a, b \in IR, a \neq 0$$
 (36)

Uma análise de *wavelets* é definida pela aplicação da transformada de *wavelet* com diversos valores para os parâmetros de dilatação e translação representando a decomposição da série original em seus respectivos componentens localizados no tempo e na frequência para estes parâmetros.

A forma de onda Ψ é chamada de *mother wavelet* (ondaleta mãe) e os parâmetros a e b tomam os seguintes valores especiais, $a=2^{-j}$ e $b=k2^{-j}$, que compõem a equação:

$$\Psi_{j,k}(x) = 2^{\frac{j}{2}} \Psi(2^{j} x - k), j, k \in \mathbb{Z}$$
(37)

obtida pela dilatação binária 2^{-j} e pela translação diática $k2^{-j}$, conforme sugere Morettin (1999, p. 163).

Pode-se gerar uma wavelet pela função escala, ou chamada de ondaleta pai, Φ , que é uma solução da equação

$$\Phi(t) = \sqrt{2} \sum_{k} l_k \Phi(2t - k) \tag{38}$$

onde
$$l_k = \sqrt{2} \int_{-\infty}^{\infty} \Phi(t) \Phi(2t - k) dt$$

Essa função gera uma família ortonormal de $L^2(IR)$,

$$\Phi_{jk}(x) = 2^{\frac{j}{2}} \Phi\left(2^{j} x - k\right), j, k \in \mathbb{Z}$$

$$\tag{39}$$

A análise de wavelet é feita pela aplicação sucessiva da transformada de wavelet com diversos valores para *a* e *b*, representando a decomposição do sinal original em diversos componentes localizados no tempo e na freqüência, de acordo com estes parâmetros.

Gençay, Selçuk e Whitcher (2002, p. 133) diz que uma série temporal financeira pode ser decomposta por uma análise de *wavelets*, por uma sequência de projeções de *wavelets* pai e mãe, a partir das funções $\Phi \in \Psi$, como seguem as equações (37) e (39).

A *wavelet* mãe funciona como uma janela de cobertura finita que persegue a série temporal. A captação dos pontos de alta e baixa freqüência se dá pela translação e dilatação da forma da *wavelet*.

Essa representação para série temporal y, pode ser dada por:

$$y_{t} = \sum_{k} a_{j,k} \Phi_{j,k}(t) + \sum_{k} d_{j,k} \Psi_{j,k}(t) + \sum_{k} d_{j-1,k} \Psi_{j-1,k}(t) + \dots + \sum_{k} d_{1,k} \Psi_{1,k}(t)$$

$$(40)$$

onde j é o número de componentes e k que varia de 1 ao número de coeficientes do componente específico. Os coeficientes $a_{j,k}, d_{j,k}, \ldots, d_{1,k}$ são os coeficientes das transformadas de *wavelets* dadas pelas projeções

$$a_{j,k} = \int \Phi_{j,k}(t).y_t dt$$
, chamada parte de aproximação (41)

e

$$d_{j,k} = \int \Psi_{j,k}(t).y_t dt, \text{ chamada parte de detalhe}$$
 (42)

Dessa forma, o objetivo expresso pelo uso das *wavelets*, conforme descrito por Donoho e Tohnstone (1994, p. 439), é o de redução do ruído, também conhecido como *denoising*, que a análise de *wavelets* realiza na escolha dos coeficientes que devem ser mantidos para preservar a informação e consistência dos dados da série temporal financeira original.

Em outras palavras, segundo os autores, a idéia central é a supressão daqueles coeficientes que provocam ruído na série de tempo, fazendo com que a série fique mais lisa do ponto de vista técnico e visual. A seguir ilustra-se este processo para uma série bruta de informações do Ibovespa no período de 03/01/2000 a 31/12/2009.

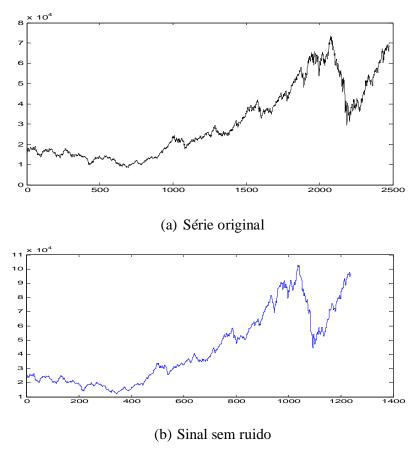


Figura 2.6: Processo de análise de wavelets aplicado ao Ibovespa

Analisando as figuras acima, pode-se notar que o ruído foi eliminado da série temporal. Isso ocorre a partir da aplicação da transformada de *wavelet* com a escolha da forma de onda, no caso foi usada a forma Daubesch 1³ em um único nível. Obtém-se dessa forma a análise dos coeficientes obtidos pela aplicação do filtro em alta frequência. O objetivo principal foi de escolher um número mínimo de coeficientes que conserve as características iniciais da série.

Como se observa acima, a análise de *wavelets* segue características próprias de funcionamento e redução de ruídos com objetivo de alisar a série, mas com preservação das características gerais de tendência, ciclos e sazonalidades.

A seguir, será demonstrado o uso de outro processo de filtragem aplicável a séries temporais, chamado Filtro de Kalman. Este processo considera justamente as composições da série em forma de tendência, comportamento cíclico e sazonalidade.

³ Para conhecimento das formas de *wavelets* pode-se verificar no Apêndice A.

2.4.2 Filtro de Kalman

O filtro de kalman foi introduzido por Rudolph Emil Kalman e apareceu na literatura em 1960 quando o autor descreveu seu algoritmo para solução de aplicação de filtro de dados discretos. (GREWAL e ANDREWS, 2008, p. 21).

Segundo os mesmos autores, filtro de Kalman é um conjunto de equações matemáticas desenvolvidas em forma de um algoritmo computacional que formam um processo iterativo desenvolvido para realizar previsões futuras e estimar variâncias de modelos para séries temporais.

Harvey (2001, p. 22) lembra que este algoritmo é um processo aplicável quando se pode escrever a série temporal como uma forma de espaço de estado. Grewal e Andrews (2008, p. 22) afirmam que quase todos os modelos convencionais de séries temporais têm uma representação nesta forma.

Isso significa que através do processo de filtragem pelo filtro de Kalman, partese de uma variável observável (a série temporal financeira) e consegue-se estimar outra variável não observável chamada variável de estado, podendo estimar os estados passados, presente e futuros através da previsão de valores.

Essa estimação, segundo Harvey (2001, p. 24), dos parâmetros que são desconhecidos, ocorre pelo processo de maximização da verossimilhança por decomposição do erro de previsão, conforme já comentado neste trabalho.

Oliveira (2007, p. 74) cita que o filtro de Kalman é extremamente útil e tem bons resultados encontrados na literatura, devido a sua otimalidade e estrutura de fornecer formulações de fácil implementação e processamento em tempo real.

A estruturação⁴ do modelo na forma de espaço de estados é feita a partir de uma série temporal $\{y_t\}_{t=1}^n = \{y_1, y_2, ..., y_t, ..., y_n\}$ com n elementos. Tais variáveis são denominadas variáveis observáveis e representam um vetor $n \times 1$ e se relacionam com

_

⁴ O desenvolvimento desta parte é baseado em HARVEY (2001, p. 100-166).

as variáveis de estado x_t por um processo de Markov⁵ gerando uma equação chamada de equação de medição ou observação:

$$y_{t} = A_{t}x_{t} + \varepsilon_{t} \tag{43}$$

com t=1,...,T, sendo A_t é uma matriz $n\times m$, ε_t um vetor serialmente não correlacionado com média zero e matriz de covariância M_t e x_t é um vetor $m\times 1$ que contém as variáveis de estados não observáveis.

Como visto na equação (43) acima, o cálculo de um estado de um sistema dinâmico linear no tempo t, x_t é calculado recursivamente partindo-se das estimativas anteriores do estado no tempo t-1, x_{t-1} e dos novos dados fornecidos na entrada y_t não sendo necessário o armazenamento de todos os dados anteriores para se estimar o estado atual do sistema.

Entende-se por estado de um sistema um vetor coluna $m \times 1$ contendo variáveis que são de interesse do analista. Para isso, usa-se o filtro de Kalman, muitas vezes, combinado com redes neurais, onde estas variáveis descritas como de interesse do analista são dadas pelos pesos da rede neural, com objetivo de encontrar as melhores estimativas para estas variáveis.

As variáveis de estado x_t são geradas no modelo por um processo de Markov de primeira ordem, gerando uma nova equação matemática denominada equação de transição:

$$x_t = B_t x_{t-1} + \zeta_t \tag{44}$$

com $t=1,\ldots,T$, sendo B_t é uma matriz $n\times m$, ς_t um vetor serialmente não correlacionado com média zero e matriz de covariância V_t .

Os vetores ε_t e ζ_t são assumidos como ruído branco com média zero com matrizes de covariância Q_t e não serialmente correlacionados entre si e não correlacionados com o estado inicial. Além disso, o estado inicial do sistema x_0 tem média \hat{x}_0 e matriz de covariância Q_0 .

_

⁵ Para mais informações sobre os processos de Markov, veja apêndice B.

Assim sendo, ao se definir \hat{x}_{t-1} uma estima do estado anterior no momento t dado que se conhece todo o processo anterior a t, e \hat{x}_{t+1} a estimativa do estado posterior, podese conhecer o erro de medição ou de observação como:

$$e_{t-1} = x_{t-1} - \hat{x}_{t-1} \tag{45}$$

$$e_t = x_t - \hat{x}_t \tag{46}$$

Chegam-se então as matrizes de covariância do erro anterior P_{t-1} e posterior P_t todas $m \times m$ e definidas por:

$$P_{t-1} = E \left[e_{t-1} e_{t-1}^T \right] \tag{47}$$

$$P_{t} = E \left[e_{t} e_{t}^{T} \right] \tag{48}$$

Dessa forma, pode-se então descrever as equações aplicáveis ao filtro de Kalman:

ganho de Kalman

$$K_{t} = \frac{P_{t|t-1}A_{t}^{T}}{A_{t}P_{t|t-1}A_{t}^{T} + V_{t}}$$
(49)

• equação de atualização

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t \left[y_t - A_t \hat{x}_{t|t-1} \right]$$
(50)

$$P_{t|t} = P_{t|t-1} - K_t A_t P_{t|t-1}$$
 (51)

equação de previsão

$$\hat{x}_{t+1|t} = B_{t+1|t} \hat{x}_{t|t} \tag{52}$$

$$P_{t+1|t} = B_{t+1|t} P_{t|t} B_{t+1|t}^T + Q_t$$
 (53)

Assim, as equações do filtro de Kalman se mantém interligadas através das estimativas do vetor de estado \hat{x} e pela matriz de correlação estado do erro P_t . As equações de atualização fazem a correção dos y_t para cada passo t, enquanto que as equações de previsão realizam a estimativa futura para o instante t+1, um passo a

frente, antes da próxima medida seja tornada disponível no sistema. Tal processo é repetido recursivamente até a convergência do estado.

Por conseguinte, se \hat{x}_t é dado como o estado atual do sistema, então $\hat{x}_{t|t-1}$ referese a estimativa do estado para o passo t dado o conhecimento das variáveis no passo anterior t-1, enquanto que $\hat{x}_{t+1|t}$ refere-se a estimativa do estado no instante t+1 dadas as informações no instante anterior t.

Fato semelhante ocorre com as matrizes de correlação do estado de erro sendo descritas pelas sentenças:

$$P_{t|t-1} = E \left[\left(x_t - \hat{x}_{t|t-1} \right) \left(x_t - \hat{x}_{t|t-1} \right)^T \right]$$
(54)

$$P_{t|t} = E\left[\left(x_t - \hat{x}_{t|t}\right)\left(x_t - \hat{x}_{t|t}\right)^T\right]$$
(55)

A descrição do filtro de Kalman está apoiada no constructo de que tanto os ruídos das equações de medição e de transição seguem uma distribuição normal. Em outras palavras estatísticas, bastaria dizer que os dois primeiros momentos são suficientes para descrever os estados do sistema, sendo $\hat{x}_t = E[x_t]$ e $P_t = E[(x_t - \hat{x}_t)(x_t - \hat{x}_t)^T]$. O estimador é muitas vezes dito ótimo por minimizar a variância do erro.

Esse princípio de iteração, descrito acima, é o coração do filtro de Kalman, onde cada nova observação, em um dado instante de tempo, é tornada disponível pelo sistema, tanto o vetor do espaço do estado como a matriz de covariância dos estados são atualizadas.

Conforme foi descrito, o processo do filtro de Kalman é um filtro por obter o valor filtrado em um instante *t* que nada mais é do que o valor esperado da variável de estado condicional para as informações que estão à disposição até aquele instante *t*.

Todavia, sabe-se que os pressupostos operacionais do filtro de Kalman implicam que o sistema deve estar descrito na forma de espaços de estado linear e que os ruídos são brancos e seguem uma distribuição gaussiana e não são correlacionados entre si e

com o estado inicial. Satisfeitas estas condições, o filtro de Kalman minimiza o erro quadrático médio sendo, portanto, um estimador ótimo (Harvey, 2001, p. 30).

Ao se tratar de séries temporais financeiras, que apresentam características de não lineariedade, rompe-se uma das suposições básicas do uso do filtro de Kalman. Para solucionar este problema, aplica-se então o filtro de Kalman estendido que será descrito a seguir.

2.4.3 Filtro de Kalman estendido

Conforme descrito anteriormente, a aplicação do filtro de Kalman fica prejudicada sob certas condições. Além das descritas acima, tem a questão das matrizes definidas em (43) e (44) já que, quando são ditas estocásticas, ou seja, dependem da informação anterior disponível, o modelo é considerado condicionalmente gaussiano.

Caso as matrizes não sejam estocásticas, ou seja, quando os ruídos não seguem uma distribuição normal, o filtro de Kalman, na forma clássica descrita acima, não pode mais ser aplicado. Há ainda o caso da função linear dos estados não ser linear e nem a equação do vetor de estado também não ser linear. A solução passa a ser a linearização dos modelos.

Considere então, segundo Harvey (2001, p. 155), um sistema não linear dado pelas equações:

$$y_t = h_t x_t + r_t \tag{56}$$

$$x_t = f_t x_t + \zeta_t \tag{57}$$

em que h_t e f_t são funções vetoriais não lineares das equações definidas em (43) e (44).

O processo de linearização destas equações é feito aplicando a elas a expansão em séries de Taylor, sobre suas médias condicionais $\hat{x}_{t|t-1}$ e $\hat{x}_{t|t}$ conforme descrito a seguir:

$$h_{t}(x_{t}) = h_{t}(\hat{x}_{t|t-1}) + A_{t}(x_{t} - \hat{x}_{t|t-1}) + \dots$$
(58)

$$f_t(x_t) = f_t(\hat{x}_{t|t}) + B_t(x_t - \hat{x}_{t|t}) + \dots$$
 (59)

onde as matrizes A_t e B_t são matrizes Jacobianas definidas por:

$$A_{t} = \frac{\partial h_{t}(\widehat{x}_{t|t-1})}{\partial x} \tag{60}$$

$$B_{t} = \frac{\partial f_{t}(\widehat{x}_{t|t})}{\partial x} \tag{61}$$

Desconsiderando os termos de ordem superior a dois na expansão de Taylor, e substituindo as expressões (58) e (59) em (56), tem-se:

$$y_{t} = A_{t} x_{t} + h_{t}(\hat{x}_{t|t-1}) - A_{t} \hat{x}_{t|t-1} + r_{t}$$
(62)

$$x_{t} = B_{t}x_{t} + f_{t}(\widehat{x}_{t|t}) - B_{t}\widehat{x}_{t|t} + \zeta_{t}$$

$$\tag{63}$$

Dessa forma, as formulações para o filtro de Kalman estendido ficam:

• ganho de Kalman

$$K_{t} = \frac{P_{t|t-1}A_{t}^{T}}{A_{t}P_{t|t-1}A_{t}^{T} + V_{t}}$$
(64)

• equação de atualização

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t \left[y_t - h_t \hat{x}_{t|t-1} \right]$$
(65)

$$P_{t|t} = P_{t|t-1} - K_t A_t P_{t|t-1}$$
(66)

• equação de previsão

$$\hat{x}_{t+1|t} = f_t \hat{x}_{t|t} \tag{67}$$

$$P_{t+1|t} = B_{t+1|t} P_{t|t} B_{t+1|t}^T + Q_t$$
(68)

Comparando-se as equações (49) a (53) e (64) a (68), podem-se notar poucas alterações nas suas formulações. A presença dos elementos linearizados pelas equações de Taylor $h_t \hat{x}_{t|t-1}$ no lugar de $A_t \hat{x}_{t|t-1}$ e $f_t \hat{x}_{t|t}$ no lugar de $B_{t+1|t} \hat{x}_{t|t}$, além da substituição

das matrizes de transição dos estados são trocadas pelas suas respectivas Jacobianas $A_t = \frac{\partial h_t(\widehat{x}_{t|t-1})}{\partial x} e^{-B_t} = \frac{\partial f_t(\widehat{x}_{t|t})}{\partial x} \text{ no lugar de } A_t e^{-B_t} \text{ respectivamente.}$

Dessa forma, o filtro de Kalman e o filtro de Kalman estendido contribuem sensivelmente para a redução dos erros da previsão, principalmente quando usado em combinação com outras ferramentas como as redes neurais, conforme consta em trabalhos recentes no Brasil e na comunidade acadêmica internacional.

Diante deste fato, passa-se, a seguir, a moldar-se as diferentes formas de combinação do uso de filtros com metodologias de previsão para séries temporais financeiras.

2.4.4 Combinação de filtros

O uso combinado de filtros é uma das principais contribuições deste trabalho no sentido de se verificar a contribuição que este uso combinado tem para reduzir as medidas de qualidade das previsões para séries temporais financeiras.

Descreveu-se nos itens acima o uso de dois dos filtros com maior quantidade de pesquisas encontradas conforme trabalho levantado por Gooijer e Hyndman (2006, p. 451). Todavia, são poucos ainda os trabalhos que usaram essa combinação de filtros, só que aplicadas a área de engenharias como Oliveira (2005, p. 28) e Soares (2007, p. 89). Assim, nesta pesquisa, será destacado o uso combinado de filtros aplicados a séries temporais financeiras.

Assim, primeiramente, tem-se o modelo destacado pela aplicação da transformada de *wavelets* primeiro e posteriormente o uso do filtro de Kalman para previsão.

Seja então uma série temporal representativa dos log retornos denotada por $\left\{R_{t}\right\}_{t=1}^{n} = \left\{R_{1}, R_{2}, \ldots, R_{t}, \ldots, R_{n}\right\}$. Aplicando sobre esta série a transformada de *wavelets* (W) em um nível tem-se:

$$WR_{t} = \sum_{k} a_{j,k} \Phi_{j,k}(t) + \sum_{k} d_{j,k} \Psi_{j,k}(t) + \sum_{k} d_{j-1,k} \Psi_{j-1,k}(t) + \dots + \sum_{k} d_{1,k} \Psi_{1,k}(t)$$
(69)

onde j é o número de componentes e k que varia de 1 ao número de coeficientes do componente específico. Os coeficientes $a_{j,k}, d_{j,k}, \ldots, d_{1,k}$ são os coeficientes das transformadas de wavelets dadas pelas projeções

$$a_{j,k} = \int \Phi_{j,k}(t) R_i dt$$
, chamada parte de aproximação (70)

e

$$d_{j,k} = \int \Psi_{j,k}(t) R_i dt, \text{ chamada parte de detalhe}$$
 (71)

Aplicando-se sobre a equação (69) as equações do filtro de Kalman, tem-se:

• equação de medição

$$WR_{t} = A_{t}x_{t} + \varepsilon_{t} \tag{72}$$

equação de transição:

$$x_t = B_t x_{t-1} + \zeta_t \tag{73}$$

Agora, descrevendo-se as equações aplicáveis ao filtro de Kalman tem-se:

• ganho de Kalman

$$K_{t} = \frac{P_{t|t-1}A_{t}^{T}}{A_{t}P_{t|t-1}A_{t}^{T} + V_{t}}$$
(74)

equação de atualização

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t \left[WR_t - A_t \hat{x}_{t|t-1} \right]$$
(75)

$$P_{t|t} = P_{t|t-1} - K_t A_t P_{t|t-1}$$
 (76)

• equação de previsão

$$\hat{x}_{t+1|t} = B_{t+1|t} \hat{x}_{t|t} \tag{77}$$

$$P_{t+1|t} = B_{t+1|t} P_{t|t} B_{t+1|t}^T + Q_t (78)$$

em que P_t segue definido pelas equações (47) e (48).

Analogamente, pode-se definir a aplicação contrária, ou seja, aplicar primeiro o filtro de Kalman sobre a série dos log retornos $\left\{R_{t}\right\}_{t=1}^{n}=\left\{R_{1},R_{2},...,R_{t},...,R_{n}\right\}$ e posteriormente fazer a transformada de *wavelets*. Suas equações ficam assim definidas:

• equação de medição

$$R_{t}KF = A_{t}x_{t} + \varepsilon_{t} \tag{79}$$

equação de transição:

$$x_t = B_t x_{t-1} + \zeta_t \tag{80}$$

Agora, descrevendo-se as equações aplicáveis ao filtro de Kalman tem-se:

ganho de Kalman

$$K_{t} = \frac{P_{t|t-1}A_{t}^{T}}{A_{t}P_{t|t-1}A_{t}^{T} + V_{t}}$$
(81)

• equação de atualização

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t \left[R_t KF - A_t \hat{x}_{t|t-1} \right]$$
(82)

$$P_{t|t} = P_{t|t-1} - K_t A_t P_{t|t-1}$$
(83)

• equação de previsão

$$\hat{x}_{t+1|t} = B_{t+1|t} \hat{x}_{t|t} \tag{84}$$

$$P_{t+1|t} = B_{t+1|t} P_{t|t} B_{t+1|t}^T + Q_t (85)$$

em que P_t segue definido pelas equações (47) e (48).

Posteriormente, usando a transformada de *wavelets* definida pela equação (40), tem-se:

$$WR_{t}KF = \sum_{k} a_{j,k} \Phi_{j,k}(t) + \sum_{k} d_{j,k} \Psi_{j,k}(t) + \sum_{k} d_{j-1,k} \Psi_{j-1,k}(t) + \dots + \sum_{k} d_{1,k} \Psi_{1,k}(t)$$
(86)

onde j é o número de componentes e k que varia de 1 ao número de coeficientes do componente específico. Os coeficientes $a_{j,k}, d_{j,k}, \ldots, d_{1,k}$ são os coeficientes das transformadas de *wavelets* dadas pelas projeções

$$a_{j,k} = \int \Phi_{j,k}(t) R_t K F dt$$
, chamada parte de aproximação (87)

e

$$d_{j,k} = \int \Psi_{j,k}(t) R_t K F dt, \text{ chamada parte de detalhe}$$
(88)

E sobre estas duas possíveis combinações dos filtros aplicados primeiro sobre a série dos dados, usam-se os modelos tradicionais econométricos de previsão e também combinado com as redes neurais. O uso combinado dos filtros de Kalman com redes neurais será descrito a seguir.

2.4.5 Redes Neurais com filtro de Kalman estendido

Segundo Haykin (2001, p. 823), a necessidade do uso do filtro de Kalman em conjunto com as redes neurais advém do poder de explorar suas propriedades únicas em realizar treinamento supervisionado de uma rede recorrente. Singhall e Wu (1989, p. 1188) talvez tenham sido os primeiros a demonstrar a melhora do desempenho de uma rede neural supervisionada que utilizou o filtro de Kalman estendido.

Ainda segundo Haykin (2001, p. 824), o treinamento de redes neurais pode ser abordado como um problema de construção de espaços de estados, uma vez que as redes neurais trabalham de uma forma não linear. Sendo assim, a estimação dos estados é realizada via filtro de Kalman estendido.

Tal aplicação, segundo o autor, os pesos da rede são organizados em forma de vetor w_t em um dado instante de tempo t que ocuparão o lugar do vetor de estados x_t nas equações dos espaços de estados do filtro de Kalman. O processo de aplicação do algoritmo irá atualizar os pesos a cada instante de tempo e também, com isso, atualizando a matriz de correlação P_t dos erros conforme descrito nas equações (47) e (48).

O processo inicia-se definindo um vetor de resultado d_t como a resposta desejada organizado da forma $n \times 1$. O objetivo do treinamento é encontrar o vetor de

pesos w que a diferença entre a saída e o vetor esperado, ou seja, o erro, conforme definido nas equações (45) e (46) seja minimizado pelo erro quadrático médio.

Tendo em mente o processo de filtragem adaptativa, as equações do modelo de espaço de estados para a rede neural pode ser modelado pelas equações:

$$d_{t} = h_{t}(w_{t}, u_{t}, v_{t}) + r_{t} \tag{89}$$

$$W_{t+1} = W_t \tag{90}$$

onde u_t representa o vetor de entrada, v_t o vetor de atividades para cada nó da rede e r_t o vetor de ruído do sistema.

Associando as equações (89) e (90) com as equações (56) e (57) a função f_t fica sendo a matriz identidade e ζ_t segue a matriz nula. O vetor d_t assume o papel da equação de medição e a função h_t assume a medida de relação entre o vetor de entrada, saída e pesos da rede neural.

Com o paralelo traçado e associação feita pelas equações da rede e do filtro de Kalman, pode-se reescrever as equações do filtro de Kalman estendido com funções dos vetores de entrada, saída e pesos da rede neural, como é feito a seguir:

$$K_{t} = \frac{P_{t|t-1}A_{t}^{T}}{A_{t}P_{t|t-1}A_{t}^{T} + V_{t}}$$
(91)

$$\widehat{w}_{t|t} = \widehat{w}_{t|t-1} + K_t \left[d_t - h_t \left(\widehat{w}_{t|t-1} \right), u_t \right]$$
(92)

$$P_{t|t} = P_{t|t-1} - K_t A_t P_{t|t-1}$$
(93)

$$\widehat{W}_{t+1|t} = \widehat{W}_{t|t} \tag{94}$$

$$P_{t+1|t} = P_{t|t} \tag{95}$$

e a matriz Jacobiana C_t é calculada como segue:

$$C_{t} = \frac{\partial h_{t}(w_{t}, u_{t})}{\partial w} = \begin{bmatrix} \frac{\partial h_{1}}{\partial w_{1}} & \frac{\partial h_{1}}{\partial w_{2}} & \cdots & \frac{\partial h_{1}}{\partial w_{t}} \\ \frac{\partial h_{2}}{\partial w_{1}} & \frac{\partial h_{2}}{\partial w_{2}} & \cdots & \frac{\partial h_{2}}{\partial w_{t}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_{n}}{\partial w_{1}} & \frac{\partial h_{n}}{\partial w_{2}} & \cdots & \frac{\partial h_{n}}{\partial w_{t}} \end{bmatrix}$$

$$(96)$$

Na prática, essas derivadas parciais são calculadas usando o algoritmo de aprendizagem em tempo real ou por retropropação através do tempo. Dessa forma, o filtro de Kalman usa esse algoritmo para testar a convergência do erro.

De posse de todas as informações e combinações de técnicas descritas neste capítulo, passa-se agora a uma revisão dos trabalhos que se utilizaram de tais ferramentas para modelar e prever dados futuros a partir de séries temporais financeiras.

3. REVISÃO DA LITERATURA

Este capítulo expõe a revisão de estudos científicos, voltados para previsão de séries temporais e uso combinado de filtros de previsão, publicados no Brasil e na comunidade acadêmica internacional.

A revisão cobriu as seguintes bases de dados que foram pesquisadas: SciELO⁶ – *Scientific Electronic Library Online*; ISI – Web of Knowledge⁷; EBSCO Host⁸; ProQuest⁹; Base de Periódicos da Capes¹⁰; Portal Domínio Público¹¹; Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo¹²; Science Direct¹³; Portal Gale Cengage Learning¹⁴ e Portal de Estudos Econômicos da FEA-USP¹⁵.

3.1 Breve histórico sobre previsão de séries temporais

O periódico mais importante da área de modelagem para previsão futura é o *International Journal of Forecasting*, sendo respeitado na comunidade acadêmica como um *journal* de referência e excelência na área. Em 2006, Gooijer e Hyndman (2006, p. 443) lançaram um *paper* comemorativo de 25 anos de análise e previsão de séries temporais, onde fizeram uma varredura nas publicações deste periódico, cobrindo o período 1982 a 2005, onde selecionaram 340 artigos publicados no periódico e tiraram conclusões importantes acerca do passado da produção científica e principalmente sobre o futuro desta área.

A conclusão geral que chegaram foi de que foram realizados progressos enormes em muitas áreas onde foram realizadas previsões, mas acreditam que ainda há um grande número de tópicos que precisam de um maior desenvolvimento dos modelos de previsão.

⁶ http://www.scielo.org/php/index.php

⁷ http://pcs.isiknowledge.com

⁸ http://search.ebscohost.com

⁹ http://www.proquest.com

¹⁰ http://www.periodicos.capes.gov.br/

¹¹ http://www.dominiopublico.gov.br

¹² http://www.teses.usp.br

¹³ http://www.sciencedirect.com/

¹⁴ http://www.gale.cengage.com/

¹⁵ http://www.estecon.fea.usp.br/index.php/estecon/issue/archive

Os modelos de alisamento exponencial se originaram na década de 50 e anos 60 com os trabalhos de Brown (1959 *apud* Gooijer e Hyndman, 2006, p. 444). Mas foi somente 10 anos depois que Pegels (1969, p. 312) atribuiu uma classificação simples, definindo os padrões de tendência, sazonalidade definindo modelos aditivos (linear) e multiplicativos (não-linear).

Mais tarde, Snyder (1985 *apud* Gooijer e Hyndman, 2006, p. 445) revelou que os métodos de alisamento exponencial poderiam ser considerados como resultantes de um modelo de espaço de estado (ou seja, um modelo de uma única fonte de erro). Ainda segundo os autores, embora essa informação tenha passado desapercebida no momento, nos últimos anos forneceu a base de uma grande quantidade de trabalhos com modelos de espaço de estado.

Porém, foi Box, Jenkins e Reinsel (1994, p. 22) que forneceram condições de colocar os modelos de alisamento exponencial dentro de um quadro estatístico mais rigoroso ao demonstrar que algumas previsões feitas por tais modelos surgem como casos especiais dos modelos ARIMA.

Foi Yule (1927, p. 271) que, estudando o comportamento de séries deterministas, lançou o conceito de séries estocásticas por ter identificado toda série que poderia ser considerada um processo estocástico. Desde então, vários trabalhos são publicados na área de séries temporais que lidaram com estimação estocástica de parâmetros, identificação, verificação e previsão futura.

Essa percepção de se realizar previsão respeitando coerentemente a sequência de identificação, estimação e previsão foi a marca histórica deixada por Box e Jenkins (1994, p. 25) com o livro *Time Series Analisys: forecasting and control*, quando popularizaram os chamados modelos ARIMA. Esses modelos se expandiram por várias ciências, passando por estudos de natureza empírica e chegando às séries temporais financeiras, principalmente com a evolução do poder computacional.

O avanço e a criação de softwares cada vez mais potentes para análise de dados colaborou veementemente para o avanço das técnicas de análise de séries temporais. A análise de *outliers* e detecção de quebras estruturais são exemplos desse avanço no processo de análise.

Os estudos voltados para análise e previsão de séries temporais financeiras são usados para vários propósitos como estudo da estrutura dinâmica de uma série de informações estatísticas, para investigar a relação entre duas variáveis financeiras, ajustar comportamento sazonais de séries econômicas, melhorarem a qualidade de regressões quando os erros são serialmente correlacionados e gerar intervalos de previsão para volatilidade (TSAY, 2000, p. 638).

O sucesso dos modelos ARIMA gerou um volume considerável de trabalhos acadêmicos relativos à análise de séries temporais. De qualquer modo, a história dos modelos de séries temporais é dividida em pesquisa acerca do domínio do tempo e da frequência. São duas escolas bem distintas de modelos.

No domínio do tempo, usam-se as funções de autocorrelação dos dados e modelos paramétricos, tais como ARIMA, para descrever o comportamento da série. No domínio da frequência o foco é na análise espectral ou distribuição de potência para estudar suas aplicações. (TSAY, 2000, p. 639).

Ainda segundo este mesmo autor, estes avanços colaboraram também para o surgimento de novos modelos quando se passaram a detectar não linearidades nas séries temporais. Na literatura econométrica financeira evoluiu o conceito de variância condicional até o surgimento dos modelos da família ARCH e GARCH para a modelagem da volatilidade condicional heterocedástica, muito embora o início da análise de séries temporais não lineares foi atribuído a Volterra (1930, *apud* Gooijer e Hyndman, 2006, p. 433).

Gooijer e Hyndman (2006, p. 451) destacam com grande importância que a partir do início da década de 80, duas importantes técnicas avançadas de análise de sereis temporais tomaram corpo junto às pesquisas na área: os modelos de espaços de estados e o uso de filtros para redução de ruídos. Entre esses filtros, destacam-se as wavelets e o filtro de Kalman.

Afirmam ainda que mesmo sendo uma idéia recente, Kalman (1960 apud Gooither e Hyndman (2006, p. 452)) já havia divulgado esse poderoso instrumento de análise. A grande contribuição de Kalman foi deixar o algoritmo onde descreve o posicionamento de uma série temporal na estrutura de um modelo de espaço de estados inclusive descrevendo algoritmo de Kalman estendido para dados não lineares.

A detecção de comportamento não linear também trouxe para a área de séries temporais o uso das redes neurais artificiais. Para Haykin (2001, p. 29), o desenvolvimento das redes neurais ocorreu a partir dos trabalhos de McCulloch e Pitts (1943), Hebb (1949) e Rosemblatt (1958) que introduziram o primeiro modelo de redes neurais de auto-organização, e o modelo *perceptron* de aprendizado supervisionado, respectivamente.

Desde então, diversos modelos de redes e configurações foram adotadas e discutidas na literatura, entre eles as redes recorrentes e uso combinado de filtros para melhora da etapa de treinamento e desempenho das previsões com redes neurais (TSAY, 2000, p. 641).

Na linha do tempo pode-se destacar na figura a seguir os principais responsáveis pela evolução continuada dos estudos, envolvendo análise de séries de tempo e as técnicas destacadas em negrito sendo as usadas nesta pesquisa.

Cholesky Galileo Fisher Fermat Wierner Pascal Kolmogorov Huygens Itô Newton Schmidt Bernoulli Redes Swerling Neurais Riccati Kalman Stratnovich Piazzi Kailath Laplade Bucy Legendre Potter Gauss Bierman Maxwell Carlson

Figura 3.1: Linha do tempo dos principais contribuintes para análise de estimação de dados

Fonte: Adaptado de Grewal e Andrews (2008, p. 03)

Embora exista na literatura um consenso claro de que as previsões de modelos não lineares substancialmente superam os dos modelos lineares, existe espaço ainda para se trabalhar com os modelos multivariados e híbridos, que usam combinação de vários modelos mesmo para os problemas já resolvidos e os que ainda estão por vir no futuro (GOIJER e HYNDMAN, 2006, p. 462).

Diante destes fatos históricos, comenta-se a seguir os trabalhos relevantes que trouxeram grandes contribuições para a área e para o delineamento desta pesquisa.

A revisão dos textos empíricos aqui citados está organizada a partir dos objetivos de uso combinado dos filtros e modelos de previsão que são objeto de estudo desta tese. Para cada combinação possível entre uso de filtros e modelos de previsão, são discutidas as principais justificativas teóricas apresentadas por esses autores e apresentados os principais resultados encontrados.

3.2 Previsão com filtro de Kalman

No que diz respeito aos modelos combinados de previsão, também chamados modelos híbridos, Souza (2008, p. 4) estudou tais modelos de previsão de séries temporais de curto, médio e longo prazo, confrontando modelos lineares e não lineares. Por modelos não lineares híbridos, considerou o uso de redes neurais com função de base radial – RNs-RBF, com treinamento baseado no filtro de Kalman estendido, ou seja, na fase de treinamento, usou os dados filtrados pelo algoritmo de Kalman. Para os modelos lineares, considerou o modelo de Box, Jenkins e Reinsel (1994, p. 33).

O autor destaca o reduzido tempo de processamento com a utilização do filtro de Kalman como algoritmo de treinamento. Esta constatação é consistente com a hipótese de redução gradual de tempo elaborada e discutida em Haykin (2001, p. 837), devido ao fato do algoritmo seguir critérios de arquitetura recorrente o que facilita a convergência.

A base de dados utilizada foi para a *commoditie* soja com dados do CEPEA (Centro de Estudos Avançados em Economia Aplicada) e realizou previsões de curtíssimo prazo (um passo a frente), médio prazo (20 passos a frente) e longo prazo (687 passos a frente). Os resultados demonstraram que os erros foram menores para as operações de curtíssimo prazo, ficando instáveis para longo prazo. O melhor método foi as redes neurais com treinamento pelo filtro de Kalman.

Na mesma linha de uso das redes neurais para previsão, Oliveira (2007, p. 8) mostrou melhor desempenho para realizar previsões comparando modelos ARIMA-GARCH, redes neurais *feedforward* e redes neurais treinada com algoritmo de Kalman e aplicou a ações dos setores financeiro, alimentos, indústria e serviços.

Abrindo parênteses para explorar um pouco mais este trabalho sobre as séries analisadas e uso combinado do filtro de Kalman com redes neurais, o autor faz, a

exemplo também de Sabino e Bressan (2009, p. 9), todo o tratamento estatístico detalhado dos testes aplicados bem como descreve suas hipóteses para análise das séries financeiras estudadas. Os resultados apontaram que em todas as séries, os testes estatísticos aplicados (raiz unitária, normalidade pelo Jarque Bera e BDS para detectar presença de não lineariedade) mostraram o caminho para se usar os modelos ARIMA-GARCH e redes neurais. Utilizou critérios de análise das previsões como o coeficiente de Theil e a raiz do erro quadrático médio, além de construir intervalos de previsão para as previsões como fez Lima (2004, p. 164).

Voltando a discussão principal, para Oliveira (2007, p. 262), os resultados apontaram para um maior conservadorismo (maior amplitude) dos intervalos de confianças para as previsões realizadas pelas redes neurais do que para os modelos ARIMA-GARCH. Dos resultados obtidos acerca das previsões, as redes neurais treinadas com o filtro de Kalman apresentaram melhores resultados que as redes neurais feedforward e os modelos ARIMA-GARCH.

O autor revela ainda algumas limitações do uso do filtro de Kalman para aplicações com redes neurais que advém da necessidade de escolhas de parâmetros da transformação das equações *unscented* que calcula as estatística da variável *x* propagada através da função não linear *f*(.).

Singhall e Wu(1989, p. 1188) talvez tenham sido os primeiros a demonstrar a melhora do desempenho de uma rede neural supervisionada que utilizou o filtro de Kalman estendido.

Estes autores demonstraram que o seu algoritmo, embora de grande esforço computacional, convergia com menos iterações do que os tradicionais métodos de retroprogação. Após este trabalho, diversos outros autores fizeram simplificações e melhorias no algoritmo e diversificaram o seu uso em problemas ligados à área de engenharia, saúde e transporte como os artigos de Shah e Palmieri (1990, p. 42) Williams (1992, p. 244) e Puskorus e Feldcamp, (1994, p. 288), no Brasil também muito utilizado em pesquisas recentes como em Oliveira (2007, p. 80) e Pereira (2009, p. 103).

Corsini e Ribeiro (2008, p. 11) fizeram uso da aplicação do filtro de Kalman em modelos de apreçamento de *commoditites* agrícolas, utilizando os preços futuros para

definir o comportamento dos preços à vista do açúcar. Utilizou um modelo de estimação multivariado com a variável do preço do petróleo como explicativa e o açúcar como dependente.

Os autores trabalharam com 1157 observações diárias das variáveis e construiuse um modelo simplificado de suavização por média móvel para prever o preço do açúcar um passo a frente, usando o filtro de Kalman e sem o mesmo para realizar as previsões. Foi considerado o preço à vista do açúcar como variável não observável. Os resultados apontaram para uma melhora, passando de 6,8% para os modelos tradicionais para 6,2%, números esses medidos pelo MAPE.

Aiube (2005, p. 108) utilizou filtro de Kalman para estimação dos preços da commodities do petróleo. Segundo o autor, as commodities possuem características próprias como a de que seus preços são formados nos mercados futuros. Isso faz com que as variáveis de estados (estocásticas), muitas vezes, não são observáveis e necessitam de ser estimadas. O autor usa o chamado filtro de partículas para estimar o modelo desejado.

Segundo o autor, o filtro de partículas é um procedimento recursivo para integração, dentro da classe dos métodos sequenciais de Monte Carlo que dispensa as condições de linearidade e normalidade. Cabe ressaltar que esta metodologia exige um esforço computacional extremamente sofisticado.

A justificativa dada pelo autor pela opção do filtro de Kalman está na possibilidade de se fazer inferência sobre as variáveis consideradas não observáveis. Assim, dada uma distribuição de probabilidade da variável de estado e um modelo que descreva o processo estocástico das variáveis observadas, o filtro de Kalman gera distribuições um passo a frente para as variáveis de estado. Para a análise dos valores previstos o autor empregou o RMSE que é a raiz do erro quadrático médio. Essa medida é pouco detectada em trabalhos ligados a séries temporais financeiras.

Conclui a tese afirmando que o filtro de partículas é viável na prática e mostrou que o erro obtido é próximo daquele do filtro de Kalman para problemas cujas séries seguem uma distribuição normal e a estimação paramétrica é coerente com diversos trabalhos da literatura. Porém, fazendo uma análise empírica, o autor verificou que o

filtro de Kalman produz variáveis filtradas com um índice de erros de previsão menor do que os filtros de partículas, quando se compara com os dados reais.

3.3 Previsão com filtro de Kalman e wavelets

Em relação ao uso dos filtros de Kalman e de *wavelets* para o processo de filtragem Postalcioglu, Erikan e Bolat (2005, p. 951) afirmam que o filtro de Kalman remove distúrbios ou falhas de uma série de tempo (ou um sinal), usando a inicialização e transmissão das estatísticas de covariância de erro. Comenta ainda que a aplicação do filtro de Kalman se torna impraticável em modelos de grande escala, como demonstrou para o sistema oscilador. A alternativa para esse tipo de sistema é o filtro de *wavelet*. Em sua pesquisa usou a *wavelet* de *Coiflet* 2¹⁶, com 9 níveis de decomposição e mostrou que a resposta do filtro *wavelet* é melhor quando comparado com o resultado do filtro de Kalman no que se refere a filtragem da supressão do ruído chamada *denoising*. Embora o autor não tenha feito este trabalho com intuito de previsão de dados, fica claro o uso da melhora que as técnicas podem trazer no caso de supressão de ruído de séries temporais.

Em recente publicação Xie, Zhang e Ye (2007, p. 326) investigam a aplicação do filtro de Kalman com o uso da transformada discreta de *wavelet* para realização de previsão de curto prazo para uma série temporal do volume de tráfego obtido de certa região em um intervalo de cinco minutos. Os autores usam a transformada de *wavelet* para separação dos dados em alta e baixa frequência e usa o filtro de Kalman conforme sugerido por Welch e Bishop (2006, p. 3), para extrair os ruídos dos dados das séries decompostas a fim de melhorar a qualidade das previsões. Usou dois tipos de *wavelets*: a *wavelet* de Daubeschies¹⁷ número quatro e a *mother wavelet* de Haar e comparou os resultados usando apenas o filtro de Kalman para prever.

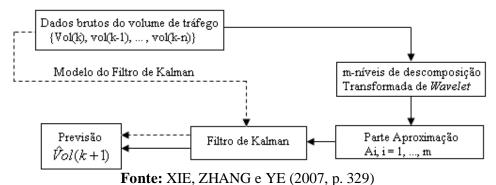
Os autores usaram a decomposição por *wavelet* em dois níveis de dados. Após a decomposição via transformada de *wavelet*, a parte da série *denoised*, chamada aproximação, é fornecida como *input* para o filtro de Kalman descartando o ruído. Os

¹⁷ Para verificar a forma das wavelets consulte o Apêndice A.

¹⁶ Para verificar a formas das wavelets consulte o Apêndice A.

autores elaboraram o seguinte esquema de previsão usando filtro de Kalman e de wavelets conforme descrito na figura a seguir.

Figura 3.2: Modelo de Kalman e filtro de Kalman-*Wavelet* para previsão de XIE, ZHANG e YE(2007)



Os resultados apontaram que o uso combinado, fazendo a decomposição por wavelets e posteriormente aplicando o filtro de Kalman para denoising, supera a aplicação direta do filtro de Kalman para previsão em termos do erro absoluto percentual médio e pela raiz do erro quadrático médio.

Apontaram ainda que a *mother wavelet* de Daubesch 4 obteve melhor desempenho no sentido de reduzir o erro absoluto percentual médio das previsões. Justificam esse resultado com os mesmos argumentos de Zandonade e Morettin (2003, p. 210) em que provavelmente a *wavelet* de Daubesch 4 realiza melhor a supressão/separação do ruído melhor que a *wavelet* de Haar por preservar o perfil do comportamento dos dados brutos.

De qualquer forma, é um resultado interessante por fazer uso de *wavelets* primárias e não outras possíveis combinações de padrões de dilatação e translação das formas de ondas.

3.4 Previsão com filtro de wavelets

Na linha de previsão de séries temporais com uso de *wavelets* destacam-se os trabalhos internacionais de Granger (1992, p. 3), Tak (1995, p. 43), Ariño (1995), Ukil e

Zivanovic (2001, p. 103), Ma, Wong e Sankar (2004, p. 5824) e Aminghafari (2007, p. 715).

Tak (1995, p. 43) realizou previsões de curto prazo para o *Standard and Poor's* 500 (S&P 500), com dados diários da série temporal financeira no período compreendido entre maio de 1980 a dezembro de 1990, baseado na teoria de segregação da série temporal decomposta por *wavelets*.

O autor adotou dois subníveis para aplicação da transformada de *wavelet*, e comparou as previsões dos modelos ARIMA de redes neurais *feedforward*. Buscou atestar se a decomposição via *wavelets* traria ou não melhora na qualidade das previsões comparativamente com os modelos tradicionais sem decomposição.

Já Ma, Wong e Sankar (2004, p. 5824) também realizaram previsões com modelos ARIMA-GARCH e usaram decomposição de *wavelets* para o S&P100 e obtiveram redução dos erros para previsão em 3% com uso de *wavelets* primárias de Haar. Os resultados de Tak (1995, p. 44) foram relativamente piores com a utilização das *wavelets* de Morlet's e chapéu mexicano¹⁸ e comprovou que a filtragem por *wavelets* reduziu o MAPE em apenas 9%.

Já Ariño (1995) analisou a previsão de vendas de automóveis com modelos ARIMA e decomposição de *wavelets* e conseguiu redução de 4% no erro quadrático médio utilizando *wavelet* de daubechies numero 8, com previsões nas duas sub-séries em um nível. Resultados semelhantes foram encontrados em AMINGHAFARI (2007, p. 715) onde contou com a decomposição de *wavelets* por Daubechies número 7 em conjunto com modelos ARIMA, o menor erro ficou com a decomposição via *wavelets*, contra 2,0% medidos pelo MAPE para séries de varejo.

No Brasil, destacam-se os trabalhos de previsão de séries temporais de Chiann (1997, p. 32), Homsy, Portugal e Araújo (2000, p. 10), Zandonade e Morettin (2003, p. 205), Lima (2004, p. 133), e Rocha (2008, p. 120) que utilizaram a decomposição de *wavelets* com objetivo de melhora na qualidade do tratamento das séries temporais. Dos autores citados acima, Chiann (1997, p. 31) e Homsy, Portugal e Araujo (2000, p. 10), trabalharam com análise de *wavelets* em séries temporais não financeiras.

_

¹⁸ Para verificar o formato das ondas consulte Apêndice A.

Lima (2004, p. 133) aplicou um método de decomposição de *wavelets* para o comportamento do Ibovespa em conjunto com modelos ARIMA GARCH e redes neurais recorrentes e conseguiu uma redução do erro de 3% para previsões de 21 passos a frente com destaque para as redes neurais e formas de onda de Deubechies numero 1.

Na mesma linha, Rocha (2008, p. 121) citando Lima (2004, p.155), conseguiu bons resultados combinando modelos de previsão de suavização exponencial e modelo ARIMA reduzindo o erro em 7,08% medido pelo MAPE. Utilizou a função de onda de Daubechies número 2.

Nos diversos trabalhos analisados, nota-se que o uso de filtros se mostrou de grande aplicação interdisciplinar, o que, em parte, explica a forte presença das *wavelets* nos trabalhos pesquisados nos últimos anos. Essas conclusões enriquecem a relevância da adoção de tais ferramentas de filtros à previsão de séries financeiras aplicadas ao mercado brasileiro. Também corroboram para a idéia de que se pretende obter uma melhora na qualidade das previsões, provendo o investidor de dados para que possa tomar sua decisão e posicionamento no mercado futuro.

Neste capítulo, não se teve a pretensão de se fazer uma revisão ampla de todos os inúmeros trabalhos publicados acerca dos modelos econométricos existentes para previsão de séries temporais. Muito menos trazer à tona toda uma discussão sobre quais modelos são melhores ou piores. Optou-se pelo foco no uso combinado das técnicas de filtragem em conjunto com os modelos de previsão. Diversos trabalhos foram deixados de lado, pela opção do autor em trazer para a discussão apenas aqueles necessários para a discussão que tiveram uma ligação direta com os objetivos e proposta desta pesquisa.

A partir das pesquisas aqui descritas e discutidas, assentaram-se trabalhos empíricos e aqueles essencialmente práticos, de áreas próximas à área de previsão de séries temporais financeiras e outras mais distantes como engenharia, eletrônica e física. Todavia, todos tiveram o claro propósito de trazer a confirmação da hipótese formulada de que o uso combinado das técnicas de filtragem e previsão pode trazer benefícios aos resultados finais.

Dessa forma, passa-se a revelar a seguir a metodologia empregada para a discussão desta pesquisa, apoiada nos diversos trabalhos aqui discutidos.

4. METODOLOGIA

Nesta seção são apresentados a metodologia desenvolvida nesta pesquisa que envolve a classificação da pesquisa, os métodos empregados, os procedimentos para análise, previsão e avaliação dos resultados e as técnicas utilizadas como os filtros de volatilidade.

O principal objetivo desta pesquisa é avançar no entendimento do uso combinado dos filtros de volatilidade com a perspectiva de redução de ruídos e consequente queda do erro de previsão.

A literatura indiretamente incentiva o desenvolvimento desta pesquisa à medida que, de um lado, tem-se mostrado inconclusiva sobre a questão da contribuição do uso combinado de técnicas de previsão e, por outro lado, tem indicado a necessidade de se aprofundar nas minúcias do uso dos diversos tipos de filtros na análise e na previsão de séries temporais financeiras.

4.1 Natureza da Pesquisa

Do ponto de vista metodológico, segundo Beuren (2006, p. 80), esta pesquisa pode ser classificada como exploratória, descritiva, bibliográfica e quantitativa de acordo com cada abordagem descrita.

Quanto aos seus objetivos, é uma pesquisa exploratória e descritiva. É exploratória porque segundo a autora procura desenvolver uma visão geral dos grandes temas emplacados no trabalho de modo a tornar claro para formulação de suas hipóteses de pesquisa. É também descritivo porque busca descrever as principais características presentes numa série temporal como grupamento de volatilidade, medidas estatísticas do comportamento da normalidade e não lineariedade que permitam construir e descrever o modelo que rege o seu comportamento.

Do ponto de vista dos seus procedimentos adotados para checagem do construto da tese, fez-se necessário uma pesquisa bibliográfica para dar suporte teórico e visualização do estado da pesquisa que combina uso de filtros de volatilidade e modelos que permitem melhor adequação a séries de tempo com características próprias como é o caso das séries financeiras.

A maior parte das referências são internacionais aplicadas as mais diferentes séries em distintos campos do conhecimento, particularmente explorando com maior rigor de detalhes àquelas voltadas ao campo das séries temporais financeiras, cujo objetivo é conhecer melhor os trabalhos e a forma como foram explorados para melhor escolha dos modelos aplicados para os dados brasileiros.

As fontes desses dados foram as bases internacionais do acervo acadêmico e também bases nacionais para averiguar o estado das pesquisas no Brasil relativas ao campo em que lhe cabe. Tais bases foram citadas no capítulo sobre revisão da literatura. Foram levantadas teses, dissertações tanto nacionais como internacionais, *papers* dos mais renomados *journals* que são referência à área.

Não se teve a pretensão de se chegar a uma revisão completa de todos os inúmeros trabalhos publicados acerca dos modelos aqui desenvolvidos, mas buscou-se chegar o mais próximo do estado da arte que foi possível para as técnicas de previsão de séries temporais e seus filtros que são aplicáveis.

Optou-se pelo foco no uso combinado das técnicas de uso encadeado de filtros que foram usados para algum tipo de previsão. Diversos trabalhos foram deixados de lado, pela opção do autor de trazer para a discussão apenas aqueles que tiveram uma ligação direta com os objetivos e proposta desta pesquisa.

No que diz respeito a abordagem do problema, é uma pesquisa quantitativa de base econométrica, caracterizado pela forte presença de modelos de origem e desenvolvimento matemático e estatístico. As análises contaram com ferramentas e softwares específicos da área, conforme será descrito a seguir.

4.2 Operacionalização da pesquisa

Para consecução da pesquisa, optou-se por uso de softwares onde cada um teve o seu devido papel. O uso de mais de um software fez-se necessário por não existir ainda na comunidade acadêmica um software que enquadre todas as etapas descritas nesta pesquisa.

A figura 4.1 descreve todo o procedimento operacional do tratamento e da análise dos dados através de um fluxograma cujo objetivo foi elucidar cada uma das etapas a ser seguida.

Inicialmente, forma-se a série dos log retornos através da expressão (2) e analisa-se as estatísticas descritivas, buscando verificar os fatos estilizados na série temporal financeira como assimetria, curtose, grupamento de *clusters* de volatilidade, normalidade etc.

Aplica-se então o teste de normalidade, dado pela expressão (3) e testa-se a estacionariedade da série, uma vez que os dados devem estar estacionarizados. Caso esta hipótese seja violada, aplicam-se diferenças sucessivas na série temporal para torná-la estacionaria e aplica-se novamente o teste até que o mesmo seja verificado.

A partir da série estacionária, verifica-se a presença de autocorrelações seriais nos dados, aplicando o teste de independência BDS dado pela expressão (11). Se independente, a série temporal pode representar um comportamento de passeio aleatório ou caótico. Estes passos não serão verificados por não ser objeto desta pesquisa.

Obtido o comportamento dependente da série temporal, verifica-se o tipo de dependência, se linear ou não linear através do teste de McLeod-Li. Se a série apresentar comportamento linear, poderão ser extraídos modelos lineares. Caso contrário, tem-se que detectar qual o nível de não lineariedade, se na média, na variância ou total, onde os melhores modelos são os de redes neurais.

O próximo passo seria então realizar as previsões com modelos ARIMA-GARCH e com redes neurais recorrentes para a série dos log retornos puros, ou seja, sem nenhum processo de filtragem nos dados. No caso das redes neurais, utilizaram-se as redes recorrentes proposta por Williams e Zipser (1989), conforme já implementado em Lima (2004).

Como se tem a série temporal financeira $\left\{y_{t}\right\}_{t=1}^{n}=\left\{y_{1},y_{2},...,y_{t},...,y_{n}\right\}$ que pode ser um índice de bolsa, preços de ações, cotações de moedas ou *commoditites* etc, e a partir desta série, obtém os retornos pela expressão $R_{t}=\ln\left(\frac{y_{t}}{y_{t-1}}\right)$ que fornece os n valores dos retornos.

Após feitas as previsões
$$\left\{R_{t}\right\}_{t=1}^{n+t} = \left\{R_{1}, R_{2}, \dots, R_{t}, \dots, R_{n}, \underbrace{R_{n+1}, R_{n+2}, \dots, R_{n+t}}_{previsões}\right\}$$
 dos

retornos, deve-se voltar à série para o nível bruto de valores.

Assim, dado o valor previsto R_{n+1} , que representa $R_{n+1} = \ln\left(\frac{y_{n+1}}{y_n}\right)$, desenvolvese a expressão para se chegar ao valor de y_{n+1} , aplicando-se o processo inverso da logaritmização que é a exponenciação:

$$R_{n+1} = \ln\left(\frac{y_{n+1}}{y_n}\right) \tag{97}$$

$$e^{R_{n+1}} = e^{\ln\left(\frac{y_{n+1}}{y_n}\right)} \tag{98}$$

De onde se obtém:

$$e^{R_{n+1}} = \frac{y_{n+1}}{y_n} \tag{99}$$

$$y_{n+1} = y_n \times e^{R_{n+1}} \tag{100}$$

Chegando-se assim ao valor previsto que pode ser aplicado ao modelo ARIMA-GARCH.

Em se tratando dos dados inseridos na rede neural recorrente, foi utilizado o processo de normalização proposto por Azoff (1994), dado pela expressão a seguir:

$$Z_{t} = \frac{R_{t} - \min\left\{R_{t}\right\}_{t=1}^{N}}{\max\left\{R_{t}\right\}_{t=1}^{N} - \min\left\{R_{t}\right\}_{t=1}^{N}} \in [0;1]$$
(101)

Dessa forma, ao obterem-se os valores previstos para Z_t , onde

$$\left\{Z_{t}\right\}_{t=1}^{n+t} = \left\{Z_{1}, Z_{2}, \dots, Z_{t}, \dots, Z_{n}, \underbrace{Z_{n+1}, Z_{n+2}, \dots, Z_{n+t}}_{previs\tilde{o}es}\right\} \text{ pode-se ter:}$$

$$Z_{t+1} = \frac{R_{t+1} - \min(R_t)}{\max(R_t) - \min(R_t)}$$
(102)

$$R_{t+1} = Z_{t+1} \left[\max(R_t) - \min(R_t) \right] + \min(R_t)$$
(103)

Para que sejam obtidos os valores da série no nível bruto (y_t) , é necessário realizar uma transformação dos dados resultantes da análise da série dos retornos que segue:

$$\ln\left(y_{t+1}\right) - \ln\left(y_t\right) = \left\{Z_{t+1}\left\lceil \max\left(R_t\right) - \min\left(R_t\right)\right\rceil + \min\left(R_t\right)\right\}$$
(104)

$$\ln\left(y_{t+1}\right) = \left\{Z_{t+1}\left[\max\left(R_{t}\right) - \min\left(R_{t}\right)\right] + \min\left(R_{t}\right)\right\} + \ln\left(y_{t}\right)$$
(105)

Aplicando exponencial na expressão acima, chega à reconstrução da série prevista no nível original.

$$y_{t+1} = e^{\left\{Z_{t+1}\left[\max(R_t) - \min(R_t)\right] + \min(R_t)\right\} + \ln(y_t)}$$
(106)

Destaca-se aqui de antemão que o objetivo desta parte é tratar das ferramentas necessárias para suportar a discussão proposta ao longo da tese, sem pretender esgotar o que há de melhor nas boas publicações acadêmicas, sejam elas pioneiras ou mais recentes.

Assim, os testes empregados na sequencia descrita a seguir foram aqueles verificados na literatura como os mais presentes e considerados essenciais para análise e previsão de séries temporais.

Em suma, a discussão presente reforça a importância do roteiro aqui descrito no que diz respeito a contribuição da pesquisa para área de estimação de previsões futuras para uma tomada de decisão consciente em mercados futuros.

Série Temporal Finaneira Série dos Log-Retomos Estatística Descritiva e Teste de Normalidade Diferencia-se a série Teste de Estacionariedade Não Estacionária Se Estacionária Se Independente Teste Independência (BDS) Se Dependente Passeio Aleatório Se Normal Teste Normalidade Verificar Tipo de Dependência Teste de McLeod-Li Teoria do Caos Se não Normal Linear Não Linear Modelos Lineares Verificar Tipo de Não Lineariedade Teste de Hsieh Média Variância Total GARCH Redes Neurais Sem filtro Filtros Previsão Wavelets Kalman Kalman Wavelets Previsão Previsão Qualidade da Previsão

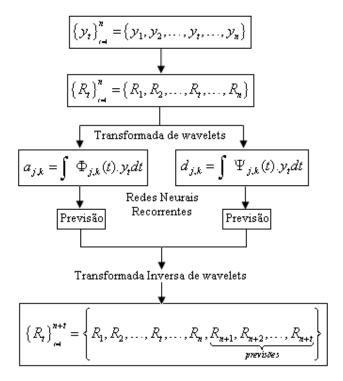
Figura 4.1: Fluxograma descritivo das etapas adotadas na pesquisa

A partir daí entra a utilização dos filtros.

4.2.1 Previsão com filtro de *wavelets* e redes neurais recorrentes

No caso da aplicação do filtro de *wavelets* tem-se para um nível de decomposição o esquema operacional seguido:

Figura 4.2: Fluxograma descritivo para previsão do redes neurais e filtro de wavelets



4.2.2 Previsão com filtro de Kalman

A partir das séries de retornos, aplica-se o filtro de Kalman estendido, uma vez que já foi verificada a presença de não lineariedade da série, deve-se fazer as previsões chegando-se a série abaixo:

$$\left\{R_{t}\right\}_{t=1}^{n+t} = \left\{R_{1}, R_{2}, \dots, R_{t}, \dots, R_{n}, \underbrace{EKFR_{n+1}, EKFR_{n+2}, \dots, EKFR_{n+t}}_{previsões}\right\}$$
(107)

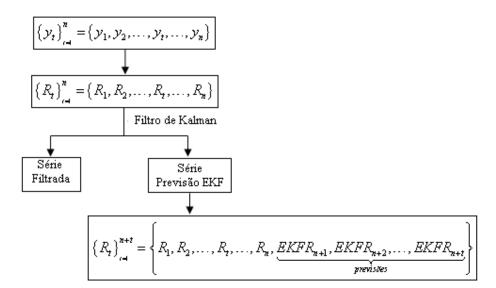


Figura 4.3: Fluxograma descritivo para previsão com filtro de Kalman

Lembrando que o filtro de Kalman quando aplicado sobre uma série de dados, solta dois tipos de informação. O filtro *denoised* e a série prevista. Nesta etapa de previsão, considera-se apenas a série prevista.

4.2.3 Previsão com filtro de Kalman e de wavelets

Para realizar este tipo de previsão, adota-se primeiramente a aplicação do filtro de Kalman sobre a série dos dados. Posteriormente, aplica-se sobre a série filtrada por Kalman, o filtro de *wavelets*.

Após aplicada a decomposição pela transformada de *wavelets*, realiza-se a previsão utilizando-se as redes neurais recorrentes. Nesta etapa, assim como na previsão feita somente com as redes neurais, deve-se escolher a forma de onda.

A estrutura desta aplicação conjunta dos filtros e das previsões está descrita na figura 4.4 a seguir.

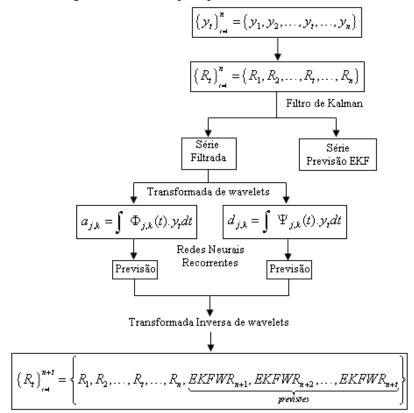
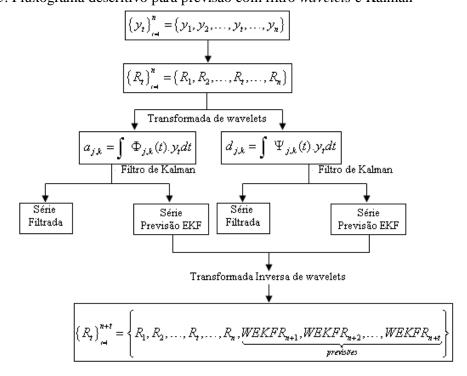


Figura 4.4: Fluxograma descritivo para previsão com filtro de Kalman e wavelets

4.2.4 Previsão com filtro de wavelets e de Kalman

Figura 4.5: Fluxograma descritivo para previsão com filtro wavelets e Kalman



Nesta pesquisa, considerou-se ainda a aplicação reversa, ou seja, decompões-se primeiro usando o filtro de *wavelets* e posteriormente o de Kalman. A operacionalização desta aplicação conjunta dos filtros está descrita na figura 4.5 acima.

4.2.5 Ferramentas de análise

O quadro 4.1 abaixo demonstra as ferramentas utilizadas para confecção da pesquisa. Todas as etapas descritas acima foram conduzidas através da utilização de softwares como MATLAB® 6.5, EVIEWS ® 4. A seguir, descrevem-se cada um dos processos utilizados e a ferramenta necessária para se chegar aos resultados.

Quadro 4.1: Ferramentas utilizadas em cada etapa da pesquisa.

Etapa	Ferramenta Empregada		
Teste de Normalidade	EVIEWS ® 4 através do Menu QUICK, SÈRIES		
	STATISTICS; HISTOGRAM and STATS obtém o		
	histograma com curva normal e teste de JB;		
Teste de Estacionariedade	Gerado no EVIEWS ® 4 através do Menu QUICK,		
	ESTIMATE EQUATION, onde se gera um modelo		
	AR(1);		
Teste de Independência (BDS)	Gerado a partir do histograma feito no teste de		
	normalidade e no menu VIEW, BDS		
	INDEPENDENCE TEST;		
Teste de McLeod Li	Faz-se a análise do correlograma e a partir do modelo		
	AR(1) gerado aplica-se a partir do menu VIEW –		
	RESIDUAL TESTS – CORRELOGRAM SQUARED		
	RESIDUALS		
Redes Neurais Recorrentes	Algoritmo em MATLAB desenvolvido por Oliveira		
	(2003) e já testado em Lima (2004).		
Transformada de wavelets	Desenvolvido pelo autor através dos comandos do		
	MATLAB ® 6.5		
Filtro de Kalman	Desenvolvido pelo autor através dos comandos do		
	MATLAB ® 6.5		
Estatísticas de previsão	Desenvolvido pelo autor através dos comandos do		
	MATLAB ® 6.5		

4.3 Estatísticas de erro e intervalo de confiança para as previsões

As estatísticas de acuraria das previsões são calculadas conforme Gooijer e Hyndman (2006, p. 457) que levantaram as principais medidas de erros nos trabalhos de

previsão de séries temporais. O termo "acurácia" designa "o quão bem ajustado" é o modelo, ou seja, "quanto" o modelo é capaz de reproduzir os dados já conhecidos.

Dentre as diversas medidas possíveis para analisar a eficiência de um modelo de séries temporais, citam-se, a seguir, as principais medidas de desempenho conforme os autores citados acima.

Todos os modelos descritos nesta tese soltam como resultado final um valor esperado para a previsão futura. Nomeando $E_t[y_{t+1}]$ como a esperança condicionada de y_{t+1} dada a informação até período t (inclusive), sendo esta medida o estimador ótimo pelo critério do erro quadrático médio condicionado.

Segundo Enders (2004, p. 21), define-se o erro de previsão j passos à frente, a partir de t - $e_t(j)$ como sendo a diferença entre a medida y_{t+j} e o previsor esperado condicionado:

$$e_t(j) \equiv y_{t+j} - E_t \left[y_{t+j} \right] \tag{108}$$

Logo, o erro da previsão para um passo à frente é: $e_t(1) \equiv y_{t+1} - E_t \left[y_{t+1} \right] = \varepsilon_{t+1}$, isto é, a parte "não prevista" do valor histórico y_{t+1} , conhecendo-se a informação contemporânea medida no período t e o previsor de y_t .

Como se tem uma série temporal com n períodos de tempo, logo ter-se-ão n termos de erro, e então se pode calcular para esta pesquisa:

• Erro Percentual Absoluto Médio (Mean Absolute Percentage Error) (MAPE): é um valor absoluto médio em percentual, para se verificar a margem de acerto em comparação com o valor previsto. É mais adequado para comparação entre modelos. Quanto menor for o seu valor, melhor é o ajuste do modelo.

$$MAPE_{h} = \frac{\sum_{i=1}^{h} \left| \frac{\hat{\mathcal{E}}_{i}}{y_{i}} \right|}{h} \times 100\% \qquad MAPE \ge 0$$
(109)

r: Coeficiente de Correlação linear de Pearson. O coeficiente de correlação linear r mede o grau de relacionamento linear entre os valores emparelhados x e y em uma amostra. O coeficiente de correlação linear é chamado, às vezes, coeficiente de correlação momento-produto de Pearson. Varia entre -1 e 1, sendo estes extremos indicação de associação linear negativa e positiva perfeita, respectivamente.

$$r_h = \frac{\sum_{i=1}^h (y_i - \overline{y})(\hat{y}_i - \overline{\hat{y}})}{S_y S_{\hat{y}}} \in [-1;1]$$

$$(110)$$

em que \hat{y} é a esperança condicionada de y_{t+1} dada a informação até período t (inclusive).

• TIC – Coeficiente de Desigualdade de Theil (*Theil Inequality Coefficient*): Este coeficiente sempre estará entre zero e um, sendo que zero indica um ajuste perfeito.

$$TIC_{h} = \frac{\sqrt{\sum_{i=1}^{h} \hat{\varepsilon}_{i}^{2}}}{\sqrt{\sum_{i=1}^{h} \hat{y}_{i}^{2} + \sqrt{\sum_{i=1}^{h} y_{i}^{2}}}} \in [0,1]$$
(111)

Segundo Ahlburg (1992, p. 99), a escolha da medida de análise dos erros de previsão não pode ser baseada no gosto pessoal do pesquisador nem a popularidade estatística do indicador. Segundo o autor, a contribuição do indicador a ser usado deve ser vista de acordo com o tamanho da amostra a ser prevista. Para pequenos passos a frente na previsão, deve-se usar o MAPE e para grandes passos o RMSE dentro da pesquisa em que levantou e atestou que poucos trabalhos apresentam alguma justificativa do método de comparação das previsões.

Para todas as previsões feitas, serão considerados ainda os intervalos de confianças centrados na previsão feita para a esperança condicionada dos valores.

O intervalo de confiança para as previsões será o sugerido por Prankratz e Dudley (1987, p. 245) definido para y_{t+h} , com nível de confiança de 95% como:

$$\hat{y}_t(h) \pm 1,96\sqrt{Var_y[h]} \tag{112}$$

onde $Var_y[h] = Variância[\widehat{e_t(h)}]$, sendo que, $e_t(h) \equiv y_{t+h} - E_t[y_{t+h}]$, conforme define Enders (2004, p. 25).

Passa-se a seguir a definir a arquitetura da rede neural utilizada para as previsões.

4.4 Arquitetura da rede neural recorrente

Para as redes neurais, segundo Haykin (2001, p. 123), define-se a sua classificação de acordo com a arquitetura empregada. A mais simples de todas, segundo o autor, é a rede neural com alimentação adiante ou *feedforward*. Tais redes constituem-se por múltiplas camadas, onde a entrada de cada camada é a saída da camada anterior e a interconexão dos neurônios é acíclica.

Outra modalidade são as redes recorrentes que usam ciclos para treinamento. As três redes recorrentes mais utilizadas, segundo ainda Gooijer e Hyndman (2006, p. 452) são a rede de Jordan, a rede de Elman e a rede de Elman estendida.

O que diferencia uma da outra é que a rede na rede de Jordan, a realimentação ocorre apenas no sentido dos nós de saída para os nós de entrada, ao passo que na rede de Elman a realimentação sai apenas dos nós das camadas escondidas. Na rede de Elman estendida, a realimentação sai dos nós da camada escondida e também dos nós (ou neurônios) de saída. As diversas arquiteturas podem ser vistas em Haykin (2001, p. 150).

Para esta pesquisa, usa-se as redes neurais recorrentes de Elman com quatro neurônios na camada intermediária e função de ativação logística com 200 épocas de treinamento e epsilon de convergência de 0,001 e taxa de aprendizagem 0,005. Diversas outras configurações da rede foram testadas conforme sugestão de Oliveira (2003, p. 166), mas a descrita neste parágrafo foi a que apresentou melhor desempenho na capacidade de aprendizagem, verificação e previsão.

Para a separação dos valores adotou-se os valores sugeridos conforme os trabalhos de William e Zipser (1989, p. 285), com 60% dos dados para treinamento, 20% para validação e 20% para teste.

A seguir, desenvolve-se toda a aplicação prática descrita até aqui.

5. RESULTADOS EXPERIMENTAIS

Nesta seção são apresentados os resultados das análises comparativas de diversos modelos de previsão e combinação de filtros aplicados a séries temporais financeiras.

5.1 Análise da série do Ibovespa

O Índice da Bolsa de Valores de São Paulo (BM&FBOVESPA), chamado também de IBOVESPA, é considerado o termômetro do mercado de capitas brasileiro. É o mais importante indicador de desempenho médio das ações negociadas na bolsa brasileira. Dois fatores justificam seu uso no mercado: ele retrata a movimentação média dos negócios realizados com as ações mais negociadas na BM&FBOVESPA e pela sua tradição, já que manteve a integridade da sua série histórica pelo fato de não sofrer modificações em sua metodologia desde a sua implantação em 02/01/1968.

Em outras palavras, o IBOVESPA representa o valor atual, em moeda corrente, de uma carteira teórica de ações inicialmente definida em sua implantação com valor de 100 pontos, não sofrendo nenhum outro tipo de novos investimentos ou resgates. Dessa forma, o índice reflete não somente as variações dos preços das ações, mas também está ajustado pelo impacto da distribuição dos proventos.

Em termos de liquidez, o IBOVESPA compreende mais de 80% do número de negócios e do volume financeiro verificados no mercado à vista (lote-padrão) da BM&FBOVESPA. Em termos de capitalização bursátil, as ações das empresas integrantes da carteira teórica do IBOVESPA são responsáveis, em média, por aproximadamente 70% do somatório da capitalização bursátil de todas as empresas com ações negociáveis na BM&FBOVESPA.

Além disso, o IBOVESPA é empregado também nos contratos de derivativos financeiros de seu próprio índice. Sua cotação de fechamento em pontos de índice é negociada amplamente em contratos cheios e mini a futuro, demonstrando uma tendência do mercado como parâmetro para negócios futuros.

A série do IBOVESPA foi tomada no período de 03/01/2000 quando representava 16.930 pontos até 30/12/2009, quando apresentou 68.588 pontos. Nesse período, a valorização foi de 305,13%. As cotações foram consideradas de fechamento no período diário num total de 2477 observações.

Mar/06 - Inflação Argentina 80000 Mar/00 - Estouro Set/08 - Quebra Lehman Brothers da bolha da Nasdaq 70000 60000 Set/01 - Atentado WTC Nov/01 - Crise Argentina 50000 Out/09 - Rio 2016 40000 Jul/02 – Risco Lula Out/02 - Dólar R\$ 4,00 Abr/08 -30000 Investment Grade20000 Mar/03 - Ataque ao Iraque 10000 0 311/2003 311/2003 311/2004 311/2005 311/2005 311/2008 311/2008

Figura 5.1: Série temporal nominal do IBOVESPA diário

Fonte: BM&FBOVESPA, Sistema Enfoque

A figura 5.1 destaca o gráfico do IBOVESPA com suas cotações de fechamento e alguns destaques de fatos mundialmente relevantes que tiveram de alguma forma influência no comportamento na volatilidade do índice.

Como se nota, a série do IBOVESPA é fortemente volátil e que, constantemente, sofre a influência de variáveis exógenas como crises, boatos, que não estão ligados diretamente à sua estrutura, mas que influenciam na oscilação de seus retornos. A figura a seguir destaca os log retornos.

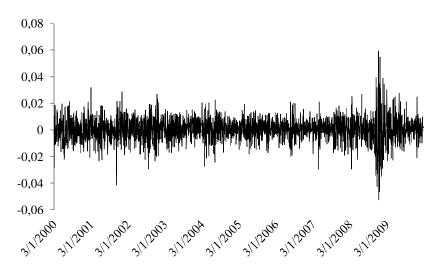


Figura 5.2: Série temporal dos log retornos do IBOVESPA diário

Observa-se que os retornos diários oscilam ao redor de zero apresentando uma variabilidade que depende do tempo, chamada volatilidade, com períodos de alta e de baixa variabilidade e dias em que o retorno é um valor anormal, chamado *outlier*. Além disso, nota-se diversos *clusters* de volatilidade que acontecem devido as incertezas do mercado provocadas por fenômenos econômicos e sociais como os destacados na figura 5.1.

Esse resultado está coerente com o observado por Mandelbrot e Hudson (2004, p. 115) e amplamente confirmado na literatura de que a autocorrelação positiva da volatilidade indica que ela se apresenta em *clusters*.

Tais fatos são comumente chamados de estilizados que provocam na série um comportamento não linear, já que responde de maneira diferente a impactos grandes e pequenos, ou negativos e positivos.

A figura a seguir, destaca a inspeção visual do histograma e do gráfico da densidade de probabilidade dos log retornos.

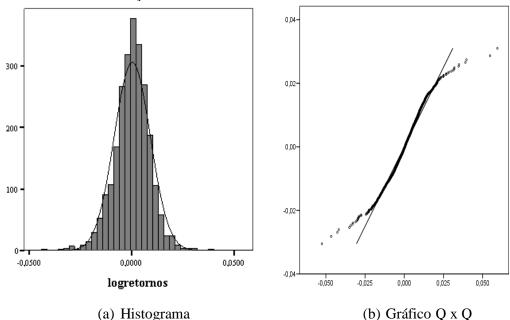


Figura 5.3: Histograma dos log retornos diários do IBOVESPA diário e retornos diários contra distribuição normal

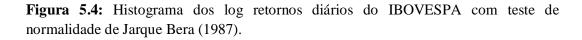
Nota-se que o histograma tem a sua parte central mais alta do que uma normal e há presença de valores bem afastados da tendência central dos dados. Essas informações são características das séries temporais financeiras e são descritos por comportamento leptocúrtico com caudas mais pesadas que a normal pelo fato dos valores se afastarem da média a vários múltiplos do desvio padrão.

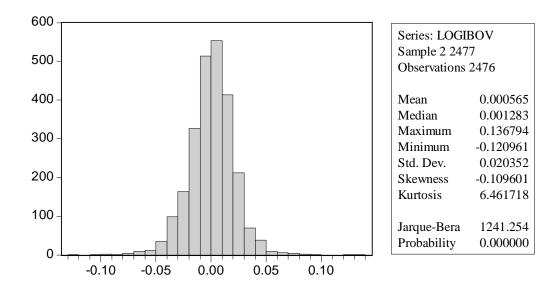
Na figura 5.3, observa-se ainda no item (b) o gráfico Q x Q, que é um gráfico que traz no eixo das abscissas os quantis da normal padrão e no outro os quantos dos dados. Se os dados tendessem a apresentar uma distribuição normal, os pontos estariam sobre a reta, o que não acontece no caso do IBOVESPA no período analisado.

O passo seguinte descreve o teste de normalidade dos log retornos de Jarque Bera (1987). As hipóteses para este teste são definidas como:

 H_0 : a série segue uma distribuição normal

 H_1 : a série não segue uma distribuição normal





A estatística de teste indica o valor de JB = 1241,254 com p-valor igual a zero. O nível de confiança adotado foi de 95%, sendo sua probabilidade de significância inferior a 5%, indicando a rejeição da hipótese nula que revela que a série não segue uma distribuição normal. Ressalta-se que a mesma hipótese também seria rejeitada a ao nível de significância de 1%.

As estatísticas descritivas para a série do IBOVESPA apontam para uma média próximo de zero, que vai de acordo com a teoria financeira clássica de que a média dos retornos de um ativo financeiro é sempre zero. O desvio padrão incondicional retrata as oscilações média dos log retornos. O excesso de curtose com valores acima de 3 é um dos principais fatores que pode ter levado a rejeição da hipótese nula de normalidade.

Segundo Christoffersen e Diebold (2000, p. 15), uma vez que a distribuição incondicional dos retornos não segue uma distribuição normal de probabilidade, a estimação dos retornos médios por máxima verossimilhança fornece os parâmetros para a média e a variância que convergirão para os verdadeiros parâmetros quanto mais elementos contiver a amostra de dados.

A não normalidade da série histórica dos log retornos do índice é um possível indicativo de que modelos como CAPM, Black e Scholes, que exigem a normalidade dos retornos, podem provocar resultados não confiáveis para as variáveis de retorno de

mercado. Para melhorar a análise dos dados, deve-se verificar a estacionariedade dos dados.

Para o teste de estacionariedade, verificou-se os diversos tipos de testes na literatura. Os principais são o teste de Dickey e Fuller Aumentado (Dickey e Fuller, 1979), Phillips e Peron (1988) e Kwiatkowski-Phillips-Schmidt-Shin – KPSS (Kwiatkowski et AL., 1992).

Optou-se por aplicar o teste de Dickey Fuller Aumentado (ADF) por ser o mais indicado e utilizado na literatura. Como visto, o teste verifica se a série já se encontra estacionária no nível ou se é necessário fazer diferenças entre ela para se tornar estacionária. As hipóteses do teste são:

 H_0 : 1RU: existe raiz unitária, isto é, a série é não estacionária

 $H_1: 0RU$: não existe raiz unitária, isto é, série estacionária

A tabela 5.1 mostra a estatística de teste para a série dos log retornos do IBOVESPA com significância de 5%.

Tabela 5.1 Teste ADF para os log retornos do IBOVESPA

Hipótese	Estatística t	p-valor
$H_0: 1RU$ $H_1: 0RU$	-109,1018	0,000

O teste ADF apresentado indica que o p-valor é inferior a 5%, revelando a rejeição de hipótese nula. Logo, a série dos log retornos do IBOVESPA é estacionária.

O passo seguinte é verificar se a série temporal segue um comportamento independente e identicamente distribuído, chamado i.i.d., isto é, todos os termos tem a mesma distribuição de probabilidade. O objetivo deste teste é identificar lineariedade ou não lineariedade determinista ou estocástica nos dados dos log retornos do IBOVESPA.

Para isso, aplica-se o teste desenvolvido por Brock, Dechert e Scheinkman (1996) – BDS. Suas hipóteses são:

 H_0 : a série é estocasticamente independente

 H_1 : a série é estocasticamente dependente

Tabela 5.2 Valores obtidos pelo teste BDS nas respectivas dimensões para o IBOVESPA

<u>Dimensão</u>	<u>BDS</u>	Erro-padrão	Estatística Z	<u>p-valor</u>
2	0.009046	0.001596	5.669845	0.0000
3	0.018695	0.002528	7.394849	0.0000
4	0.027910	0.003001	9.299171	0.0000
5	0.033667	0.003119	10.79547	0.0000
6	0.036541	0.002998	12.18758	0.0000

O teste BDS indica que, em todas as dimensões, os log retornos do IBOVESPA não seguem um comportamento independente e identicamente distribuído como se observa na tabela 5.2 os baixos p-valores onde se rejeita a hipótese nula.

Dessa forma, a série não tem comportamento independente e identicamente distribuído. Isso significa que existe uma dependência temporal entre os log retornos. Em outras palavras, os retornos futuros são influenciados pelos retornos passados. Para o fato de não serem identicamente distribuídos, refere-se que eles possuem intervalos de tempos com diferentes distribuições de probabilidade para os log retornos. Essa característica mostra a presença de não lineariedade dos log retornos.

Por se tratar de dados com periodicidade diária, as séries temporais apresentam maior persistência na volatilidade. Para verificar a presença de não linearidade, pode-se aplicar o teste de McLeod-Li (1983). O teste é baseado na autocorrelação amostral dos quadrados dos resíduos de uma sucessão cronológica estacionária gaussiana. As hipóteses do teste são:

 H_0 : a série temporal é linear

 H_1 : a série temporal é não-linear

O teste é feito após ajustar um modelo autoregressivo, AR(1) no caso, que remova a dependência linear da série dos log retornos, e faz-se a análise dos resíduos dessa regressão.

Tabela 5.3 Valores obtidos pelo teste de McLeod-Li para o IBOVESPA

Série Log Retornos IBOVESPA			
Estatística F	48,020	p-valor	0,00000

O teste de McLeod-Li, para 5 defasagens, rejeitou a hipótese nula pelo fato do p-valor ser inferior a 5%. Isto indica que a série temporal dos log retornos do IBOVESPA possui comportamento não-linear. A série dos log retornos ao quadrado apresenta forte autocorrelação, o que dá indícios de que o modelo auto-regressivo de heterocedasticidade condicional generalizada pode ser utilizado para a melhor modelagem da série.

Figura 5.5: Correlograma dos quadrados dos quadrados dos resíduos do IBOVESPA

Sample: 3 2477

Included observations: 2475

Q-statistic probabilities adjusted for 1 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
<u> </u>		1	0.139	0.139	48.020	
ı	ı ı	2	0.355	0.342	360.42	0.000
ı <mark></mark>		3	0.181	0.116	441.64	0.000
ı	•	4	0.212	0.081	553.48	0.000
ı		5	0.257	0.167	717.18	0.000
ı <mark>=</mark>	•	6	0.230	0.123	849.00	0.000
ı <u> </u>		7	0.280	0.139	1043.1	0.000
' <mark>=</mark>	•	8	0.116	-0.056	1076.4	0.000
ı		9	0.306	0.150	1309.2	0.000
' <u> </u>		10	0.233	0.145	1444.4	0.000
ı		11	0.320	0.149	1699.4	0.000
' =	ψ	12	0.201	0.006	1799.7	0.000
'	-	13	0.120	-0.120	1835.7	0.000
' =	"	14	0.122	-0.083	1872.9	0.000
' <u> </u>	1	15	0.170	0.044	1945.2	0.000
' <u> </u>	1	16	0.200	0.032	2044.5	0.000
ı <mark>=</mark>	1	17	0.205	0.052	2149.2	0.000
ı <mark>=</mark>	•	18	0.243	0.103	2297.1	0.000
•	 	19	0.094	-0.058	2319.3	0.000
ı —		20	0.224	0.033	2445.0	0.000

Um fato estilizado e bem conhecido das séries temporais de retornos financeiras é o comportamento heterocedástico, ou seja, se a variância dos dados não é constante para diferentes intervalos de tempo. Resumidamente, este efeito é caracterizado pelas

fortes oscilações do mercado que ocorrem quando este se comporta de forma inesperada e inconstante.

Uma vez estabelecido que exista algum tipo de não-linearidade em uma série temporal, é preciso identificar o tipo de não-linearidade existente, se na média ou na variância condicionadas.

Pode-se encontrar séries com dependência não linear na média nos modelos de média móveis e nos modelos TAR (*Threshold Autoregressive*) propostos por Tong e Lim (1980). Como exemplos de modelos com não lineariedade na variância cita-se os modelos da família GARCH.

Para identificar onde se encontra a não-lineariedade, aplicou-se o teste de Hsieh (1989). As hipóteses do teste são:

 H_0 : dependência não linear na variância condicionada

 H_1 : dependência não linear na média

Tabela 5.4 Teste de Hsieh para o IBOVESPA

IBOVESPA			
Coeficiente	Estatística de Hsieh	p-valor	
1 1	0,1762	0,6110	
1 2	0,1791	0,6635	
1 3	-0,0416	0,5844	
1 4	0,0867	0,7061	
1 5	-0,0772	0,6221	
2 2	0,1304	0,5744	
2 3	-0,1068	0,6354	
2 4	-0,0636	0,5452	
2 5	0,1481	0,5617	
3 3	0,1314	0,6295	
3 4	0,0108	0,5337	
3 5	0,1255	0,5239	
4 4	-0,0483	0,5077	
4 5	-0,0826	0,4519	
5 5	-0,0749	0,5360	

Embora não seja consenso na literatura sobre quantos valores de i,j para o teste deva ser empregado, usou-se 5 combinações, gerando 15 pares, por ser o mais empregado nos trabalhos pesquisados.

Conforme se observa na tabela 5.4, para quinze pares (*i,j*), os valores dos coeficientes do teste de Hsieh amostrais para os log retornos do IBOVESPA e o correspondente p-valor para um nível de confiança de 95%, podendo-se dizer que não se rejeita a hipótese nula dos coeficientes iguais a zero para nenhum valor, o que indica uma não-linearidade na variância condicionada.

De agora em diante, sabe-se da não lineariedade dos log retornos. Isso descarta quaisquer modelos que não sejam os de características não lineares. Assim, os modelos GARCH e as redes neurais de tempo recorrente são aplicados para realização de previsões.

Para se realizar as previsões estática com 4 passos a frente, estimou-se inicialmente um modelo de volatilidade condicional ARIMA-GARCH. O modelo estimado de acordo com as análises das funções de autocorrelação e autocorrelação parcial foi um modelo AR(1)-GARCH(1,1). A seleção da ordem p, q do modelo foi feita minimizando os critérios de informação do AIC (*Akaike Information Criteria*).

O critério AIC é de fundamental importância na seleção do modelo de melhor ajuste, por ponderar a quantidade de parâmetros com a função de verossimilhança. O modelo melhor ajustado é o que possuir menor número de parâmetros, isto é chamado parcimonioso, relacionados a altos valores da função de verossimilhança, provocando baixos valores do AIC. O modelo a ser escolhido é aquele que apresentar o menor AIC.

As variáveis são descritas a seguir para o modelo escolhido com base neste critério.

Análise dos parâmetros estimados para o modelo AR(1)-GARCH(1,1):

$$LOGIBOV_{t} = \phi_{1}LOGIBOV_{t-1} + \varepsilon_{t}$$

$$\varepsilon_t^2 \left| I_{t-1} \sim N\left(0; h_t\right) \right.$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

Tabela 5.5 Parâmetros do modelo AR(1)-GARCH(1,1) para os log retornos do IBOVESPA

Parâmetros	Coeficiente	Erro-padrão	Estatística z	p-valor
$\phi_{ m l}$	0,006490	0,022150	0,292988	0,7695
α_0	9,84E-06	2,09E-06	4,708958	0,0000
α_1	0,065818	0,008411	7,825356	0,0000
$eta_{ m l}$	0,907825	0,011714	77,50009	0,000

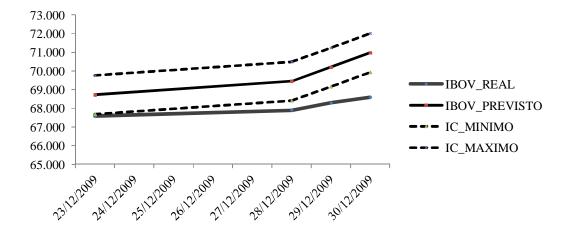
De forma geral, pode-se observar que o modelo apresenta bom ajuste à série dos log retornos para o modelo ARIMA-GARCH, visto que os p-valores são significativos com significância de 5% e até mesmo de 1% por rejeitar a hipótese nula de serem iguais a zero.

Outro ponto a ser observado é que a soma dos parâmetros α_1 (ARCH) e β_1 (GARCH) próximo de 1, indica o grau de persistência da série temporal. Esse indicador é crucial para se fazer previsão, tanto de curto quanto de longo prazo. Quanto maior a soma desses parâmetros, maior o grau de persistência e isso revela uma menor capacidade preditiva de longo prazo do modelo.

A figura 5.5 ilustra os valores previstos, os valores reais para uma previsão estática de 4 passos a frente, bem como o intervalo de confiança (IC) de 95% para os valores previstos. Observa-se que o intervalo de confiança está 100% fora dos valores reais do IBOVESPA.

Cabe aqui ressaltar que nos gráficos a seguir, embora apareçam oito datas, a previsão foi de quatro passos a frente. Manteve-se a data dos valores previstos apenas por uma questão de didática, uma vez que houve feriados e pontes de períodos sem pregão entre as datas de 24/12/2009 e 27/12/2009, voltando a ter pregão em 28/12/2009.

Figura 5.6: Gráfico do IBOVESPA real e previsto 4 passos a frente com previsão estática para o modelo ARIMA-GARCH



A baixa capacidade preditiva deste modelo, já identificada acima, apresenta os indicadores de análise de previsão dados na tabela 5.6.

Tabela 5.6 Estatísticas de acurácia do modelo AR(1)-GARCH(1,1) para o IBOVESPA

Estatísticas de acurácia do modelo preditivo	Valores
TIC	0,013089
MAPE	2,56%
Correlação	0,227729

Observa-se que a correlação entre os valores reais e previstos ocorre pelo movimento de tendência verificado na série real e na série prevista. O valor do TIC mostra um bom ajuste, mas um erro relativamente alto.

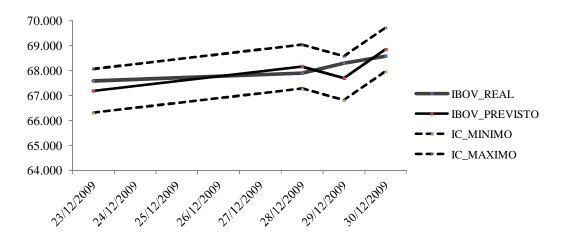
Devido à baixa eficácia do modelo da família GARCH, usou-se também os modelos não paramétricos de redes neurais. As redes neurais representam uma saída quando não se consegue criar modelos adequados à realidade dos dados. Essa dificuldade é encontrada nas séries temporais financeiras devido à mudança contínua na volatilidade para curtos períodos de tempo.

Foi utilizada aqui uma rede neural recorrente onde existe uma conexão de realimentação da rede entre os processadores em uma mesma camada e também em camadas diferentes.

O ajuste da rede neural teve duas fases: a primeira é a fase de treinamento onde foram usados 2300 valores dos 2477 disponíveis, o que corresponde 92,85% dos dados. A segunda fase é o desempenho da rede, onde foram utilizados os demais dados para testes, para posterior utilização da rede. A rede contou com um neurônio na camada de entrada, quatro neurônios na camada intermediária e um na camada de saída. A função de ativação utilizada foi a função logística com 200 épocas de treinamento.

Os resultados encontrados estão na figura 5.7.

Figura 5.7: Gráfico do IBOVESPA real e previsto 4 passos a frente com previsão estática para Redes Neurais Recorrentes



Pode-se observar que a qualidade do ajuste e das previsões melhora sensivelmente com uso das redes neurais recorrentes. O intervalo de confiança compreende quase todos os valores reais. Apenas um valor ficou de fora.

A tabela 5.7 apresenta os indicadores de análise de previsão para as redes neurais recorrentes.

Tabela 5.7 Estatísticas de acurácia com uso de redes neurais recorrentes para o IBOVESPA

Estatísticas de acurácia do modelo preditivo	Valores
TIC	0,005389
MAPE	0,84%
Correlação	0,544006

Posteriormente, acredita-se que o uso combinado das redes neurais com filtros de previsão, tenderiam a melhorar, uma vez que os filtros teriam a capacidade de diluir a turbulência dos retornos.

O primeiro filtro aplicado foi de *wavelets*, como já estudado em Lima (2004). Os resultados podem ser vistos no gráfico 5.7. A forma de onda escolhida foi a *wavelet* de "daubesch" número 1, por ser uma *wavelet* primária.

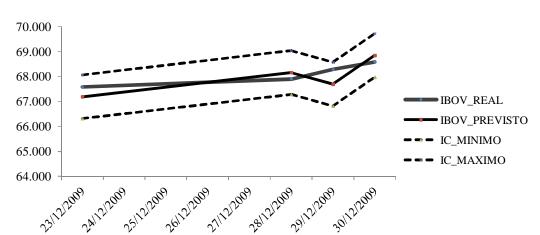


Figura 5.8: Gráfico do IBOVESPA real e previsto 4 passos a frente com previsão estática para Redes Neurais Recorrentes com filtro de *wavelets*

Os resultados apontam para uma melhora na qualidade dos valores previstos e consequente melhora dos indicadores de previsão como pode ser visto na tabela 5.8.

Tabela 5.8 Estatísticas de acurácia com uso de redes neurais recorrentes para o IBOVESPA com filtro de *wavelets*

Estatísticas de acurácia do modelo preditivo	Valores
TIC	0,004417
MAPE	0,83%
Correlação	0,705138

Nota-se que o ajuste das previsões foi melhor, como dado pelo TIC menor. O erro médio das previsões reduziu 0,01% e a correlação entre os valores previstos e reais aumentou significativamente.

Dessa forma, o processo de uso do filtro de *wavelets* colaborou para a redução do erro nas previsões. Outro processo de decomposição da série temporal para uso de previsão é através dos modelos estruturais. O modelo aplicado aqui é o de espaços de estados cujo avanço se deu a partir da obra Harvey (2001), onde sua teoria foi descrita.

Um modelo de espaços de estados é construído, baseado na independência do evento futuro do processo em relação a seu estado passado, dado que existe o estado presente. Isto é, as informações do passado estão contidas no estado do processo. O estado de um processo significa encontrar o menor número de variáveis independentes tal que a partir do conhecimento dessas variáveis no instante inicial, determina-se o comportamento do sistema para os instantes futuros.

A representação desse estado é feita por duas equações dinâmicas, onde uma é a equação que indica os valores observados do processo obtidos em função do vetor de estado e a outra é a equação de transição que indicará a evolução dinâmica do vetor de estado não observado. (Souza, 1989, p. 54).

Como visto, estes modelos supõem que os movimentos característicos suportam ser decompostos em parcelas não-observáveis, como por exemplo, tendência, sazonalidade, parte cíclica e parte aleatória (erro). A contribuição desse processo é que cada componente poderá ser interpretado diretamente devido à forma como o modelo é estimado.

As previsões futuras para a série temporal, para o modelo estimado na forma de espaço de estados, são feitas através de estimadores atualizados do vetor de estado que não é claramente observado.

A ferramenta para estimação e previsão por modelos estruturais utilizada é o filtro de Kalman, que estima equações de previsão e de atualização das posições em cada instante de tempo.

Vale ressaltar que, segundo Harvey (2001, p. 231), as variáveis não representam necessariamente medidas de quantidades físicas, além de não ser única. Como as variáveis de estado estimadas são independentes, não podem ser expressas como funções algébricas de outras variáveis de estado.

A seguir, os resultados obtidos segundo o uso de modelos estruturais com filtro de kalman.

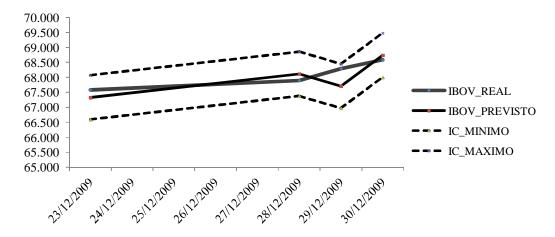
Tabela 5.9 Estatísticas de acurácia com filtro de Kalman para o IBOVESPA

Estatísticas de acurácia do modelo preditivo	Valores
TIC	0,005505
MAPE	0,86%
Correlação	0,598134

Pode-se observar que o ajuste das previsões pelo filtro de Kalman não foram superiores às obtidas com uso de redes neurais e *wavelets*. O erro das previsões aumentou e a correlação com os valores reais diminuiu, sendo ainda mais eficiente que o uso sozinho de redes neurais.

A figura a seguir, ilustra as previsões feitas.

Figura 5.9: Gráfico do IBOVESPA real e previsto 4 passos a frente com previsão estática para filtro de Kalman



O uso separado dos filtros de Kalman e *wavelets* identificou que as *wavelets* com uso de redes neurais recorrentes apresentou melhor ajuste e melhor qualidade das previsões estáticas, embora se ressalta que o filtro de Kalman foi mais eficiente que os modelos econométricos da família GARCH.

Agora, tem-se a seguinte questão: o que aconteceria se combinar-se as técnicas de filtragem? Ou seja, se utilizar o filtro de Kalman e sobre seu filtro usar as *wavelets* para realizar as previsões com redes neurais recorrentes, uma vez que as redes melhoram o ajuste não linear dos dados. Ou também, se o reverso traria melhores resultados.

Os resultados dessa combinação são encontrados a seguir.

A tabela a seguir ilustra a aplicação combinada das previsões estáticas com 4 passos a frente, utilizando-se primeiro o algoritmo de Kalman e sobre ele aplica-se o filtro de *wavelets* com decomposição em um nível.

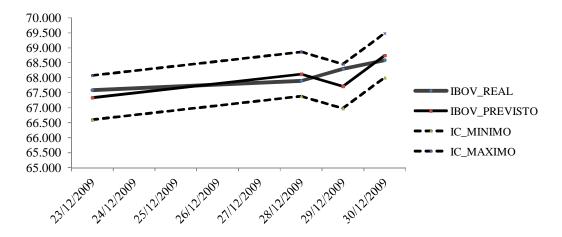
Tabela 5.10 Estatísticas de acurácia combinando redes neurais recorrentes com filtro de Kalman primeiro e *wavelets* depois para o IBOVESPA

Estatísticas de acurácia do modelo preditivo	Valores
TIC	0,004546
MAPE	0,72%
Correlação	0,671659

O ajuste foi o melhor obtido, bem como a estatística de erro das previsões. A correlação é estatisticamente mais eficiente de todos os modelos verificados até aqui.

O gráfico das previsões é dado abaixo:

Figura 5.10: Gráfico do IBOVESPA real e previsto 4 passos a frente com previsão estática pelo uso de redes neurais recorrentes com filtro de Kalman primeiro e *wavelets* depois.



A seguir, fez-se a aplicação do uso combinado das técnicas ao reverso, ou seja, aplicou-se primeiro a decomposição de *wavelets* em um nível sobre cada um as previsões com filtro de Kalman.

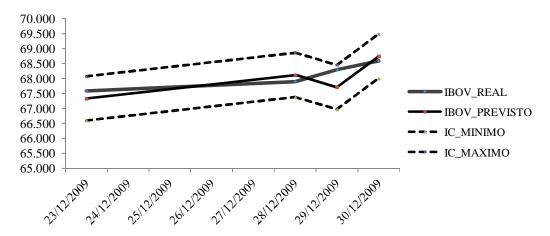
Tabela 5.11 Estatísticas de acurácia combinando redes neurais recorrentes com filtro de *wavelets* primeiro e Kalman depois para o IBOVESPA.

Estatísticas de acurácia do modelo preditivo	Valores
TIC	0,005321
MAPE	0,93%
Correlação	0,563688

O reverso da técnica combinada do uso de filtros não melhora a qualidade e ajuste das previsões como verificado pelas estatísticas acima.

O gráfico das previsões é dado abaixo:

Figura 5.11: Gráfico do IBOVESPA real e previsto 4 passos a frente com previsão estática pelo uso de redes neurais recorrentes com filtro *wavelets* primeiro e Kalman depois



Também se verificou se as alterações na forma das *wavelets* trariam algum benefício para a melhora na qualidade das previsões. Dessa forma, realizaram-se as decomposições para posterior previsão com redes neurais variando os diferentes tipos de *wavelets*.

Detectou-se que o tipo de *wavelet* altera o resultado. Isto é, cada forma de onda apresenta melhor ou pior ajuste, maior ou menor MAPE e correlação. Um resultado interessante foi que, embora haja mudança nos indicadores, nenhum resultado superou o da *wavelet* primária 'db1', ou seja, cada *wavelet* no processo de decomposição atrai mais ou menos pontos, pois de acordo com a sua forma, capta mais ou menos pontos ao passar pela série temporal. Esse fato provoca alterações nos pesos da rede que reagem alterando o ajuste das previsões e, consequentemente, na transformação inversa para voltar a previsão no nível.

Harter (2004, p. 33) encontrou resultados semelhantes ao verificado aqui para análise de séries temporais de precipitação na cidade de Pelotas-RS, trabalhando também com a transformada de *wavelets*. A autora, porém, utilizou apenas a *wavelet* de Morlet para realização de previsões.

A tabela abaixo mostra as estatísticas de erro para as opções de ondas.

Tabela 5.12 Estatísticas de acurácia com uso de redes neurais recorrentes para o IBOVESPA com filtro de Kalman e filtro de *wavelets*.

Forma		RN com	Wavelet	sem Kalman	RN com K	alman e <i>Wavelet</i>
de onda	Pontos na divisão	tic	mape	correlação	tic	mape correlação
haar	1238	0,004417	0,83%	0,705138	0,004546	0,72% 0,671659
db1	1238	0,004417	0,83%	0,705138	0,004546	0,72% 0,671659
db6	1243	0,006341	1,03%	0,508117	0,01000	1,70% -0,01756
db7	1244	0,005684	0,96%	0,53107	0,009729	1,69% -0,04%
db8	1245	0,005685	0,64%	0,5172	0,009716	1,72% -0,05755
db9	1246	0,00644	1,01%	0,427585	0,01000	1,74% -0,07063
db10	1247	0,007621	1,14%	0,319726	0,01000	1,77% -0,09061
sym6	1243	0,006934	1,13%	0,27725	0,011252	1,88% -0,25469
sym7	1244	0,006252	1,03%	0,448249	0,01102	1,89% -0,11739
sym8	1245	0,005918	0,99%	0,460502	0,009804	1,75% -0,08894
coif2	1243	0,006181	1,07%	0,433283	0,011202	1,92% -0,15089
coif3	1246	0,005292	0,89%	0,574482	0,009523	1,69% -0,03772
coif4	1249	0,006362	1,00%	0,433372	0,009811	1,74% -0,07392
coif5	1252	0,006341	1,00%	0,432275	0,00980	1,74% -0,07552
bior2.6	1244	0,006815	1,11%	0,367767	0,011345	1,95% -0,16896
bior2.8	1246	0,005797	0,96%	0,475459	0,009574	1,72% -0,08549
bior3.5	1243	0,006752	1,10%	0,311082	0,011443	1,92% -0,24911
bior3.7	1245	0,006279	1,04%	0,416263	0,016007	1,79% -0,1061
bior3.9	1247	0,006099	0,94%	0,468296	0,009607	1,71% -0,06317
bior6.8	1246	0,005734	0,96%	0,487712	0,00961	1,73% -0,0805
rbio2.6	1244	0,006708	1,11%	0,380267	0,011348	1,96% -0,16384
rbio2.8	1246	0,00562	0,94%	0,504082	0,009622	1,72% -0,07711
rbio3.5	1243	0,006633	1,12%	0,299398	0,011338	1,92% -0,25395
rbio3.7	1245	0,006398	1,05%	0,419519	0,010295	1,82% -0,10087
rbio3.9	1247	0,006175	0,99%	0,449249	0,00981	1,75% -0,08391
rbio5.5	1243	0,006522	1,11%	0,34263	0,011607	1,95% -0,23877
rbio6.8	1246	0,005738	0,95%	0,489741	0,009639	1,73% -0,06545
dmey	1288	0,006334	1,00%	0,435683	0,00981	1,74% -0,07376

Foram selecionadas 28 formas de onda e rodada a rede neural para previsão 4 passos a frente, fazendo uso da combinação de filtro de Kalman e de *Wavelet*. Primeiramente usou-se a decomposição por cada forma de onda e sobre a série decomposta, aplicou-se o filtro de Kalman para previsão.

Posteriormente, aplicou-se inicialmente o filtro de Kalman e sobre ele a decomposição de *Wavelet*. Os resultados apontaram que o erro somente é inferior para a forma de onda primária, no caso para a *wavelet* de "Haar" e "daubesch 1". Nas demais,

o erro se propaga. A ordem que se obteve melhor resultado foi a de usar o filtro de Kalman primeiro e sobre ele, a decomposição de *wavelets*.

A tabela 5.13 mostra o efeito da ordem contrária, sendo feito primeiro o filtro de *wavelets* e posteriormente a previsão pelo filtro de Kalman. Os resultados foram inferiores ao feito na ordem contrária.

Uma observação importante é quanto à decomposição das sub-séries. Dependendo da forma que se utiliza as sub-séries, contém mais ou menos pontos. Esse fato é devido ao tipo e a forma da *wavelet* que, ao passar pela série temporal, capta mais ou menos pontos de acordo com a sua dilatação ou contração.

A transformada de *wavelet* fornece seus resultados de acordo com a quantidade de dados captados na análise, havendo então uma relação entre o número de observações e a escala dada na sua forma de onda.

O vetor de *wavelet* depende da forma adequada de sua escolha. Detectou-se que a forma de onda escolhida deve possuir características semelhantes ao encontrado na série temporal que se está utilizando.

No caso das séries temporais financeiras com agrupamentos de volatilidade, como caso do IBOVESPA em estudo, uma boa escolha seria a *wavelet* de Haar ou de "daubesch1", ou seja, escolher uma forma de onda primária.

Tabela 5.13 Estatísticas de acurácia com uso de redes neurais recorrentes para o IBOVESPA com filtro de *wavelets* e filtro de Kalman

Forma de		RN com Wavelet e Kalman				
onda	Pontos na divisão	tic	mape	correlação		
haar	1238	0,005321	0,93%	0,563688		
db1	1238	0,005321	0,93%	0,563688		
db6	1243	0,009740	1,60%	0,047039		
db7	1244	0,009849	1,65%	-0,011657		
db8	1245	0,009846	1,70%	-0,035632		
db9	1246	0,00978	1,73%	-0,064341		
db10	1247	0,009864	1,78%	-0,098635		
sym6	1243	0,011756	2,00%	-0,299302		
sym7	1244	0,01112	1,89%	-0,071799		

Forma de		RN com Wavelet e Kalman		
onda	Pontos na divisão	tic	mape	correlação
sym8	1245	0,019293	1,82%	-0,114365
coif2	1243	0,011607	1,98%	-0,131841
coif3	1246	0,009165	1,61%	0,008895
coif4	1249	0,009793	1,74%	-0,074515
coif5	1252	0,009805	1,75%	-0,077529
bior2.6	1244	0,011551	2,00%	-0,143812
bior2.8	1246	0,009504	1,72%	-0,09328
bior3.5	1243	0,012125	2,05%	-0,286781
bior3.7	1245	0,010571	1,88%	-0,123581
bior3.9	1247	0,009744	1,72%	-0,073114
bior6.8	1246	0,009491	1,71%	-0,070862
rbio2.6	1244	0,01173	2,03%	-0,141263
rbio2.8	1246	0,009436	1,70%	-0,060745
rbio3.5	1243	0,01189	2,04%	-0,289256
rbio3.7	1245	0,010625	1,87%	-0,088728
rbio3.9	1247	0,009989	1,77%	-0,095173
rbio5.5	1243	0,01202	2,05%	-0,248366
rbio6.8	1246	0,009488	1,71%	-0,065451
dmey	1288	0,00981	1,74%	-0,073764

Resumindo todo o processo de análise, pode-se notar que o uso combinado das técnicas de filtragem trouxe relativo benefício no ajuste das previsões e na qualidade das mesmas. A tabela a seguir resume todas as medidas, lembrando que volatilidade do IBOVESPA de 21 dias foi de 12,71%.

Tabela 5.14 Resumo das estatísticas de previsão para o IBOVESPA

Medida	ARIMA	RN-	RN-	Filtro	FK_WAV	WAV_FK
	GARCH	RECORR	REC_WAV	Kalman		
TIC	0,013089	0,005389	0,004417	0,005505	0,004546	0,005321
MAPE	2,56%	0,84%	0,83%	0,86%	0,72%	0,93%
CORREL	0,227729	0,54406	0,705138	0,598134	0,671659	0,563688

Dessa forma, o uso do filtro de Kalman, com posterior decomposição de *wavelets* e previsão com redes neurais trouxe mais benefícios do que as demais combinações de modelos.

A seguir, testa-se a mesma metodologia para uma série de uma commoditite.

5.2 Análise da série da commoditie do barril de petróleo

Aplica-se a seguir, todo o processo de análise para uma *commoditie*. A escolhida foi o preço em dólar americano do barril de petróleo bruto pelo fato da sua importância não só para a economia brasileira como mundial. A série analisada compreende o período de 04/01/2000 a 08/01/2010, num total de 2511 observações bruta dos preços *spot* cotados na bolsa americana.

Segundo dados da OPEP (2010), a produção anual de petróleo cru do mundo é de cerca de 24 bilhões de barris, para um comsuno anual de 23 bilhões, o que garante uma sobra de 1 bilhão que ficam em estoque. Ainda segundo a organização, as reservas existentes no mundo são calculadas em aproximadamente 1 trilhão de barris de petróleo. Dessa forma, o petróleo é uma das commodities mais importantes de todo o planeta, responsável pela geração de seus derivados que atendem diversas áreas do setor produtivo. Logo, o conhecimento de seu comportamento futuro é de extrema relevância para toda a comunidade.

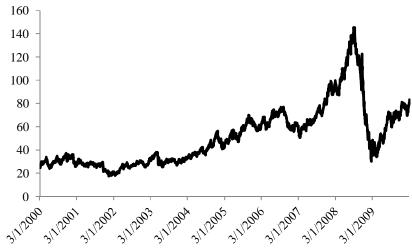


Figura 5.12 Comportamento do preço do barril de petróleo (\$ US\$)

Fonte: US Energy Information Administration http://tonto.eia.doe.gov/dnav/pet/pet_pri_spt_s1_d.htm

A figura 5.12 destaca o gráfico do petróleo bruto com suas cotações de fechamento em dólares americanos. Como se nota, a série do petróleo cru apresenta algumas oscilações por sofrer influência de variáveis exógenas como crises, boatos,

novas descobertas de novos poços que influenciam na oscilação de seus retornos. A figura a seguir destaca os log retornos.

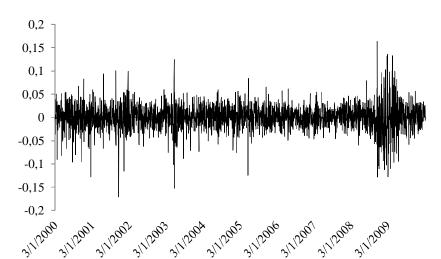


Figura 5.13: Série temporal dos log retornos do petróleo cru diário

Observa-se também que os retornos diários oscilam ao redor de zero, apresentando uma variabilidade que depende do tempo, chamada volatilidade, com períodos de alta e de baixa variabilidade e dias em que o retorno é um valor anormal, chamado *outlier*. Além disso, notam-se diversos *clusters* de volatilidade e picos que demandaram fortes oscilações em seu comportamento representado no gráfico como linhas e picos de grande variação que acontecem devido às incertezas do mercado.

O barril de petróleo que chegou a passar dos US\$ 100 no início de 2008 passou neste período por fortes oscilações no mercado internacional, chegando a pouco mais de US\$ 30 no final de 2008 e está mais de US\$ 83 no início de 2010.

A figura a seguir, destaca a inspeção visual do histograma e do gráfico da densidade de probabilidade dos log retornos.

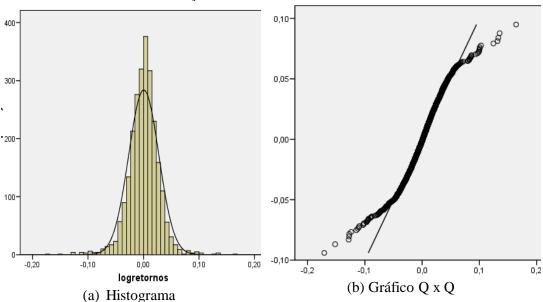


Figura 5.14: Histograma dos log retornos diários da série do petróleo cru diário e retornos diários contra distribuição normal

Verifica-se que o histograma tem a sua parte central mais alta do que a própria curva normal e há presença de valores bem afastados da tendência central dos dados. Essas informações são características das séries temporais financeiras e são descritos por comportamento leptocúrtico com caudas mais pesadas que a normal, pelo fato dos valores se afastarem da média a vários múltiplos do desvio padrão, conforme também verificado no exemplo do Ibovespa.

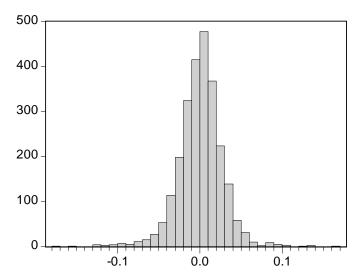
Na figura 5.14 observa-se ainda no item (b) o gráfico Q x Q, que serve para verificar se os dados tendessem a apresentar uma distribuição normal, os pontos estariam sobre a reta, o que não acontece no caso do petróleo cru no período analisado.

O passo seguinte descreve o teste de normalidade dos log retornos de Jarque Bera (1987). As hipóteses para este teste são definidas como:

 H_0 : a série segue uma distribuição normal

 H_1 : a série não segue uma distribuição normal

Figura 5.15: Histograma dos log retornos diários do petróleo cru com teste de normalidade de Jarque Bera (1987).



Series: LOGOLEO Sample 1 2510 Observations 2510				
Mean	0.000468			
Median	0.001370			
Maximum	0.164137			
Minimum	-0.170918			
Std. Dev.	0.027138			
Skewness	-0.271185			
Kurtosis	7.168388			
Jarque-Bera	1847.949			
Probability	0.000000			

A estatística de teste indica o valor de JB = 1.847,949 com p-valor igual a zero. A hipótese é rejeitada tanto a 5% quanto a 1% de significância, o que revela que a série não segue uma distribuição gaussiana.

O próximo passo é o teste de raízes unitárias. A tabela 5.14 mostra a estatística de teste para a série dos log retornos do petróleo cru com significância de 5%.

Tabela 5.15 Teste ADF para os log retornos do Petróleo

Hipótese	Estatística t	p-valor
$H_0:1RU$	-108,5839	0,000
$H_1: ORU$		

Rejeita-se a hipótese nula de existência de raízes unitárias, sendo, portanto, uma série já com comportamento estacionária.

Aplica-se agora o teste desenvolvido por Brock, Dechert e Scheinkman (1996) – BDS. O objetivo deste teste é identificar lineariedade ou não lineariedade determinista ou estocástica nos dados dos log retornos do petróleo cru.

Tabela 5.16 Valores obtidos pelo teste BDS nas respectivas dimensões para o petróleo

<u>Dimensão</u>	<u>BDS</u>	Erro-padrão	Estatística Z	<u>p-valor</u>
2	0,015158	0,001670	9,076904	0,0000
3	0,029722	0,002649	11,21842	0,0000
4	0,039040	0,003150	12,39536	0,0000
5	0,042411	0,003277	12,94152	0,0000
6	0,043314	0,003155	13,72898	0,0000

O teste BDS revela que, em todas as dimensões, os log retornos do petróleo cru não seguem um comportamento independente e identicamente distribuído como se observa na tabela 5.15 os baixos p-valores onde se rejeita a hipótese nula. Essa característica mostra a presença de não lineariedade dos log retornos.

Para comprovar a presença de não linearidade, pode-se aplicar o teste de McLeod-Li (1983). Como descrito anteriormente, o teste é feito após ajustar um modelo autoregressivo, AR(1) no caso, que remova a dependência linear da série dos log retornos, e faz-se a análise dos resíduos dessa regressão.

Tabela 5.17 Valores obtidos pelo teste de McLeod-Li para o Petróleo

Série Log Retornos Petróleo bruto						
Estatística F	62,01141	p-valor	0,00000			

O teste de McLeod-Li, para 5 defasagens, rejeitou a hipótese nula pelo fato do p-valor ser inferior a 5%. Isto indica que a série temporal dos log retornos do petróleo possui comportamento não-linear. A série dos log retornos ao quadrado apresentou forte autocorrelação, o que dá indícios de que o modelo auto-regressivo de heterocedasticidade condicional generalizada pode ser utilizado para a melhor modelagem da série.

Figura 5.16: Correlograma dos quadrados dos quadrados dos resíduos para o petróleo

Sample: 2 2510

Included observations: 2509

Q-statistic probabilities adjusted for 1 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
Adiocolleration		1 2 3 4 5 6 7 8	0.227 0.150 0.176 0.223 0.189 0.152 0.130 0.101	0.227 0.104 0.130 0.161 0.098 0.057 0.032 -0.002	129.79 186.30 264.56 390.17 479.64 538.12 580.94 606.54	0.000 0.000 0.000 0.000 0.000 0.000 0.000
		9 10 11 12 13 14 15 16 17 18 19 20	0.122 0.136 0.099 0.092 0.192 0.143 0.123 0.135 0.156 0.137 0.110	0.033 0.049 0.010 0.017 0.129 0.041 0.033 0.044 0.048 0.023 -0.002	644.08 690.61 715.17 736.36 829.04 880.94 919.24 965.61 1026.9 1074.2 1104.7 1195.5	0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

Uma vez estabelecido que exista algum tipo de não-linearidade em uma série temporal, é preciso identificar o tipo de não-linearidade existente, se na média ou na variância condicionadas. O teste de Hsieh define tal processo como pode ser visto a seguir:

Tabela 5.18 Teste de Hsieh para o Petróleo

	PETRÓLEO				
Coeficiente	Estatística de Hsieh	p-valor			
1 1	-0,2141	0,4152			
1 2	-0,3637	0,3580			
1 3	-0,3784	0,3526			
1 4	-0,4597	0,3229			
1 5	0,2453	0,5969			
2 2	0,3592	0,6403			
2 3	0,2593	0,6023			
2 4	0,1611	0,5640			
2 5	-0,0573	0,4772			
3 3	-0,047	0,4813			
3 4	-0,7017	0,2414			
3 5	0,0248	0,5099			

PETRÓLEO		
Coeficiente	Estatística de Hsieh	p-valor
4 4	-0,2602	0,3974
4 5	0,4042	0,6570
5 5	0,0371	0,5148

Conforme se observa, na tabela 5.17, para quinze pares (*i,j*), os valores dos coeficientes do teste de Hsieh amostrais para os log retornos do Petróleo e o correspondente p-valor para um nível de confiança de 95%, sendo todos superiores a 0,05 (5%), o que revela a não rejeição da hipótese nula dos coeficientes iguais a zero para nenhum valor, ou seja, existe não-linearidade na variância condicionada. Assim, os modelos GARCH e as redes neurais de tempo recorrente podem ser aplicados para realização de previsões.

Realizaram-se as previsões estáticas com quatro passos à frente por um modelo de volatilidade condicional ARIMA-GARCH. O modelo estimado, de acordo com as análises das funções de autocorrelação e autocorrelação parcial, foi um modelo AR(1)-GARCH(1,1). A seleção da ordem p, q do modelo foi feita minimizando os critérios de informação do AIC (*Akaike Information Criteria*).

Análise dos parâmetros estimados para o modelo AR(1)-GARCH(1,1):

$$\begin{aligned} OLEO_t &= \phi_1 OLEO_{t-1} + \varepsilon_t \\ \varepsilon_t^2 \left| I_{t-1} &\sim N\left(0; h_t\right) \right. \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \end{aligned}$$

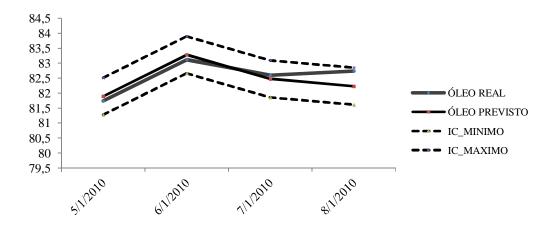
Tabela 5.19 Parâmetros do modelo AR(1)-GARCH(1,1) para os log retornos do Petróleo

Parâmetros	Coeficiente	Erro-padrão	Estatística z	p-valor
$\phi_{ m l}$	-0,035666	0,019525	-1,826665	0,0678
α_0	1,66E-05	3,52E-06	4,719082	0,0000
α_1	0,066008	0,006347	10,40009	0,0000
eta_1	0,909644	0,010603	85,79008	0,0000

De forma geral, pode-se observar que o modelo apresenta bom ajuste à série dos log retornos do petróleo para o modelo ARIMA-GARCH, visto que os p-valores são significativos com significância de 5% e até mesmo de 1% por rejeitar a hipótese nula de serem iguais a zero com exceção do parâmetro ϕ_1 .

A figura 5.18 ilustra os valores previstos, os valores reais para uma previsão estática de 4 passos a frente, bem como o intervalo de confiança (IC) de 95% para os valores previstos. Observa-se que o intervalo de confiança está 100% fora dos valores reais do IBOVESPA.

Figura 5.17: Gráfico do Petróleo real e previsto 4 passos a frente com previsão estática para o modelo ARIMA-GARCH



A tabela a seguir demonstra as estatísticas de erro adotadas na pesquisa para a série do petróleo.

Tabela 5.20 Estatísticas de acurácia do modelo AR(1)-GARCH(1,1) para o Petróleo

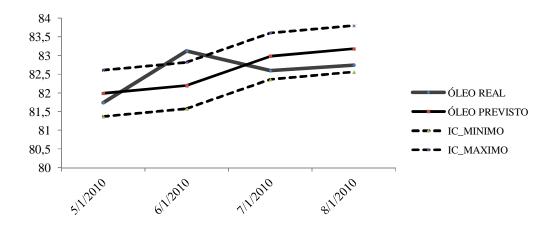
Estatísticas de acurácia do modelo preditivo	Valores
TIC	0,005968
MAPE	1,033%
Correlação	0,395243

Observa-se uma baixa correlação verificada entre a série real e a série prevista. O baixo valor do TIC mostra um bom ajuste, mas um erro relativamente baixo.

Devido à baixa eficácia do modelo da família GARCH, usou-se também os modelos não paramétricos de redes neurais. Foi utilizada aqui uma rede neural recorrente em duas fases: a primeira é a fase de treinamento onde foram usados 2311 valores dos 2511 disponíveis, o que corresponde 92,04% dos dados. A segunda fase é o desempenho da rede, onde foram utilizados os demais dados para testes para posterior utilização da rede. A rede contou com um neurônio na camada de entrada, quatro

neurônios na camada intermediária e um na camada de saída. A função de ativação utilizada foi à função logística com 200 épocas de treinamento.

Figura 5.18: Gráfico do Petróleo real e previsto 4 passos a frente com previsão via redes neurais recorrentes



O gráfico 5.18 acima mostra os valores reais e previstos para o barril de petróleo com previsão de quatro passos a frente.

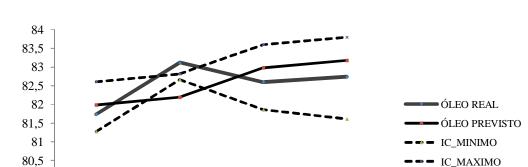
A tabela a seguir demonstra as estatísticas de erro adotadas na pesquisa para a série do petróleo.

Tabela 5.21 Estatísticas de acurácia com previsão feita por redes neurais recorrentes para o petróleo

Estatísticas de acurácia do modelo preditivo	Valores
TIC	0,004164
MAPE	0,66%
Correlação	0,401368

Observa-se uma baixa correlação verificada entre a série real e a série prevista. O baixo valor do TIC mostra um bom ajuste e erro considerado bom.

Dessa forma, passa-se a partir de agora a fazer uso combinado dos filtros de previsão. O primeiro filtro aplicado foi de *wavelets*, com a forma de onda "daubesch" número 1. Os resultados são demonstrados a seguir.



80

Figura 5.19: Gráfico do Petróleo real e previsto 4 passos a frente com previsão para Redes Neurais Recorrentes com filtro de *wavelets*

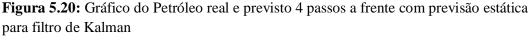
Os resultados apontam para uma melhora na qualidade dos valores previstos e consequente melhora dos indicadores de previsão como pode ser visto na tabela 5.20.

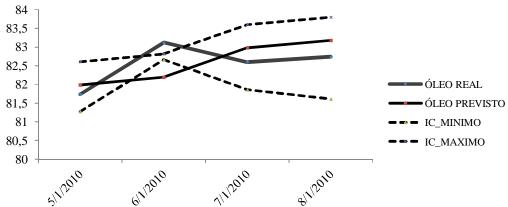
Tabela 5.22 Estatísticas de acurácia com uso de redes neurais recorrentes para o petróleo com filtro de *wavelets*

Estatísticas de acurácia do modelo preditivo	Valores
TIC	0,003132
MAPE	0,49%
Correlação	0,498437

Nota-se que o ajuste das previsões foi melhor, como dado pelo TIC menor. O erro médio das previsões reduziu de 0,66% para 0,49% e a correlação entre os valores previstos e reais aumentou.

Dessa forma, o processo de uso do filtro de *wavelets* colaborou para a redução do erro nas previsões. Aplica-se a seguir o filtro de Kalman para tratamento da série segundo o uso desta ferramenta.





Pode-se observar que o ajuste das previsões pelo filtro de Kalman não foi superior ao obtido com uso de redes neurais e *wavelets*. O erro das previsões aumentou e a correlação com os valores reais diminuiu, sendo ainda mais eficiente que o uso sozinho de redes neurais, como pode ser visto na tabela a seguir:

Tabela 5.23 Estatísticas de acurácia com filtro de Kalman para o petróleo.

Estatísticas de acurácia do modelo preditivo	Valores
TIC	0,004192
MAPE	0,70%
Correlação	0,480144

Fazendo a partir de agora o uso combinado dos filtros, tem-se a seguir a aplicação primeiro do algoritmo de Kalman e sobre ele aplica-se o filtro de *wavelets* com decomposição em um nível. Os resultados podem ser vistos a seguir.

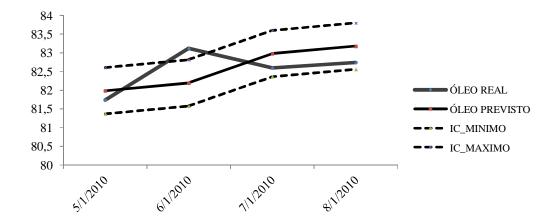
Tabela 5.24 Estatísticas de acurácia combinando redes neurais recorrentes com filtro de Kalman primeiro e *wavelets* depois para o petróleo

Estatísticas de acurácia do modelo preditivo	Valores
TIC	0,004223
MAPE	0,68%
Correlação	0,54132

O ajuste apresentou ligeiro aumento, enquanto a estatística de erro dada pelo MAPE teve pequena queda. A correlação é estatisticamente mais eficiente de todos os modelos verificados até aqui.

O gráfico das previsões é dado abaixo:

Figura 5.21: Gráfico do petróleo real e previsto 4 passos a frente com previsão estática pelo uso de redes neurais recorrentes com filtro de Kalman primeiro e *wavelets* depois.



Fez-se a aplicação do uso combinado das técnicas ao reverso, ou seja, aplicou-se primeiro a decomposição de *wavelets* em um nível sobre cada uma das séries alisadas, realizando-se as previsões com filtro de Kalman. Os resultados são apresentados a seguir.

Tabela 5.25 Estatísticas de acurácia combinando redes neurais recorrentes com filtro de *wavelets* primeiro e Kalman depois para o petróleo.

Estatísticas de acurácia do modelo preditivo	Valores
TIC	0,004392
MAPE	0,60%
Correlação	0,486431

O reverso da técnica combinada do uso de filtros não melhora a qualidade e ajuste das previsões como verificado pelas estatísticas acima.

O gráfico das previsões é dado abaixo:

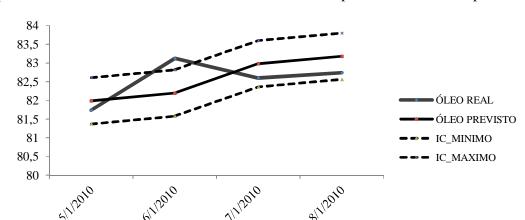


Figura 5.22: Gráfico do petróleo real e previsto 4 passos a frente com previsão estática pelo uso de redes neurais recorrentes com filtro *wavelets* primeiro e Kalman depois

A exemplo do cálculo feito para o IBOVESPA, fez-se as alterações na forma das wavelets para verificar se trariam algum benefício para a melhora na qualidade das previsões. Também chegou-se a conclusão que o tipo de wavelet altera o resultado. Os melhores resultados foram encontrados com as formas de wavelets primárias.

Foram selecionadas as mesmas 28 formas de onda e rodada a rede neural para previsão 4 passos a frente, fazendo uso da combinação de filtro de Kalman e de *Wavelet* e vice-versa. Dessa forma, ficou evidente que o vetor de *wavelet* depende da forma adequada de sua escolha.

E resumindo todo o processo de análise para o preço do barril de petróleo, podese notar que o uso combinado das técnicas de filtragem não colaborou muito para a redução dos erros e melhora do ajuste do modelo. A tabela a seguir resume todas as medidas, lembrando que volatilidade do petróleo de 21 dias foi de 12,04%.

Tabela 5.26 Resumo das estatísticas de previsão para o Petróleo

Medida	ARIMA	RN-	RN-	Filtro	FK_WAV	WAV_FK
	GARCH	RECORR	REC_WAV	Kalman		
TIC	0,005968	0,004164	0,003132	0,004192	0,004223	0,004392
MAPE	1,033%	0,66%	0,49%	0,70%	0,68%	0,60%
CORREL	0,395243	0,401368	0,498437	0,480144	0,54132	0,486431

Dessa forma, a aplicação das redes neurais em conjunto com o filtro de *wavelets* foi a combinação que trouxe melhores resultados medidos pelas estatísticas acima. Este resultado pode estar ligado a menor volatilidade da série da *commoditie* do que para um

índice de bolsa de valores. Muito embora, a correlação com o uso combinado do filtro de Kalman com posterior decomposição por *wavelets* foi o melhor resultado encontrado.

A seguir, apresenta-se as conclusões finais e propostas de novas pesquisas.

CONCLUSÃO

Na revisão da literatura, mencionou-se que os artigos que trabalham com previsão de séries temporais no clássico *International Journal of Forecasting* revisaram vinte e cinco anos de modelos, técnicas e combinação de resultados, que semearam ao longo desse período uma substancial contribuição para as decisões decorrentes dos modelos de previsão futura. Muito embora não existe ainda um circuito fechado e já devidamente testado e validado pelo mercado como o modelo a ser seguido para se realizar previsão de séries temporais financeiras.

Ressalta-se ainda que nem essa gigantesca presença de modelos de modelos presentes na literatura foi capaz de explicar crises no mercado financeiro e falhas nos modelos de previsão.

Mas, um consenso de mercado é que sem estes modelos seria muito pior. E foi com esta inspiração e motivação que se desenvolveu esta tese. Investiu-se no entendimento dos modelos e dos possíveis filtros empregados para redução de ruído e melhora nas previsões. Estudaram-se novas metodologias como filtros de Kalman e wavelets para fortalecer as técnicas já existentes, mas com grande campo de exploração. Assim, pretendeu-se nesta tese avançar no uso combinado da aplicação dos filtros com modelos de previsão como as redes neurais, que classicamente fornecem melhores resultados na previsão, mas com o objetivo de testar o efeito desta combinação de filtros na modelagem e previsão futura.

Resultados como uso de um único filtro já se encontrava presente na literatura que atestou a melhora na redução dos erros. Assim, surgiu a indagação desta pesquisa: qual a melhora que se teria se combinasse os filtros? Esta indagação alimentou e motivou a busca de recursos para se desenvolver essa junção de técnicas e metodologias para se concretizar uma pesquisa.

E em relação à questão buscada, encontrou-se suporte na literatura que já vinha apontando para a necessidade de melhoria nos filtros de volatilidade em conjunto com as técnicas já desenvolvidas e testadas. O *background* levantado apontou para a existência de dois principais filtros: *wavelets* e Kalman. Tais filtros foram aplicados em diversas séries com os mais distintos objetivos. Desde testes em imagens gráficas até

mesmo previsão de dados como correntes marítimas, fluxo de veículos e também na construção de modelos de previsão de séries temporais. E a principal conclusão dos trabalhos revisados foi a de que o uso dos filtros traz alguma melhora nos resultados pretendidos.

Em suma, juntando-se a curiosidade de pesquisador com as citações encontradas na literatura que se chegou à hipótese central desta tese: o uso combinado de filtros de wavelets e Kalman para redução de ruídos melhora a qualidade da previsão quando comparados com a simples aplicação de um só deles sem séries temporais financeiras. Posteriormente, incorporou-se na hipótese central a possibilidade de se trocar as formas de wavelets para ver o ganho obtido.

Assim, esta pesquisa teve o objetivo de estudar a realização das previsões dentro das séries temporais financeiras, comparando a aplicação de filtros de volatilidade para redução do ruído nos dados e consequente melhora na previsão.

Para se investigar o assunto, buscou-se uma metodologia quantitativa e descritiva dos modelos e das formas combinadas de uso dos filtros para previsão. Chegou-se a uma sequência de passos em um fluxograma que ajuda na transcrição das ideias e testes necessários para uso conjunto destas técnicas.

Quanto aos principais resultados desta tese, chegou-se a verificar que realmente o uso das técnicas de filtragem consegue reduzir o erro das previsões apenas na série do IBOVESPA que apresentou uma alta volatilidade no período. Já para a série de uma *commoditie* que, teoricamente, apresenta uma volatilidade menor, o uso combinado dos filtros não trouxe grande melhora na redução do erro, todavia, o erro foi menor quando do uso de uma das técnicas de filtragem foi empregada.

No caso, foram as *wavelets* que conseguiram reduzir o erro com uso de redes neurais recorrentes. Resultado este, que tem suporte na literatura existente e revisada nesta pesquisa. Dessa forma, cumpriu-se também o primeiro dos objetivos específicos a que se estabeleceu.

Nestas condições, os modelos de previsão de séries temporais com filtros de volatilidade sofrem influência da qualidade dos dados usados para previsão. Um dos elementos chaves nesse processo pode ter sido a forte influência das crises econômicas no período considerado, conforme destacado para o IBOVESPA. Tais janelas de

estudos podem ser incorporadas no intuito de revelar se para séries com comportamentos mais refinados, ou que não sofreram pressão de variáveis macroeconômicas, esse comportamento estaria presente, bem como a influência das formas de *wavelets*.

Criou-se ainda a expectativa de que as diferentes formas de *wavelets* pudessem ter alguma interferência nos resultados das previsões como o segundo objetivo específico deste trabalho.

Das formas de *wavelets* testadas, chegou-se ao resultado que o erro somente é inferior para as formas de onda primária, no caso para a *wavelet* de "Haar" e "daubesch 1". Nas demais, o erro se propaga. Observou-se ainda que a ordem que se obteve melhor resultado foi a de usar o filtro de Kalman primeiro e sobre ele, a decomposição de *wavelets*. Outro ponto relevante é quanto à decomposição das sub-séries onde a forma de onda interfere na quantidade de pontos nas sub-séries decompostas, captando mais ou menos pontos dependendo da sua arquitetura gráfica.

Cabe aqui ressaltar a forte influencia que as redes neurais recorrentes apresentaram no uso combinado de técnicas de filtragem, colaborando para a qualidade das previsões em se tratando de modelos não lineares. A arquitetura das redes neurais colabora no processo de iteração das variáveis no sentido de calcular melhor os graus (pesos) de relacionamento entre os vetores de entrada e saída da rede.

Dessa forma, este trabalho contribui para a área no sentido de que se abre um caminho para verificar esta aplicação em outras séries temporais financeiras, como câmbio, ações e outros ativos financeiros. Recomenda-se ainda, para pesquisas futuras, esta expansão da investigação, procurando inclusive séries de alta frequência, para atestar se os resultados persistem.

Ainda como fonte de inspiração para novas pesquisas, pode-se notar no trabalho que existe a flexibilidade de escolha de formas específicas de *wavelets*, então, porque não criar-se também combinações de formas de onda. Mesclar formas e equações para gerar novas *wavelets*.

Embora os resultados aqui encontrados tenham apontado para resultados diferentes, salienta-se a necessidade de maior investigação acerca desta combinação de técnicas e resultados. Desta forma, os resultados aqui encontrados não têm a pretensão

de sinalizar o encerramento das pesquisas, mas sim, colocar em discussão este tema crucial para a gestão de riscos e a precificação de operações no mercado financeiro e na área de controladoria e finanças.

Finalmente, acredita-se, de forma bem positiva que o caminho a ser seguido pelos modelos quantitativos de previsão seja mesmo o uso combinados de técnicas de modelagem, redução de ruídos, segmentação de dados, para que juntos, possam, cada um deles, dar sua contribuição para um mercado financeiro tão volátil e incerto. Assim, o conhecimento das peculiaridades dos modelos e das técnicas permitirá que as pesquisas avancem. De qualquer forma, continua sendo um campo para novamente ser seguido e refinado em pesquisas na área dos métodos quantitativos aplicados a finanças

REFERÊNCIAS BIBLIOGRÁFICAS

AIUBE, F. A. L. Modelagem dos preços futuros de commodities: abordagem pelo filtro de partículas. Pontifícia Universidade Católica do Rio de Janeiro. 2005, 183 f. Tese. Doutorado em Engenharia Industrial. 2005.

AHLBURG, D. Error measures and the choice of a forecast method. **International Journal of Forecasting**. n. 8, 1992, p. 99 – 111.

AMINGHAFARI, M. Forecasting time series using wavelets. **International Journal of Wavelets Multiresolution and Information Processing**. v. 5, n. 5, p. 709-724, 2007.

ANDERSEN, T. G.; BOLLERSLEV, T.; CHRISTOFFERSEN, P. F.; DIEBOLD, F. X. **Volatility and correlation forecasting**. In.: ELLIOT, G.; GRANGER, C. W. J.;

TIMMERMAN, A. **Handbook of economic forecasting**, Amsterdam: Nosth-Holland, p. 778-878, 2006.

ARIÑO, M. A. Time series forecasts with *wavelet*: an application to car sales in the spanish market. **Discussion Paper** 95-30, ISDS, Duke University, 1995.

BACHELIER, L. Théorie de la Speculation, **Annales de Lécole Normale Supérieure**, p. 21-86, 1900.

BEUREN, I. K. Como elaborar trabalhos monográficos em contabilidade. 3 ed. São Paulo: Atlas, 2006, 195 p.

BOLLERSLEV, T. Generalized autoregressive conditional heteroscedasticity. **Journal of Econometrics**. v. 31, p. 307-327, 1986.

BOUCHAUD, J.P.; POTTERS, M. Theory of Financial Risks: from Statistical Physics to Risk Management, **Cambridge University Press**, Cambridge, 2000.

BOX, G. E. P., JENKINS, G. M., REINSEL, G. C. **Time series analysis**: forecasting and control. 3. ed. New York: Prentice Hall, 1994.

BROCK, W.; HSIEH, D.; SCHEINKMAN, J. A test for independence based on the correlation dimension. **Econometric Reviews**, v. 15, n. 3, p. 197-235, 1996.

CAJUEIRO, D. O.; TABAK, B. M.; SOUZA, S. R. S. Investigação da memória de longo prazo na taxa de câmbio do Brasil. **Trabalhos para Discussão**, Banco Central do Brasil. v.113, 2006.

CAMPBELL, J.Y.; LO, A.W.; MACKINLAY, A. C. The econometrics of financial markets. New Jersey: Princeton University Press, 1997.

CHANG, C. **Análise de ondaletas em séries temporais**. São Paulo: IME, 1997. 117 f. Tese (Doutorado em Matemática Aplicada) – Instituto de Matemática e Estatística, Universidade de São Paulo, 1997.

CHRISTOFFERSEN, P. F.; DIEBOLD, F. X. How relevant is volatility forecasting for financial risk management? In.: *Review of Economics and Statistics*, v. 82, n.1, p. 12-22, 2000.

- CIARLINI, P., MANISCALCO, U. Wavelets and Elman neural networks for monitoring environmental variables. **Journal of Computational and Applied Mathematics**. n.221, p. 302-309, 2008.
- CORSINI, F. P., RIBEIRO, C. O. **Dinâmica e previsão de preços de commodities agrícolas com o filtro de Kalman.** Anais, XXVIII Encontro Nacional de Engenharia de Produção. Rio de Janeiro, 2008, 13p.
- COSTA, R.L.; VASCONCELOS, G.L. Long-range correlations and nonstationarity in the Brazilian stock market. **Physica A**, n. 329, p.231-248, 2003.
- DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. **Journal of the American Statistical Association**. v. 74, p. 427-431, 1979.
- DONOHO, D. L., TOHNSTONE, I. M., Ideal spatial adaptation by *wavelets* shrinkage. **Biometrika**, 81, 3. p.425 55.
- ENDERS, W. Applied econometric time series 2. ed. John Wiley & Sons, 2004.
- ENGLE, R. Autoregressive conditional heteroscedasticity with estimates of the variances of U.K. inflation. **Econometrica**, v. 50. p. 987-1008, 1982.
- FAMA, E. F. Efficiente capital markets: a review of theory and empirical work. **Journal of Finance**, Maio, p. 383-417, 1970.
- FAMA, E. F. The behavior of Stock Market Prices. **Journal of Business.** n 38, p. 34-105, 1965.
- GENÇAY, R.; SELÇUK, F.; WHITCHER, B. An introduction to wavelets and other filtering methods in finance and economics. New York: Academic Press, 2002.
- GOOIJER, J. G., HUNDMAN, R. J. 25 years of time series forecasting. **International Journal of Forecasting**. 22, p. 443-473, 2006.
- GRANGER, C. W. J. Forecasting stock market prices: lessons for forecasters. **International Journal of Forecasting**. n. 8., p. 3-13, 1992.
- GREWAL, M. S., ANDREWS, A. P. **Kalman Filtering:** theory and practice using matlab. 3. ed. John Wiley & Sons, 2008. 575 p.
- HARTER, I. B. Análise de precipitação em Pelotas-RS utilizando transformada wavelet de Morlet. Pelotas, 2004, 85f. Dissertação (Mestrado). Programa de Pós-Graduação em Meteorologia da Faculdade de Meteorologia. Universidade Federal de Pelotas. Pelotas, 2004.
- HARVEY, A. C. Forecasting, structural time series models and Kalman filter. Cambridge University Press. 2001, 554 p.
- HAYKIN, S. Redes neurais: princípios e prática. 2. ed. Porto Alegre: Bookman, 2001.
- HIRSCH, M. W, Convergent activation dynamics in continuous time networks. **Neural Networks**, v. 2, 1989, p. 331-349.

- HISIEH, D. Testing for nonlinear dependence in daily foreign exchange rates. **Journal of Business**, 62, n. 3, p. 339-368, 1989.
- HOMSY, G. V.; PORTUGAL, M. S.; ARAÚJO, J. P., Ondaletas e previsões de séries de tempo: uma análise empírica. **XXII Encontro Brasileiro de Econometria**. Campinas, 2000. 28p.
- HURST, H.E., Long term storage capacity of reservoirs. Transactions of the American society of civil engenieers, n.116, p.770-779, 1951.
- JARQUE, C. M., e BERA, A. K. A test for normality of observations and regression residuals. **International Statistical Review**. v. 55, p. 163-172, 1987.
- KALMAN, R. E. A new approach to linear filtering and prediction problems. **Trans ASME J. Basic Eng.**, 82, p. 35-45, 1960.
- KWIATKOWSK, D. PHILLIPS, P. C. B., SCHIMIDT, P., SHIN, Y. Testing the null hypothesis of stationary against the alternative of a unit root. **Journal of Econometrics**. v. 54, p. 159-178, 1992.
- LIMA, F. G. Um método de análise e previsão de sucessões cronológicas unidimensionais lineares e não-lineares. Faculdade de Economia, Administração e Contabilidade. São Paulo. 228 f. Tese (Doutorado em Administração). Faculdade de Economia, Administração e Contabilidade: Universidade de São Paulo, 2004.
- LIPSCHUTZ, Seymour. **Probabilidade**. 4. ed. São Paulo: Makron Books, 1993.
- LO, A.W, MACKINLAY, A.C. Long-term memory in stock market prices, **Econometrica**, v.59, p.1279-1313, 1999.
- LYNCH, S. **Dynamical systems with applications using matlab.** Brikhãuser, 2004. 459 p.
- MA, I., WONG, T., SANKAR, T. Forecasting the volatility of a financial index by wavelet transform and evolutionary algorithm. **IEEE International Conference on Systems, Man & Cibernetics**. P. 5824-5829, 2004.
- MANDELBROT, B., HUDSON, R. **Mercados financeiros fora de controle**: a teoria dos fractais explicando o comportamento dos mercados. Tradução de Afonso Celso da Cunha Serra. Rio de Janeiro: Elsevier, 2004.
- MANTEGNA,R.; STANLEY, H.E. An Introduction to Econophysics: Correlations and Complexity in Finance, Cambridge University Press, Cambridge, 1999.
- MCLEOD, A. J. e W. K. Li, Diagnostic checking ARMA times series models using squared residuals correlations. **Journal of Time Series Analysis**, n. 4, 1983, p. 269-73.
- MILLS, T. C. The econometric modeling of financial time series. New York: Cambridge Universisty Press, 1996.
- MISITI, M. et al. **Wavelet toolbox**: for use with Matlab. The Math Works, Inc. 1997. Disponível em: http://www.mathworks.com. Acesso em: 14 ago. 2003.
- MISITI, M. Wavelets and their applications. John Wiley, 2007, 330 p.

MOODY, J. e SAFFELL, M. Reinforcement learning for trading system and portfolios. In: AGRAWAL, R.; PIATETSKY-SHAPIRO, G. Ed. KDD98 - **Proceedings of The Fourth International Conference on Knowledge Discovery and Data Mining**. AAAI Press, California, 1998, p. 279-283.

MORAIS, I. A. C., PORTUGAL, M. S. **Modelagem e previsão de volatilidade determinista e estocástica para a série do IBOVESPA**. 1999. Disponível em: http://www.ufrgs.br/ppge>. Acesso em: 20 jul. 2003.

MORETTIN, P. A. Ondas e ondaletas. São Paulo: Edusp, 1999.

_____. **Econometria Financeira**. São Paulo: Edgard Blucher, 2008.

MORETTIN, P. A.; TOLOI, C. M. de C. **Análise de séries temporais**. São Paulo: Edgard Blucher, 2004.

OLIVEIRA, M. A. de. **Aplicação de redes neurais na análise de séries temporais econômico financeiras**. Faculdade de Economia, Administração e Contabilidade. São Paulo, 2007. 285 f. Tese (Doutorado em Administração). Faculdade de Economia, Administração e Contabilidade: Universidade de São Paulo, 2007.

OLIVEIRA, M. A. de. Previsão de sucessões cronológicas econômico-financerias por meio de redes neurais artificiais recorrentes de tempo real e de processos ARMA-GARCH: um estudo comparativo quanto à eficiência de previsão. Faculdade de Economia, Administração e Contabilidade. São Paulo, 2003. 171 f. Dissertação (Mestrado em Administração). Faculdade de Economia, Administração e Contabilidade: Universidade de São Paulo, 2003.

OLIVEIRA, M. V. de. **Metodologia para validação de sinal usando modelos empíricos com técnicas de inteligência artificial aplicada a um reator nuclear.** Universidade Federal do Rio de Janeiro. Tese. Doutorado em Ciências. 307 f. 2005.

OPEP. **Organization of the Petroleum Exporting Countries**. Disponível em http://www.opec.org/opec_web/en/data_graphs/40.htm. Acesso em 20/02/2010. 2010.

PANKRATZ, A.; DUDLEY, U. Forecasts of power-transformed series. **Journal of Forecasting**. 6, 1987. p. 239-248.

PEGELS, C. C. Exponencial Smoothing: some new variations. **Management Science**, v. 12, 1969, p. 311-315.

PERCIVAL, D. B., WALDEN, A. T. Wavelet methods for time series analysis. Canbridge, 2006.

PEREIRA, L. M. Modelo de formação de preços de commodities agrícolas aplicado ao mercado de açúcar e álcool. Faculdade de Economia, Administração e Contabilidade. São Paulo, 2009. 209f. Tese. Doutorado em Administração., 2009.

PHILLIPS, P. C. B.; PERRON, P. Testing for a unit root in time series regression. **Biometrila**. v. 75, p. 335-346, 1988.

POINCARÉ, H. **Science and Method**, New York: Dover Press. (Originally published in 1908), 1952.

- POLIKAR, R. **The wavelet tutorial.** 1999. Disponível em: < http://users.rowan.edu/~polikar/WAVELETS/WTtutorial.html >. Acesso em 10 de jan. 2010.
- POSTALCIOGLU, S., ERKAN, K., BOLAT, E. D. Comparison of Kalman filter and wavelets filter for denoising. **IEEE**, v.3, 2005, p. 951-953.
- PUSKORIUS, G. V., FELDKAMP, L. A. Neuro control of nonlinear dynamical systems with Kalman filter trained recurrent networks. **IEEE Transactions on Neural Networks**. v.5, p. 271-297m 1994.
- RENAUD, O., STARCK, J., MURTAGH, F. Wavelet-based combined signal filtering and prediction. **IEEE Transactions SMC**, Part B, v. 35, n.6, 2005, p. 1241 1251.
- ROCHA, V. B. Uma abordagem de *wavelets* aplicada a combinação de previsões: uma análise teórica e experimental. Dissertação: Mestrado em Ciências. Universidade Federal do Paraná, 155 f, 2008.
- SABINO, P. A. A.; BRESSAN, A. A. Avaliação do Desempenho Preditivo de Modelos Para A Volatilidade: Um Estudo Para O Mercado Acionário Brasileiro Entre 2000 e 2009. In: XII SEMEAD, 2009, São Paulo. Anais do XII SEMEAD. São Paulo: USP, 2009. v. Único.
- SHAH, S., PALMIERI, F. A fast, local algorithm for training feedforward neural networks, **Proceedings of International Joint Conference on Neural Networks**. San Diego, CA, v. 3, p. 41-46, 1990.
- SINGHAL, S., WU, L. Training feed-forward networks with extended Kalman filter. **IEEE Transactions on Acoustics, Speech, and Signal Processing**. P. 1187-1190, 1989.
- SOARES, A. S. **Aproximação de nuvens de pontos de dados por meio de superfícies de Bézier.** Universidade Federal de Uberlância. Tese. Doutorado em Engenharia Elétrica. 240 f. 2007.
- SOUZA, R. C. Modelos estruturais para previsão de séries temporais: abordagem clássica e bayesiana. Rio de Janeiro: IMPA, 1989.
- Previsão de séries temporais utilizando rede neural treinada por filtro de kalman e evolução diferencial. Dissertação. (Mestrado em Engenharia de Produção). Pontifícia Universidade Católica do Paraná. 85 p. 2008.
- TAK, B. A new method for forecasting stock prices using artificial neural network and ondaleta theory. 1995. 107 f. Tese (Doutorado em Economia). Universidade de Pensilvânia, Estados Unidos, 1995.
- TANGBORN, A., ZHANG, S. Q. Wavelet transform to an approximate Kalman filter system. **Applied Numerical Mathematics.** n. 33, 2000, p. 307-316.
- TONG, H.; LIM, K. S. Threshold autoregressions, limit cycles, and data, **Journal of Royal Statistical Society**, n. 42, p. 245-292, 1980.
- TSAY, R. Analysis of financial time series. Wiley series in probability and statistics. 2. ed, 2005.

Time series and forecasting: brief history and future research. **Journal of the American Statistical Association**. v. 95, n. 450, 2000, p. 638-643.

UKIL, A., ZIVANOVIC, R. Adjusted Haar wavelet for application in the power systems disturbance analysis. **Digital Signal Processing**. n. 18, p. 103-115, 2001.

WELCH, G., BISHOP, G. An introduction to the Kalman filter. Technical Report TR – 95 – 041, Departament of Computer Science. University of North Caroline at Chapel Hi, USA, 2006, 16p.

WILLIAMS, R. J. Training recurrent networks using the extended Kalman filter. **Proceedings of International Joint Conference on Neural Networks**. Baltimore, v. 4, p. 241-247, 1992.

WILLIAMS, R. J.; ZIPSER, D. A learning algorithm for continually running fully recurrent neural networks. **Neural Computation**. v. 1, p. 270-280. 1989.

WONG, H., IP. W., XIE, Z., LUI, X.. Modelling and forecasting by wavelets, and the application to exchange rates. **Journal of Applied Statistics**, v. 30, n. 5, 2003. p. 537-553.

XIE, Y. ZHANG, Y., YE, Z. Short-term traffic volume forecasting using Kalman filter with discrete wavelet decomposition. **Computer Aided Civil and Infrastrucuture engenieering**. n. 22, p. 326-334, 2007.

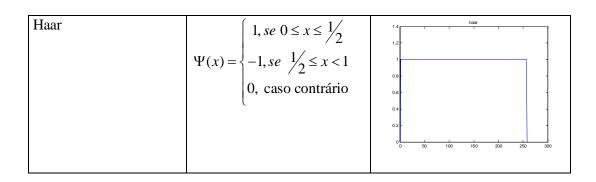
YULE, G. U. On the method for investigating periodicities in disturbe series, with special reference to Wölfer's sunspot numbers. **Philosophical Transactions of the Royal Society London, Series A**. 226, 1927, p. 267-298.

ZHANG, G.; PATUWO, B. E.; HU, M. Y. Forecasting with artificial neural networks: the state of the art. **International Journal of Forecasting**, Kent(Ohio) 14, p. 35–62, 1998.

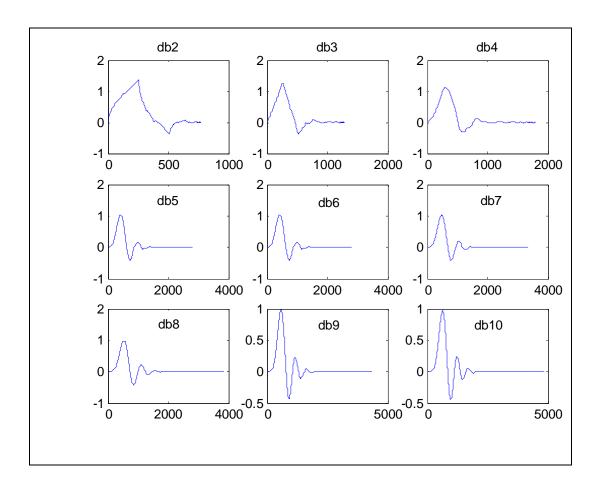
ZANDONADE, E; MORETTIN, P. A. . Wavelets in State Space Models. Applied Stochastic Models in Business and Industry, Belgica, v. 15, p. 199-219, 2003.

APÊNDICE A – Formas de *wavelets*

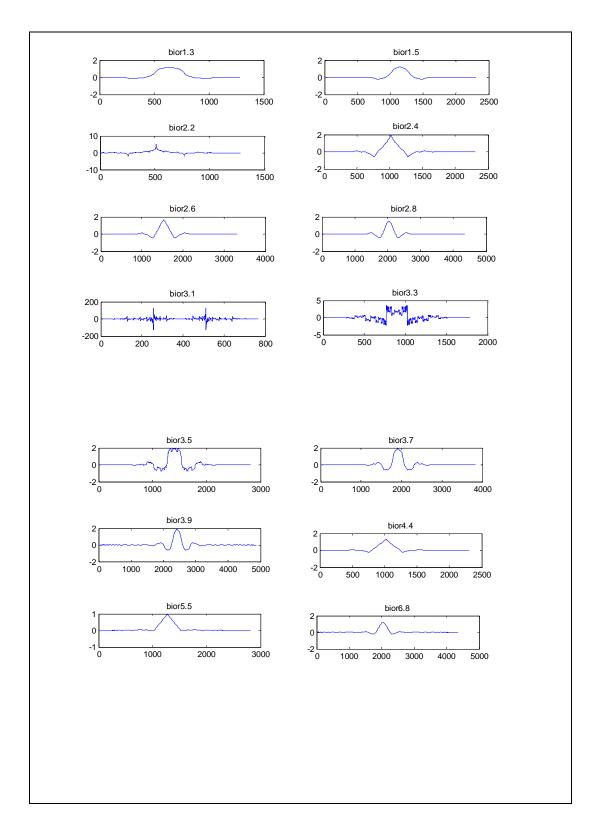
(Misiti et al., 1997)



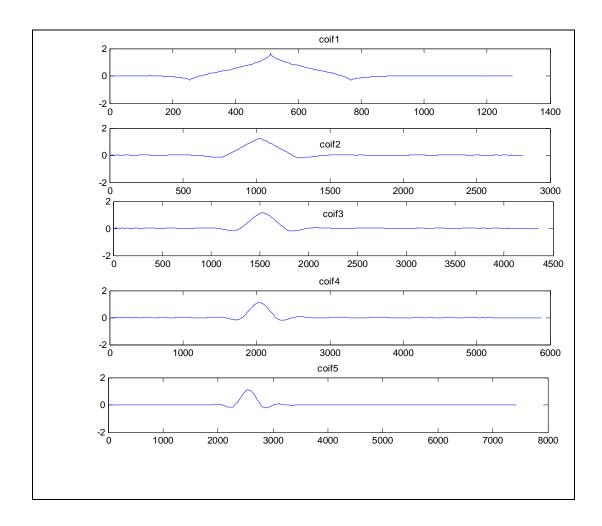
Daubechies	



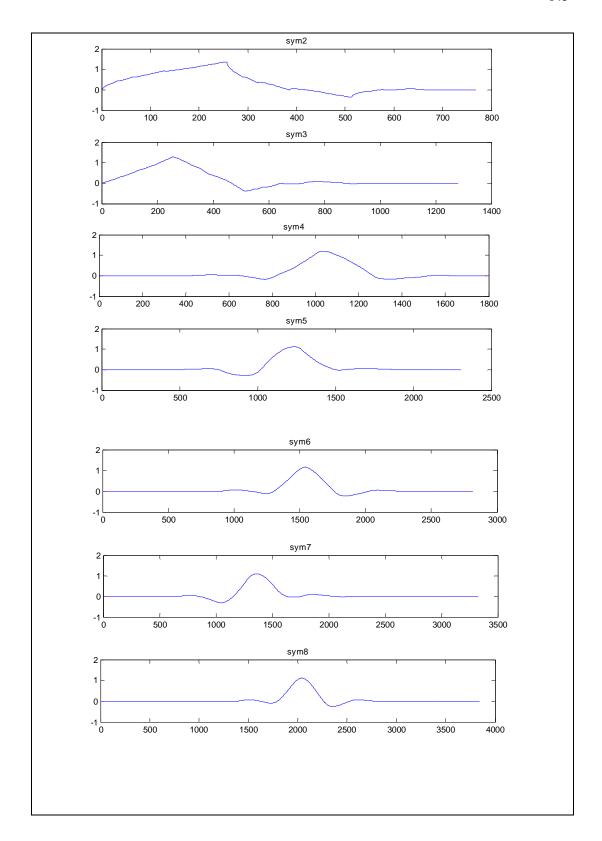
Dianthaganal	
Biorthogonal	



Coiflets	



Symlets	



Morlet	$\Psi(x) = Ce^{-\frac{x}{2}}.\cos(5x)$ sendo que C é uma constante.	0.8- 0.0- 0.4- 0.2- 0.4- 0.6- 0.8- -10 SO 100 1SO 200 250 300
Mexican Hat	$\Psi(x) = \left(\frac{2}{\sqrt{3}}\pi^{-\frac{1}{4}}\right) (1 - x^2) e^{-\frac{x}{2}}$	0.8- 0.6- 0.4- 0.2- 0.2- 0.4- 0.2- 0.4- 0.50 100 150 200 250 300
Meyer		12 1 0.8 0.6 0.4 0.2 0.2 0.3 0.4 0.50 100 150 200 250 300

APÊNDICE B – Processo de Markov

PROCESSO DE MARKOV

O objetivo deste anexo é o de demonstrar as definições e propriedades elementares sobre os processos de Markov exigidos para a completa compreensão desta tese. O conteúdo desse apêndice foi adaptado de Lipschutz (1993, p 207).

Entende-se por um vetor, r, simplesmente uma n-upla de números:

$$r = (r_1, r_2, r_3, \cdots, r_n)$$

Os r_i são chamados **componentes** de r. Se acontecer de todos os $r_i = 0$, então r é o **vetor nulo**. E dois vetores são iguais se e somente se suas componentes são iguais.

Uma $\mathbf{matriz}\ A$ é um quadro retangular de números organizados e dispostos mediante linhas e colunas:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

As *m n*-uplas horizontais

$$(a_{11} \quad a_{12} \quad \cdots \quad a_{1n}), (a_{21} \quad a_{22} \quad \cdots \quad a_{2n}), \cdots, (a_{m1} \quad a_{m2} \quad \cdots \quad a_{mn})$$

são chamadas **linhas** de A, e as *n m*-uplas verticais, suas colunas.

$$\begin{pmatrix} a_{11} \\ a_{21} \\ \dots \\ a_{m1} \end{pmatrix}, \begin{pmatrix} a_{12} \\ a_{22} \\ \dots \\ a_{m2} \end{pmatrix}, \dots, \begin{pmatrix} a_{1n} \\ a_{2n} \\ \dots \\ a_{mn} \end{pmatrix}$$

Note que o elemento a_{ij} , chamado a ij-ésima entrada, aparece na interseção da i-ésima linha com a j-ésima coluna. Representa-se tal matriz por $A = (a_{ij})$. Uma matriz com m linhas e n colunas é chamada matriz m por n e escreve-se $m \times n$. Se o número de linhas for igual ao número de colunas então é chamada de matriz quadrada. Observe que a matriz com uma única linha pode ser considerada como um vetor, e vice-versa.

Considere agora duas matrizes A e B tais que o número de colunas da matriz A é igual ao número de linhas da matriz B, isto é, A é uma matriz $m \times p$ e B é uma matriz $p \times n$. O produto de A por B, denotado por AB, é a matriz $m \times n$ cuja entrada ij é obtida pela multiplicação dos elementos da i-ésima linha de A pelos elementos correspondentes da j-ésima coluna de B e somando esses produtos.

$$\begin{pmatrix} a_{11} & \cdots & a_{1p} \\ \cdots & \cdots & \cdots \\ a_{i1} & \cdots & a_{ip} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mp} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1j} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ b_{p1} & \cdots & b_{pj} \end{pmatrix} \cdots b_{pn} = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ c_{m1} & \cdots & c_{mn} \end{pmatrix}$$

onde
$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{ip}b_{pj} = \sum_{k=1}^{p} a_{ik}b_{kj}$$
.

Se o número de colunas da matriz A não for igual ao número de linhas de B, então o produto AB não é definido.

Além disso, se u é um vetor com n componentes, pode-se formar o produto uA, que é também um vetor com n componentes. Chama-se $u \neq 0$ um **vetor fixo** (ou ponto fixo) de A, se u é "fixo a esquerda", isto é, não se altera quando multiplicado por A, ou seja uA = u.

Numericamente, pode-se considerar o seguinte exemplo: seja a matriz $A = \begin{pmatrix} 2 & 1 \\ 2 & 3 \end{pmatrix}$, o vetor u = (2, -1) é um ponto fixo de A, pois

$$uA = (2,-1)\begin{pmatrix} 2 & 1 \\ 2 & 3 \end{pmatrix} = (2.2-1.2, 2.1-1.3) = (2,-1) = u.$$

Um vetor $u=(u_1,u_2,\cdots,u_n)$ é chamado **vetor de probabilidade**, se suas componentes forem não negativas e somarem 1. Por exemplo, $u=\left(\frac{1}{4},\frac{1}{4},0,\frac{1}{2}\right)$ é um vetor de probabilidade.

Uma matriz quadrada $P = (p_{ij})$ é chamada **matriz estocástica**, se cada uma das linhas for um vetor de probabilidade, isto é, se cada entrada de P é não negativa e a soma das entradas em cada linha for 1.

Por exemplo,
$$P = \begin{pmatrix} 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$
 é uma matriz estocástica.

Se A e B forem matrizes estocásticas, então o produto AB dessas matrizes estocásticas é uma matriz estocástica, e consequentemente, todas as potências de A^n são matrizes estocásticas.

Uma matriz estocástica P é considerada **regular** se todas as entradas de alguma potência P^n são positivas. Suas propriedades são:

- (i) P tem um único vetor fixo t de probabilidade e os componentes de t são todos positivos;
- (ii) As entradas das potências P, P^2, P^3, \cdots de P convergem para as entradas correspondentes da matriz T, cujas linhas são todas iguais ao vetor fixo t;
- (iii) Se p é qualquer vetor de probabilidade, então as sequências de vetores $p^{P}, p^{P^{2}}, p^{P^{3}}, \cdots$ converge para o ponto fixo t.

Por exemplo, considere a matriz estocástica regular $P = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$. Pode-se

determinar o vetor de probabilidade com duas correspondentes que se pode representar por t = (x, 1-x), de modo que tP = t:

$$(x,1-x)\begin{pmatrix} 0 & 1\\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = (x,1-x)$$

Efetuando-se a multiplicação das matrizes do lado esquerdo da equação acima, obtém:

$$\left(0+\frac{1}{2}-\frac{x}{2},\frac{1}{2}+\frac{x}{2}\right)=(x,1-x)$$

Resolvendo as igualdades
$$\begin{cases} \frac{1}{2} - \frac{x}{2} = x \\ \frac{1}{2} + \frac{x}{2} = 1 - x \end{cases}$$
, tem-se que $x = \frac{1}{3}$.

Assim, $t = \left(\frac{1}{3}, \frac{2}{3}\right)$ é o único ponto fixo de probabilidade de *P*. Pela definição

dada acima, a sequência de vetores P, P^2, P^3, \cdots converge para a matriz

$$T = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} = \begin{pmatrix} 0.33 & 0.67 \\ 0.33 & 0.67 \end{pmatrix}, \text{ cujas linhas são o vetor } t.$$

De fato, apresenta-se a seguir, algumas das potências de P, para indicar o resultado descrito anteriormente:

$$P^{2} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix} = \begin{pmatrix} 0,50 & 0,50 \\ 0,25 & 0,75 \end{pmatrix}$$

$$P^{3} = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{3}{8} & \frac{5}{8} \end{pmatrix} = \begin{pmatrix} 0,25 & 0,75 \\ 0,37 & 0,63 \end{pmatrix}$$

$$P^{4} = \begin{pmatrix} \frac{3}{8} & \frac{5}{8} \\ \frac{5}{16} & \frac{11}{16} \end{pmatrix} = \begin{pmatrix} 0.37 & 0.63 \\ 0.31 & 0.69 \end{pmatrix}$$

$$P^{5} = \begin{pmatrix} \frac{5}{16} & \frac{11}{16} \\ \frac{11}{32} & \frac{21}{32} \end{pmatrix} = \begin{pmatrix} 0.31 & 0.69 \\ 0.34 & 0.66 \end{pmatrix}$$

Considere agora uma sequência de valores cujos resultados x_1, x_2, \cdots , satisfazem as seguintes propriedades:

- (i) Cada resultado pertence a um conjunto finito de resultados (a_1, a_2, \dots, a_n) , chamado o **espaço dos estados** do sistema; se o resultado da n-ésima tentativa é a_i diz-se que o sistema de encontra no estado a_i no instante n.
- (ii) O resultado de qualquer ensaio depende no máximo do resultado do ensaio imediatamente anterior e não de qualquer outro dos precedentes; a cada par de estados (a_i, a_j) está associada à probabilidade P_{ij} de que a_j ocorre imediatamente após ter ocorrido a_i .

O processo estocástico com as propriedades acima é chamado **processo de Markov**. Os números p_{ij} são chamados **probabilidades de transição** e podem ser expressos segundo a matriz

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix}$$

chamada matriz de transição que é uma matriz estocástica.

Geralmente os preços das ações são modelos segundo um processo de Markov uma vez que as previsões futuras sobre as oscilações dos preços desta ação não devem levar em conta as oscilações ocorridas no dia anterior, no mês anterior ou no ano anterior.

Normalmente os preços futuros são expressos em termos da distribuição de probabilidade. A propriedade de Markov nos diz que a distribuição de probabilidades dos preços em qualquer tempo depende única e exclusivamente do preço atual da ação.

A propriedade de Markov para o preço das ações é consistente com a forma fraca da eficiência de mercado, que diz que o preço atual de uma ação já reflete plenamente todas as informações que estão contidas na sequência histórica do preço.

APÊNDICE C – Programa para rodar a rede recorrente.

```
APRENDIZAGEM RECORRENTE EM TEMPO REAL (ARTR)
%
                      Fonte: Oliveira (2003).
%
                            ALGORITMO - ARTR
load c:\bilinear\Inputs.txt
load c:\bilinear\DesiredOutputs.txt
H = 4;InitialState = [0.0 zeros(1,H)];
sigmoid_type = 0;
alpha = 1.0;
learning_rate = 0.005;
epochs = 20000; epsilon = 0.001; Inputs = Inputs(801:1000);
DesiredOutputs = DesiredOutputs(801:1000);
[WeightMatrix]
artr_treino(Inputs,DesiredOutputs,H,InitialState,sigmoid_type,al
pha,learning_rate,epochs,epsilon);
[Outputs]
artr_perf(Inputs,InitialState,H,WeightMatrix,sigmoid_type,alpha)
[samples columns] = size(Inputs);
t = (1:samples)';
figure
hold on
plot(t,DesiredOutputs,'b-.');
plot(t,Outputs,'r');
legend('Valor Desejado','Valor Previsto',0)
xlabel('Janela de Previsao')
ylabel('Serie d(t) versus Serie y(t)')
figure
hold on
plot(t, DesiredOutputs - Outputs)
legend('Erro de previsao',0)
xlabel('Janela de Previsao')
ylabel('Erro de previsao, e(t)')
figure
hold on
plot(t, Outputs)
legend('Valores Previstos',0)
xlabel('Janela de Previsao')
ylabel('Valores Previstos, e(t)')
wk1write('c:\bilinear\bilinearrec',Outputs);
figure
hold on
histfit(DesiredOutputs - Outputs)
e = DesiredOutputs - Outputs
perf_mae = mae(e)
% perf2_mse = mse(e)
rmse = sqrt(mse(e))
eq = (DesiredOutputs - Outputs)*(DesiredOutputs - Outputs)'
dq = (DesiredOutputs)*(DesiredOutputs)'
oq = (Outputs)*(Outputs)'
deq = diag(eq)
ddq = diaq(dq)
doq = diag(oq)
```

```
theil = (sqrt(sum(diag(eq))/10))/((sqrt(sum(diag(oq))/10))+(sqrt(sum(diag(q))/10)))
```