

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA**

ANDRÉ BARBOSA OLIVEIRA

**Usando Redes Neurais para Estimação da Volatilidade: Redes Neurais
e Modelo Híbrido GARCH Aumentado por Redes Neurais**

PORTO ALEGRE

2010

ANDRÉ BARBOSA OLIVEIRA

Usando Redes Neurais para Estimação da Volatilidade: Redes Neurais e Modelo Híbrido GARCH Aumentado por Redes Neurais

Dissertação apresentada ao programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da Universidade Federal do Rio Grande do Sul como parte dos requisitos para a obtenção do título de Mestre em Economia, ênfase em economia aplicada.

Orientador: Prof.º Pós-Doutor Flávio Augusto Ziegelmann

PORTO ALEGRE

2010

O48u

Oliveira, André Barbosa

Usando redes neurais para estimação da volatilidade: redes neurais e modelo híbrido GARCH aumentado por redes neurais / André Barbosa Oliveira. – Porto Alegre, 2010.

73 f.

Orientador: Flávio Augusto Ziegelmann.

Ênfase em Economia Aplicada.

Dissertação (Mestrado em Economia) - Universidade Federal do Rio Grande do Sul, Faculdade de Ciências Econômicas, Programa de Pós-Graduação em Economia, Porto Alegre, 2010.

1. Mercado financeiro: Volatilidade. 2. Estatística aplicada.

I. Ziegelmann, Flávio Augusto. II. Universidade Federal do Rio Grande do Sul. Faculdade de Ciências Econômicas. Programa de Pós-Graduação em Economia. III. Título.

CDU 519.23

ANDRÉ BARBOSA OLIVEIRA

Usando Redes Neurais para Estimação da Volatilidade: Redes Neurais e Modelo Híbrido GARCH Aumentado por Redes Neurais

Dissertação apresentada ao programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da Universidade Federal do Rio Grande do Sul como parte dos requisitos para a obtenção do título de Mestre em Economia, ênfase em economia aplicada.

Orientador: Prof.º Pós-Doutor Flávio Augusto Ziegelmann

Aprovada em: Porto Alegre, 31 de Março de 2010.

Prof.º Doutor Marcelo Savino Portugal
PPGE/UFRGS

Prof.º Doutor Cléber Bisognin
IM/UFRGS

Prof.º Doutor Márcio Poletti Laurini
IBMEC/SP

Agradecimentos

Agradeço primeiramente as pessoas que foram mais importantes para a minha formação acadêmica ao longo da graduação me direcionando ao mestrado e posteriormente aos que foram decisivos para alcançar o título de mestre. Muito esforço, noites mal dormidas (ou não dormidas), cansaço, mas mesmo assim a cada momento uma euforia frente ao novo aprendizado e domínio do conhecimento desta área fascinante que é a economia, em especial a econometria e séries temporais.

Agradeço primeiramente a Professora Hildete, minha amiga e orientadora de iniciação científica na Universidade Federal Fluminense (UFF), que me iniciou e ensinou boa parte do que sei de pesquisa em economia. Seu gosto e paixão pela ciência econômica ajudando a entender e explicar fenômenos sociais é contagiante. Ainda hoje posso desfrutar de sua amizade e mesmo ser co-autor em artigos.

Ao curso CATE, coordenado pelo Hugo Boff no IE da UFRJ, atribuo grande aprendizado em economia que me permitiram ingressar no mestrado.

A CAPES minha gratidão pelo apoio financeiro com a bolsa de mestrado que me permitiu estudar e me manter em outro estado, sem este apoio o mestrado seria inviável. Este programa de apoio ao estudante é estratégico para a formação de quadros de profissionais qualificados para o Brasil.

Ao Programa de Pós Graduação em Economia da UFRGS tenho muito a agradecer. Ao professor Flávio Ziegelmann sou eternamente grato pela enorme paciência na orientação e debates onde pude aprender muito; seu rigor, formalidade e amplo conhecimento permitiram esclarecer muitas questões que passariam despercebidas, e são motivação para aumentar a qualidade do trabalho e realizar a pesquisa com testes mais rigorosos. Na UFRGS tive grande aprendizado em economia, especialmente econometria, meus agradecimentos ao Marcelo Portugal pelas excelentes aulas em diversas disciplinas e enorme destreza e experiência tanto no ensino quanto na pesquisa, bem como agradeço aos demais professores da pós-graduação: Cléber Bisognin, Fabrício, Sabino, Hélio Henkin, Eduardo Pontual, Ronald, Giácomo e Eugenio Lagemann.

A minha família ao carinho e compreensão pela ausência permanente, mas onde a amizade e a união nunca faltam.

Meus agradecimentos ao pronto atendimento e gentileza das secretárias da pós-graduação: Iara, Aline, Fátima, Raquel, Claudia e Lourdes. Ao pessoal da xerox Speed também sou grato pelo atendimento especial aos alunos da pós-graduação.

Resumo

As séries temporais financeiras são marcadas por comportamentos complexos e não-lineares. No mercado financeiro, além da trajetória das cotações, a sua variabilidade, representada pela volatilidade, consiste em importante informação para o mercado. Redes neurais são modelos não lineares flexíveis com capacidade de descrever funções de distintas classes, possuindo a propriedade de aproximadores universais. Este trabalho busca empregar redes neurais, especificamente Perceptron de múltiplas camadas com uma única camada escondida alimentada para frente (*Feedforward Multilayer Perceptron*), para a previsão da volatilidade. Mais ainda, é proposto um modelo híbrido que combina o modelo GARCH e redes neurais. Os modelos GARCH e redes neurais são estimados para duas séries financeiras: Índice S&P500 e cotações do petróleo tipo Brent. Os resultados indicam que a volatilidade aproximada por redes neurais é muito semelhante as estimativas dos tradicionais modelos GARCH. Suas diferenças são mais qualitativas, na forma de resposta da volatilidade estimada a choques de maior magnitude e sua suavidade, do que quantitativas, apresentando critérios de erros de previsão em relação a uma medida de volatilidade benchmark muito próximos.

Palavras Chave: Volatilidade. Redes Neurais. Modelos GARCH.

Abstract

The financial time series are characterized by complex and non-linear behaviors. In addition to the financial market trend in prices their variability or volatility, a risk estimate, is important information for the market players. Neural networks are flexible nonlinear models capable of describing functions of different classes, having the property of universal approximators. This paper employs neural networks, specifically one hidden layer feedforward Multilayer Perceptron, for volatility forecasting. Moreover, we propose a hybrid model that combines the GARCH model with neural networks. The GARCH and neural network models are estimated over two financial series: the S&P500 composite index and prices of Brent oil. The results indicate that the volatility approximated by neural networks is very similar to that estimated by the traditional GARCH models, while their differences are more qualitative than quantitative, with information content that differs from and complements each other for different market environments.

Keywords: Volatility. Artificial Neural Networks. GARCH Models.

Lista de Figuras

Figura 1 - Modelo de Perceptron de Múltiplas Camadas	16
Figura 2 - Aprendizado e Generalização por Redes Neurais.....	22
Figura 3 - Gráfico da Função Logística.....	23
Figura 4 – Rede Neural com Generalização Limitada (<i>Over-fitting</i>).....	26

Lista de Gráficos

Gráfico 1 – Aproximação da Função $y = \text{sen}(x)$ com ruído, pela rede neural NN(1,3,1).	24
Gráfico 2 - Neurônios da rede neural NN(1,3,1) usada para aproximar a função $y = \text{sen}(x)$	25
Gráfico 3 – Índice e Retornos: S&P 500.	39
Gráfico 4 – Índice e Retornos: Cotações Petróleo Brent.	40
Gráfico 5 – Histograma: Retornos S&P 500 e Retornos Brent.	42
Gráfico 6 – Função de Autocorrelação para Quadrados dos Retornos: S&P 500 e Brent	43
Gráfico 7 – Curva de Chegada de Informação GARCH e TGARCH: Retornos S&P 500	45
Gráfico 8 – Curva de Chegada de Informação GARCH e TGARCH: Retornos Brent..	46
Gráfico 9 – Curva de Chegada de Informação Aproximada por Redes Neurais: Retornos S&P500 e Retornos Brent	48
Gráfico 10 – Intervalos de Confiança GARCH(1,1) e Redes Neurais para Retornos S&P 500 (Período de 10/01/2008 a 08/04/2008).	54
Gráfico 11 – Intervalos de Confiança GARCH(1,1) e Redes Neurais para Retornos Brent (Período de 1/12/2008 a 26/02/2009).	55
Gráfico 12 – Intervalos de Confiança HNN-GARCH(1,1,4) para Retornos S&P 500 (Período de 10/01/2008 a 08/04/2008).	63
Gráfico 13 – Intervalos de Confiança HNN-GARCH(1,1,5) para Retornos Brent (Período de 1/12/2008 a 26/02/2009).	63

Lista de Tabelas

Tabela 1 – Estatísticas Descritivas: Retornos S&P 500 e Retornos Brent.	42
Tabela 2 – Modelos GARCH e TGARCH para os Retornos S&P 500.....	44
Tabela 3 – Modelos GARCH e TGARCH para os Retornos Brent.	46
Tabela 4 – Especificações dos Modelos de Redes Neurais para Estimação da Volatilidade: S&P500 e Brent	49
Tabela 5 - Critérios de Desempenho Preditivo GARCH(1,1) e Redes Neurais para as Séries de Retornos S&P 500 e Brent.	51
Tabela 6 – Estimação do Modelo HNN-GARCH para as Séries de Retornos: S&P500 e Brent	61
Tabela 7 – Critérios de Desempenho Preditivo Modelo HNN-GARCH para as Séries de Retornos S&P500 e Brent	62

Sumário

1 Introdução	12
2 Redes Neurais Artificiais (RNA)	14
2.1 Modelos de Redes Neurais Artificiais	14
2.2 Redes Neurais para Previsão de Séries de Tempo.....	19
2.3 Redes Neurais & Interpolação	21
3 Fatos Estilizados das Séries Temporais Financeiras e Modelos de Volatilidade .	28
3.1 Não Linearidades dos Retornos Financeiros	28
3.2 Modelos de Volatilidade.....	30
3.3 Estimação da Volatilidade por Redes Neurais e Modelos GARCH: Um Estudo Empírico	37
4 Modelo Híbrido GARCH Aumentado por Redes Neurais	56
4.1 Modelo Artificial Neural Network–GARCH	56
4.2 Modelo Híbrido GARCH Aumentado por Redes Neurais	58
4.3 Estimação do Modelo Híbrido HNN-GARCH para a Série de Retornos S&P500 e Brent	60
5 Considerações Finais	65
Referências	67
Apêndice A – Algoritmo de Retro-propagação	72

1 Introdução

Redes neurais artificiais são modelos matemáticos flexíveis com inspiração biológica, baseados na forma de processamento de informação no cérebro humano. Redes neurais artificiais têm sido tema com grande estudo e aplicações variadas nos últimos anos, apresentando bom desempenho em diversas análises e setores entre os quais setor financeiro (HAMID, IQBAL, 2004). Redes neurais são empregadas em geral para classificação e reconhecimento de padrões.

Uma área de grande aplicação de redes neurais é em previsões de séries temporais financeiras, as quais são caracterizadas por comportamentos complexos e não lineares. Redes neurais artificiais são treinadas com base em um conjunto de dados para apreender uma relação e reproduzi-la, fazendo um mapeamento de uma relação funcional entre as entradas e saídas da rede. Este trabalho aplica redes neurais para previsão da volatilidade de séries temporais financeiras.

A volatilidade constitui importante informação no mercado financeiro. A volatilidade esta associada a medidas de risco sendo empregada na precificação de ativos financeiros. Ademais, grande número de séries econômico-financeiras possuem heteroscedasticidade condicional – efeito ARCH, e a estimação da volatilidade condicional proporciona medidas de intervalos de confiança condicionais mais precisos do que intervalos de confiança fixos. Ainda, medidas de volatilidade são usadas para gestão de risco, como o valor em risco (*Value at Risk* – VaR), a perda máxima esperada para um ativo sobre um dado período de tempo.

Redes neurais são modelos não lineares flexíveis que se aproximam de métodos de estimação não paramétricos, os quais extraem a forma funcional dos dados enquanto os métodos paramétricos impõem a forma funcional da relação a ser analisada. A volatilidade possui muitas características de não linearidade sendo natural a busca de modelos não lineares para estimá-la, como os modelos de redes neurais. Mais ainda, redes neurais potencialmente permitem estudar a extensão da não linearidade presente na volatilidade por uma abordagem que não impõe restrições de especificação da forma funcional ao modelo.

Os modelos de volatilidade tradicionais da família GARCH apresentam diversas limitações para estimação e previsão da volatilidade em situações de mudanças abruptas e para a previsão da volatilidade em horizontes de tempo estendidos (LAMOUREUX, LASTRAPES, 1990; HAMILTON, SUSMEL, 1994; BILDIRICI, ERSIN, 2009). Como

alternativa aos modelos GARCH surgem modelos de volatilidade com mudança de regime, bem como modelos de volatilidade com longa memória, e diversas abordagens e extensões dos modelos GARCH (HAMILTON, SUSMEL, 1994; BAILLIE, BOLLERSLERV, MIKKELSEN, 1996; SHEPAHRD, 1996; ENGEL, 2004). Os modelos de redes neurais, enquanto modelos semi paramétricos podem ser utilizados para tentar aproximar a forma funcional correta para extração da volatilidade.

Ainda os modelos de redes neurais podem ser associados ao modelo GARCH para ampliar a sua capacidade, incorporando características de não linearidade que não são capturadas pelo modelo GARCH. Com o propósito de incorporar mais alta ordem de não linearidade no modelo GARCH é proposto um modelo híbrido GARCH estendido por redes neurais, uma variação do modelo NN-GARCH (DONALDSON, KAMSTRA, 1997).

Este trabalho busca estudar a volatilidade com modelos de redes neurais. Após esta introdução, o capítulo 2 apresenta os modelos de redes neurais e suas características, como modelos regressivos semi paramétricos e com a aproximação de relações por interpolação. O capítulo 3 trata das propriedades das séries temporais financeiras e apresenta os modelos de volatilidade, com destaque para seus aspectos não lineares e as formas de modelá-la, sobretudo para os modelos da família GARCH e com breve comentário sobre modelos de volatilidade com mudança de regime. Ainda o capítulo 3 segue com estimação da volatilidade via modelos GARCH e via redes neurais para duas séries financeiras: o índice S&P500 e as cotações do petróleo tipo Brent. O capítulo 4 propõe um modelo de volatilidade híbrido GARCH aumentado por redes neurais e o estima para as séries estudadas. O trabalho termina com as considerações finais no capítulo 5.

2 Redes Neurais Artificiais (RNA)

Redes neurais artificiais têm o propósito de fornecer um modelo que realize processamento de informações como o cérebro humano. O cérebro recebe informações e as converte em respostas, por meio de um sistema massivo de computação paralela (altamente conectado) baseado em neurônios, sua célula básica. No cérebro, estímulos do corpo ou do meio ambiente são convertidos em impulsos nervosos pelo sistema receptor passando por uma rede intrincada de neurônios onde os impulsos nervosos são combinados e transformados, produzindo-se uma resposta útil.

Redes neurais artificiais possuem uma estrutura que tenta reproduzir o comportamento funcional das redes neurais biológicas, com unidades computacionais simples que são uma representação matemática de um neurônio. Tal estrutura, por meio de suas conexões de neurônios, como um sistema paralelo e conectado, é capaz de resolver problemas complexos com operações paralelas simultâneas de suas unidades simples.

Redes neurais artificiais baseiam-se em um conjunto de dados para aprender uma relação e reproduzi-la. Estas são aplicadas em tarefas de classificação, com a associação de um padrão de entrada a uma classe entre diversas conhecidas; bem como análise de reconhecimento de padrões, como aproximação de funções e previsão de séries temporais, fazendo um mapeamento de uma relação funcional entre as entradas e saídas da rede.

Este capítulo apresenta os modelos de redes neurais artificiais e discute alguns aspectos deste modelo. São consideradas as aplicações de redes neurais como modelos regressivos semi paramétricos, sua aplicação para previsão de séries temporais e suas características de aproximação por interpolação.

2.1 Modelos de Redes Neurais Artificiais

Como motivado acima, de maneira semelhante ao sistema neural biológico, o modelo de rede neural converte os sinais de entrada do meio externo em uma resposta útil. Redes neurais são aplicadas em geral para classificação e reconhecimento de padrões.

No modelo de redes neurais artificiais queremos aprender a relação entre variáveis e reproduzi-la, mapeando o padrão de comportamento entre entradas e saídas. As entradas, chamadas de insumos, correspondem a um conjunto de variáveis explicativas, $X = \{X_1, X_2, \dots, X_m\}$, de dimensão m . Queremos usar os modelos de rede neural para aproximar a variável meta, correspondente a variável dependente, $Y = \{Y_1, Y_2, \dots, Y_n\}$, de dimensão n . Neste trabalho considera-se apenas uma variável meta, $n=1$, por simplicidade.

Uma rede neural apresenta uma estrutura de sinais de entrada, representada pelos insumos, que são propagados pela rede neural para aproximar uma variável meta, a partir do processamento de dados sobre um conjunto de unidades computacionais conectadas numa estrutura paralela.

Redes neurais artificiais são modelos baseados em um conjunto de unidades computacionais simples, neurônios artificiais, que calculam funções matemáticas. Os modelos de redes neurais artificiais são organizados em camadas, onde cada camada contém um número de neurônios que estão conectados continuamente e de forma paralela. A forma mais comum de uma rede neural é aquela com Perceptrons de múltiplas camadas, (*Multilayer Perceptrons – MLP*), que na maioria das aplicações possui uma camada de insumos, uma camada de output e uma camada escondida entre elas. A arquitetura de um Perceptron de múltiplas camadas com três camadas, com vetor de insumos X com m variáveis explicativas na camada de insumos, q neurônios na camada escondida, e uma variável meta na camada de saída, Y , com rede neural denotada por $NN(m,q,1)$ é apresentada na Figura1.

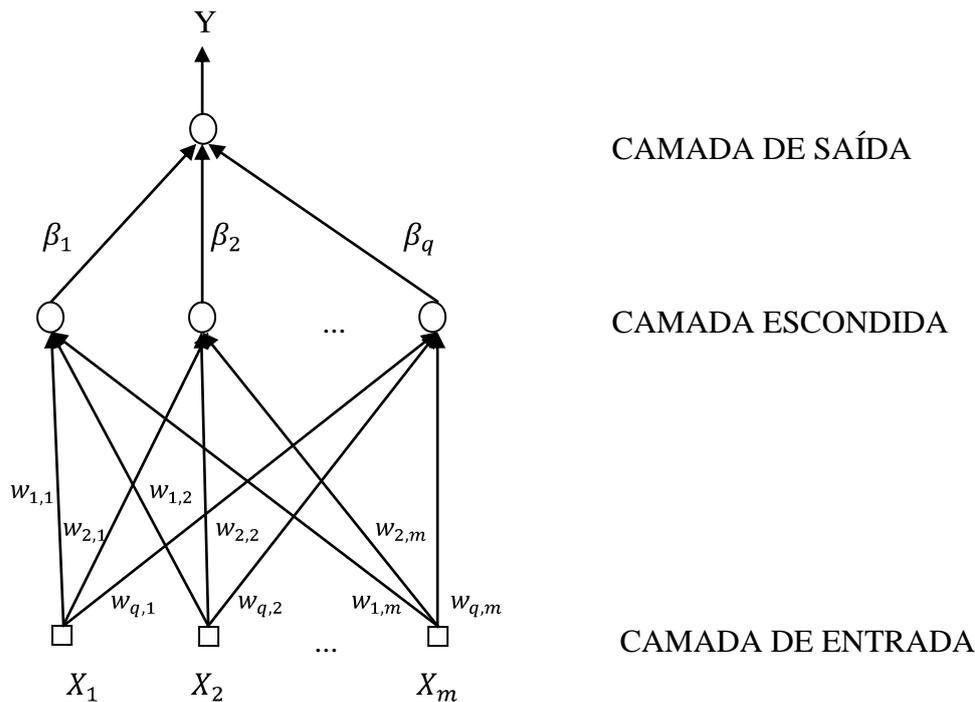


Figura 1 - Modelo de Perceptron de Múltiplas Camadas

Nota: - □ denota unidade de entrada do insumo;
 - ○ representa um neurônio, uma unidade computacional.

O neurônio é a unidade computacional básica de uma rede neural, composto por um conjunto de *pesos sinápticos*, *função de ativação* e *bias*. Um neurônio artificial na camada escondida recebe um conjunto de sinais de insumos, X , que estão conectados a entrada da sinapse do neurônio. Cada sinal é ponderado pelo correspondente peso sináptico, isto é, o sinal X_j é multiplicado pelo peso sináptico $w_{k,j}$, onde o índice k,j se refere a ligação do sinal j para o neurônio k . O neurônio artificial processa a soma dos insumos ponderados pelos pesos sinápticos adicionando um termo de *bias* e estes são avaliados pela *função de ativação*, $G(\cdot)$, resultando no sinal de saída do neurônio k , y_k . A representação matemática de um neurônio (Equação 2.1) é dada por:

$$y_k = G\left(\sum_{j=1}^m w_{k,j} * X_j + b_k\right), \quad (2.1)$$

onde, X_j é insumo j do vetor de insumos X , $w_{k,j}$, é o peso sináptico para cada sinal de entrada que se conecta ao neurônio k , b_k é o termo de bias, $G(\cdot)$, é a função de ativação do neurônio k , e y_k é o sinal de saída do neurônio k .

Os pesos sinápticos amplificam ou inibem um sinal, quanto maior um peso associado a um sinal maior sua carga num neurônio. O termo de *bias* tem o efeito de aumentar (bias positivo) ou diminuir (bias negativo) a entrada líquida da função de ativação. A função de ativação ou restringe o sinal de saída do neurônio na camada escondida, transformando o conjunto de sinais anteriores para serem novamente processados pelos neurônios que seguem na camada seguinte, ou proporciona o valor de resposta na camada final ao conjunto de sinais de entrada.

Similarmente, os sinais dos neurônios da camada escondida são processados na camada de saída e produzem uma resposta útil. Assim, o padrão de entrada, representados pelos insumos, é propagado na camada escondida, onde a soma dos insumos ponderados pelos pesos sinápticos somado ao termo de bias é avaliado na função de ativação, $G(\sum_{j=1}^m w_{k,j} * X_j + b_k)$, para cada neurônio k na camada escondida. Por sua vez, os sinais dos neurônios da camada escondida, y_k , são processados na camada de saída, ponderados pelos pesos, β_j . Ainda adiciona-se o termo de bias, β_0 , chegando-se ao argumento da função de ativação do neurônio da camada de saída, $F(\cdot)$. A saída da rede neural (Equação 2.2) é assim dada por

$$Y = F(\beta_0 + \sum_{j=1}^q \beta_j * G(\sum_{j=1}^m w_{k,j} * X_j + b_k)) \cdot \quad (2.2)$$

O tipo da função de ativação depende se o neurônio está na camada escondida ou na camada de saída, e também da aplicação do modelo de redes neurais. Os neurônios da camada escondida apresentam em geral função de ativação sigmóide, proporcionado a característica de não linearidade ao modelo de rede neural, sendo em muitos casos a função logística. O neurônio da camada de saída normalmente é especificado com função de ativação identidade, tal que o output da rede neural pode assumir um valor real em problemas de reconhecimento de padrão. Por sua vez, em problemas de classificação, com variável meta binária, assumindo valor zero ou a unidade, uma candidata apropriada a função de ativação é a curva logística.

Redes neurais tornam-se capazes de reproduzir uma relação após estimação de seus parâmetros, pesos sinápticos e níveis de bias, num processo de aprendizagem. Este processo baseia-se na minimização de uma função custo (Equação 2.3), representada pela soma do quadrado da diferença entre a resposta desejada, Y_d , e a saída da rede neural, Y , num processo iterativo de alteração dos pesos a partir de um valor inicial, w_0 , aplicando um algoritmo de aprendizagem, normalmente retro-propagação:

$$\text{Min}_{\{w\}} L(w) = \sum_{t=1}^T (Y_{d,t} - Y_t)^2. \quad (2.3)$$

A aprendizagem que determina os pesos estimados da rede neural ocorre de acordo com um algoritmo que estabelece a regra para mudança dos pesos. Numa abordagem de iterações dos pesos numa dimensão temporal, a estimação dos parâmetros da rede neural é feita a partir de um processo de aprendizagem, no sentido que os pesos são alterados de acordo com o conhecimento passado, o valor dos pesos passados, e a informação corrente, o valor das observações correntes. O ajuste dos pesos sinápticos pelo algoritmo de retro-propagação altera os pesos seguindo uma técnica de otimização pelo método do gradiente descendente, no sentido contrário do gradiente da função perda (HAYKIN, 1999). De acordo com o algoritmo de retro-propagação com termo de momento, o ajuste dos pesos sinápticos durante o processo de aprendizagem ocorre de acordo com a seguinte regra¹:

$$w_{k,j(n+1)} = w_{k,j(n)} + \eta * (-1) * \frac{\partial L(w(n))}{\partial w_{k,j(n)}} + \alpha * \Delta w_{k,j(n)}, \quad (2.4)$$

onde o índice n representa a n -ésima iteração, η é a constante de aprendizagem, que controla a taxa de alteração dos pesos na direção oposta da aproximação do gradiente descendente, $\frac{\partial L(w(n))}{\partial w_{k,j(n)}}$ é uma aproximação do gradiente local da função perda, e $\alpha * \Delta w_{k,j(n)}$ é o termo de momento com constante de momento α .

O algoritmo de retro-propagação pode ter problemas para encontrar o mínimo da superfície de erro no espaço dos parâmetros e esta busca pode ser muito lenta. O termo de momento proporciona uma inércia ao algoritmo de retro-propagação aumentando a velocidade de convergência, no sentido de que duas alterações dos pesos na mesma direção são seguidas de uma alteração adicional que aumenta sua variação nesta rotina

¹ Para uma discussão sobre o princípio de operação do algoritmo de retro-propagação ver Apêndice A.

de busca. Ademais, o termo de momento tenta incorporar informação de segunda ordem da superfície de erro.

Os algoritmos de aprendizado mais simples, como o algoritmo de retro-propagação, podem apresentar dificuldades em determinar os pesos que minimizam a superfície de erro dada a complexidade desta e a presença potencial de inúmeros mínimos locais. Algoritmos de aprendizado mais robustos acabam aplicando métodos numéricos que levam em conta informações de segunda ordem na superfície de erro, como métodos Newton e Quase - Newton. Um algoritmo de aprendizado que possui considerável eficiência relativa e tem encontrado larga aplicação é o algoritmo Levenberg-Marquardt, o qual combina características de métodos Gauss-Newton e retro-propagação (HAGAN, MENHAJ, 1994).

2.2 Redes Neurais para Previsão de Séries de Tempo

Os modelos de redes neurais são amplamente aplicados em análises de séries temporais financeiras, as quais possuem características de não linearidade (FRANSES, VAN DIJK, 2000). Redes neurais com múltiplas camadas têm bom desempenho em aproximar funções complexas, sendo aproximadores universais. Redes neurais são modelos flexíveis relacionados a métodos não paramétricos, descrevendo a relação entre a variável dependente e variáveis explicativas com base na forma funcional revelada pelas observações. Além disso, não exigem certas propriedades da série tais como estacionaridade.

Redes neurais possuem a propriedade de aproximadores universais (HORNIK, STINCHCOMBE, WHITE, 1989). Uma rede neural com múltiplas camadas pode aproximar uma grande classe de funções, para um número suficiente de neurônios na camada escondida. Redes neurais proporcionam o mapeamento de uma entrada de insumos, X , sobre a saída da rede. Seja um perceptron de múltiplas camadas, com vetor de insumos de dimensão m , X , um vetor de pesos sinápticos e termos de bias representados por W , então a saída da rede neural, $F(X, W)$, pode aproximar uma função, $f(X)$, para um nível de precisão, tal que

$$|F(X, W) - f(X)| < \varepsilon, \varepsilon > 0. \quad (2.5)$$

Para a análise de séries temporais incorporamos dependência no tempo² ao modelo neuronal utilizando como insumos a variável meta com defasagens, Y_{t-p} , $p = 1, \dots, m$, resultando no seguinte modelo não linear autoregressivo (Equação 2.6):

$$Y_t = \beta_0 + \sum_{j=1}^q \beta_j * G\left(\sum_{j=1}^m w_{k,j} * Y_{t-m} + b_k\right). \quad (2.6)$$

No modelo de rede neural para séries temporais o termo de bias assume o papel do intercepto no modelo de regressão, como a resposta do modelo quando os insumos assumem valor nulo.

Redes neurais podem ser interpretadas como modelos autoregressivos não lineares e com mudança de regime (FRANSES, VAN DIJK, 2000). A não linearidade nos modelos de redes neurais é proporcionada pela função de ativação não linear na camada escondida, como a função logística, que acomoda uma resposta distinta para diferentes valores dos regressores³. Ademais, a especificação do modelo neural permite mudanças de regime que dependem da combinação linear dos insumos, $\sum_{j=1}^m w_{k,j} * Y_{t-m} + b_k$, para valores suficientemente pequenos a função logística assume valor que tende a zero e a média condicional seria β_0 , em outro regime com valores maiores para a combinação linear dos insumos e a resposta da função logística tendendo a unidade a média condicional seria $\beta_0 + \sum_{j=1}^q \beta_j$. Para valores gerais dos insumos temos uma transição entre a média condicional dos dois regimes.

O modelo de regressão descreve uma relação estocástica entre a variável dependente, Y , e insumos, X , considerando um termo condicionalmente determinístico, $f(X)$, e um componente aleatório, ε , $Y = f(X) + \varepsilon$. Na análise de regressão é especificada uma forma funcional para relação entre a variável dependente e o regressor, que normalmente corresponde a média da variável dependente condicionada aos regressores, $E[Y|X] = f(X)$. O modelo de redes neurais pode ser considerado um modelo onde a forma funcional estimada da média da variável de saída condicionada

² Outras especificações que incorporam dinâmica ao modelo de redes neurais são a partir de redes neurais recorrentes, com realimentações de output ou saídas das unidades escondidas como insumos do modelo (KUAN & WITHE, 1994).

³ Em um modelo linear, como o modelo de regressão linear $Y=B*X + \varepsilon$, todos os valores dos regressores tem o mesmo efeito marginal sobre a variável dependente, $\frac{\partial Y}{\partial X} = B$, independente da magnitude do regressor. No modelo de redes neurais, $Y = \beta_0 + \sum B * G(X, w)$, o efeito marginal de uma variável depende de sua magnitude, $\frac{\partial Y}{\partial X} = \sum B * G'(X, w)$.

aos insumos depende dos dados e é aproximada pela rede neural, isto é, $E[Y|X] = F(X, w)$.

Métodos paramétricos especificam uma forma funcional para descrever uma relação: linear, quadrática, etc. Por outro lado, métodos não paramétricos obtêm a forma da relação a partir dos dados. Os métodos paramétricos proporcionam uma estimação precisa e com resultados ótimos que dependem da forma especificada da relação estar corretamente especificada, e podem conduzir a bons resultados mesmo na presença de poucas observações, enquanto métodos não paramétricos precisam de maior quantidade de dados para alcançar bom ajuste (GEMAN, BIENENSTOCK, DOURSAT, 1992). Os modelos não paramétricos podem aproximar a relação arbitrária subjacente à série temporal. Redes neurais são modelos não paramétricos, pois nas aplicações o número de unidades escondidas é fixado, e a aproximação livre de funções requer um número não limitado de neurônios na camada escondida (DORFFNER, 1996).

Redes neurais artificiais são métodos mais flexíveis que os modelos paramétricos de séries temporais tradicionais. Redes neurais são modelos livres para descrever relações complexas subjacentes da série temporal e não exigem características específicas da mesma, como estacionaridade (DORFFNER, 1996).

2.3 Redes Neurais & Interpolação

Redes neurais são usadas com êxito em problemas de classificação e aproximação devido a sua propriedade de generalização, proporcionando saídas muito próximas da variável meta desejada para valores dos insumos não presentes na amostra de treinamento (HAYKIN, 1999). Os modelos de redes neurais se destacam pela sua capacidade de boa generalização mesmo para relações complexas não lineares, dada a sua natureza essencialmente não linear.

Redes neurais são formas não lineares flexíveis que realizam interpolação com base em funções, tal que a saída da rede neural pode ser apresentada como uma interpolação das funções de ativação dos neurônios da camada escondida (LI, LEE, 2003). Os modelos de rede neurais realizam um mapeamento contínuo não linear de entrada saída com base nos neurônios da camada escondida, com funções de ativação contínuas e não lineares. O aprendizado de redes neurais via este mapeamento, ajusta uma curva sobre a amostra de treinamento que pode aproximar fronteiras de decisão em problemas de classificação ou formas não lineares flexíveis em análise de

reconhecimento de padrão. A generalização usa o mapeamento da relação entrada e saída do aprendizado para realizar interpolações (Figura 2).

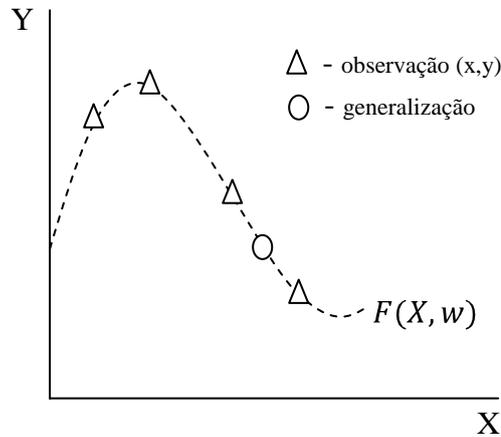


Figura 2 - Aprendizado e Generalização por Redes Neurais

Seja um perceptron de múltiplas camadas com insumo e meta escalares, uma única camada escondida, função de ativação das unidades escondidas logística e função de ativação da camada de saída linear. Neste caso, o modelo neural realiza uma transformação do insumo para a meta como uma função de uma variável, $Y = \beta_0 + \sum_{j=1}^q \beta_j * G(w_{k,1} * X + b_k)$. A grande flexibilidade das redes neurais para aproximar relações complexas está associada a sua função de ativação, que na maioria dos casos é uma função sigmóide. No caso da função logística (Equação 2.7) temos

$$G(X, w) = \frac{1}{1 + \exp(-(w_{k,1} * X + b_k))} \quad (2.7)$$

A função logística tem a forma de “s” (Figura 3) e assume valores entre zero e a unidade para os sinais de entrada. A função logística se aproxima de 1 quando o seu argumento tende ao infinito e se aproxima de zero quando seu argumento tende a menos infinito. A função logística é diferenciável e combina comportamentos linear e não linear.

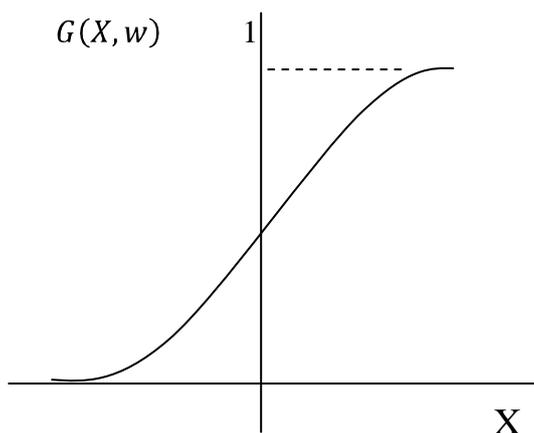


Figura 3 - Gráfico da Função Logística

A rede neural com sua estrutura paralela conectada por pesos sinápticos combina elementos locais, representados pelos neurônios na camada escondida, para reproduzir características globais, com base na interpolação dos neurônios na camada escondida que fazem aproximações de relações localizadas. A interpolação da rede neural, com aproximação de aspectos locais da relação entre variável meta e insumo para depois combiná-los e obter a aproximação global da função meta, permite grande flexibilidade para as redes neurais. Uma aproximação linear, baseada numa região semi-linear da função de ativação de um neurônio da camada escondida, pode ter sua inclinação alterada pela multiplicação do peso sináptico na camada de saída. Outro elemento que ajuda a aproximar aspectos locais da função é o termo de bias, que, desprezando o valor do peso sináptico sobre o sinal de entrada, é um indicador do ponto de interpolação. Para valores distantes de zero alteram a saída de um neurônio para as regiões curvilíneas da função de ativação; já valores nulos não alteram a correspondência de saída de um neurônio na função logística. Por sua vez, o peso sináptico, desprezando o efeito do bias, altera a inclinação da função de ativação, quanto maior seu valor mais rápido os insumos são correspondidos com as regiões saturadas da função de ativação, e mudanças de sinal promovem uma reversão do sentido de orientação dos insumos ao sinal de saída.

O modelo de redes neurais tem importante capacidade de aproximar funções, reconhecendo relações entre as variáveis, mesmo na presença de distorções por aleatoriedade da série que é observada com ruído. Um exemplo da capacidade de mapear relações entrada-saída é apresentada no Gráfico 1, com aproximação da

verdadeira função geradora dos dados, como função $y = \text{sen}(x)$, a partir de sua amostra com perturbações estocásticas. A aproximação da função senóide realizada pela rede neural, com uma arquitetura com três neurônios na camada escondida, é praticamente igual a função geradora a despeito de ter sido treinada com a forma ruidosa da função senóide.

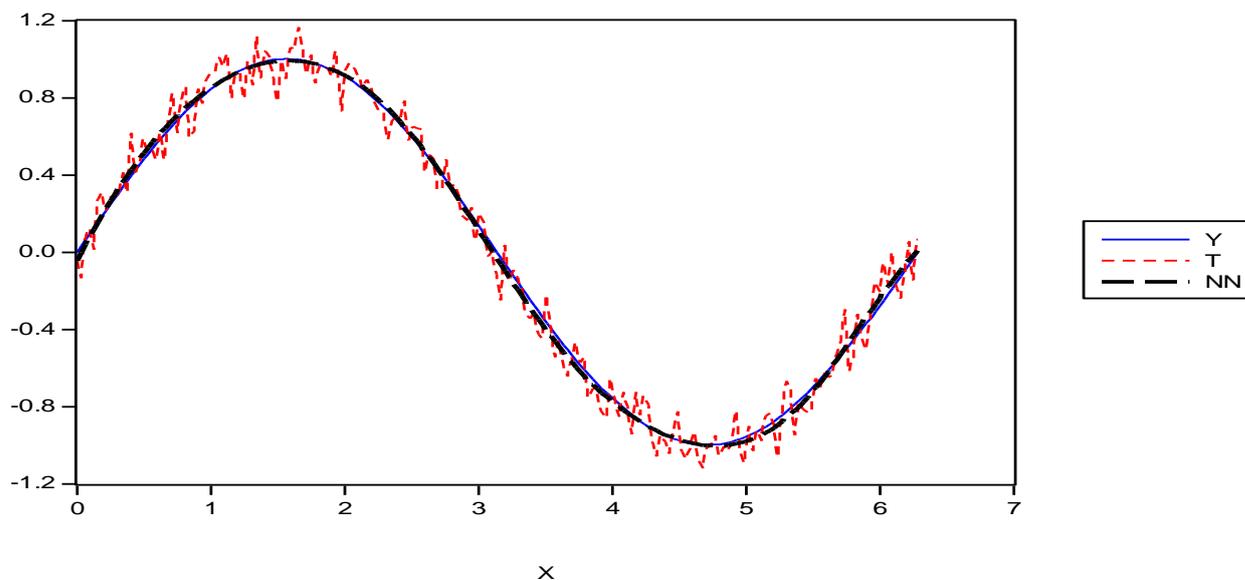
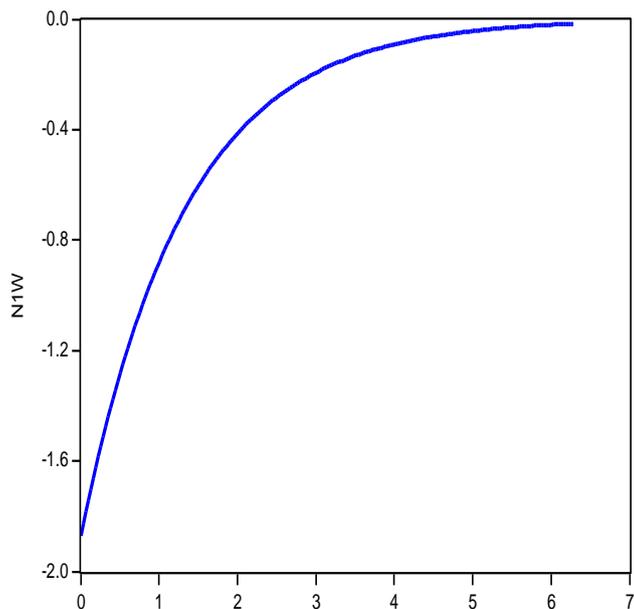


Gráfico 1 – Aproximação da Função $y = \text{sen}(x)$ com ruído, pela rede neural NN(1,3,1).

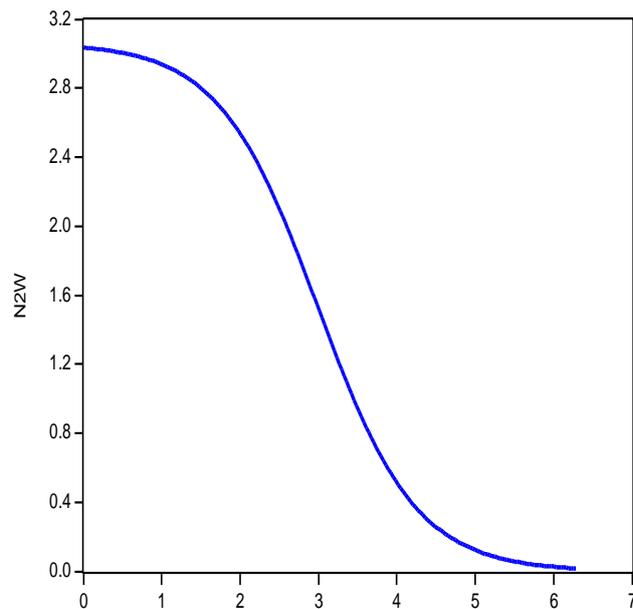
Nota: $y = \text{sen}(x)$, função seno no intervalo $[0;2\pi]$, com $\pi \cong 3.14$; T, função $\text{sen}(x)$ com ruído normal, $T = \text{sen}(x) + \varepsilon$, com $\varepsilon \sim N(0,0.10)$; NN aproximação da função $\text{sen}(x)$ para rede neural, NN(1,3,1), treinada com as observações da função senóide com ruído.

A interpolação realizada pela rede neural na aproximação da função senóide com perturbações é apresentada no Gráfico 2. O primeiro neurônio aproxima a inclinação crescente para o domínio $[1,2]$. O segundo neurônio ajusta o segmento semi linear para o domínio $[2,5]$. O terceiro neurônio assimila a inclinação crescente para a parte dos valores dos insumos no domínio $[5,6]$. A aproximação da função seno é resultante da combinação linear dos neurônios da camada escondida. Os neurônios da camada escondida aproximam relações localizadas e a saída da rede neural combina as aproximações locais para reproduzir a função geradora.



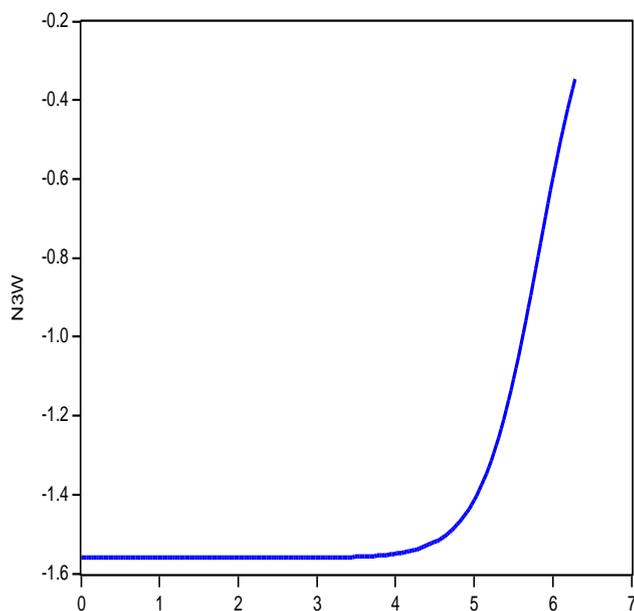
x

(a)



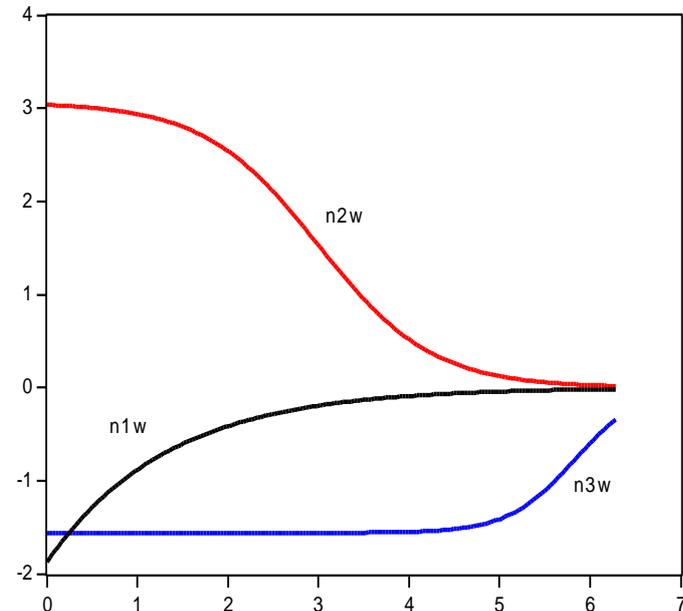
x

(b)



x

(c)



x

(d)

Gráfico 2 - Neurônios da rede neural NN(1,3,1) usada para aproximar a função $y = \sin(x)$.

Nota: (a) Neurônio 1 multiplicado pelo peso sináptico na camada de saída, $w_1 \cdot n_1$; (b) Neurônio 2 multiplicado pelo peso sináptico na camada de saída, $w_2 \cdot n_2$; (c) Neurônio 3 multiplicado pelo peso sináptico na camada de saída, $w_3 \cdot n_3$; (d) Composição das interpolações locais para reproduzir a aproximação da função $\sin(x)$.

Nem sempre a forma que é feita a interpolação pela rede neural assume uma interpretação trivial de combinação de funções. Para um dado mapeamento de entrada e saída que aproximam uma função existem arbitrárias combinações da função base da interpolação para realizá-la, algumas com interpretação gráfica longe de trivial.

Os modelos de redes neurais podem realizar aproximação de funções e classificação em problemas complexos com bom desempenho. No entanto para modelos de redes neurais inadequados a generalização pode ser muito pobre devido ao “*over-fitting*”, excesso de ajuste, onde a rede neural reproduz satisfatoriamente a amostra de treinamento, com medida de coeficiente de determinação próximos da unidade, mas alcança desempenho sofrível na amostra de teste (Figura 4).

O “*over-fitting*” nos modelos de redes neurais ocorre com modelos com grande quantidade de neurônios na camada escondida, causando super parametrização, e excesso de treinamento. Redes neurais com muitos neurônios na camada escondida podem realizar interpolações sobre regiões muito específicas da relação entre variável meta e insumo, absorvendo características locais, como ruídos no mapeamento de entrada e saída, que não correspondem à função geradora dos dados. Mais ainda, muitos neurônios na camada escondida tornam estreito um ponto base de interpolação que tem maior flexibilidade para passar pela série de observações sem uma aproximação da função geradora subjacente aos dados. Por sua vez, com excesso de treinamento os pesos acabam procurando ajustes muito específicos na amostra de treinamento que limitam a generalização.

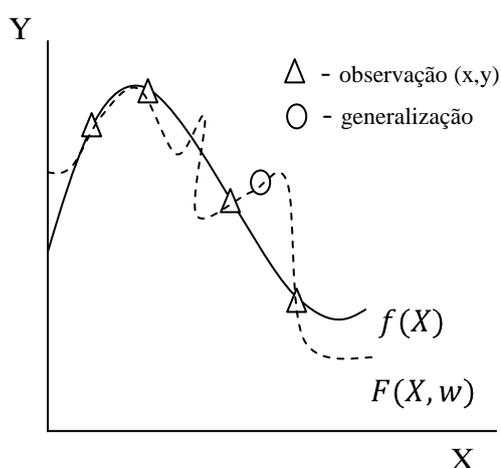


Figura 4 – Rede Neural com Generalização Limitada (*Over-fitting*)

Para um Perceptron de múltiplas camadas de arquitetura arbitrária a solução do problema de aprendizado minimizando a diferença quadrática entre a relação entre a saída da rede neural e a meta possui soluções também arbitrárias, muitas das quais resultam em pobre generalização. Uma rede neural mais parcimoniosa leva a um conjunto de soluções possíveis menores que incluem com mais frequência a verdadeira função geradora dos dados (BRAGA, CARVALHO, LUDEMIR, 2007).

3 Fatos Estilizados das Séries Temporais Financeiras e Modelos de Volatilidade

Na análise de séries temporais financeiras usamos modelos estatísticos para descrever as cotações dos ativos financeiros. Séries financeiras apresentam em especial aspectos de não linearidade. Suas trajetórias são complexas dada a natureza especulativa das mesmas. Além disso, no atual contexto de mercado financeiro integrado, tem-se um comportamento sujeito a choques externos, incorporando mais aleatoriedade às séries.

Os modelos de volatilidade descrevem a variabilidade condicional de uma série de tempo. Muitos ativos financeiros apresentam períodos estáveis com trajetória bem comportada, movendo-se para períodos onde a previsibilidade é limitada com grande variabilidade das cotações.

Para discussão dos modelos de volatilidade este capítulo começa com uma apresentação dos fatos estilizados dos retornos financeiros, enfocando seus aspectos de não linearidade, e segue com a descrição de modelos de volatilidade não lineares da família GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*). Colocar em itálico termos em outras línguas. Ainda este capítulo apresenta a comparação dos resultados de estimação da volatilidade por redes neurais e modelos GARCH para as séries de retornos S&P500 e Brent.

3.1 Não Linearidades dos Retornos Financeiros

As séries financeiras apresentam características que outras séries temporais não possuem, conhecidas como fatos estilizados dos retornos financeiros. Assim, buscamos modelos que melhor descrevam os aspectos peculiares das séries financeiras. Vários fatos estilizados são comuns aos diferentes tipos de ativos especulativos, títulos, ações, índices de mercado, commodities, etc.; e às diferentes frequências de dados, geralmente disponíveis com cotações diárias, e mais recentemente em cotações intra-diárias⁴.

⁴ Os fatos estilizados dos retornos financeiros se mantêm para séries em distintas frequências, embora é verificado que séries em alta frequência possuem características particulares. Este estudo trata apenas das séries em frequência diária.

Os fatos estilizados das séries temporais financeiras envolvem características de imprevisibilidade, com ausência ou pouca correlação entre os retornos, porém os quadrados dos mesmos são correlacionados; a distribuição dos retornos apresenta caudas pesadas ou excesso de curtose, isto é, maior probabilidade de ocorrência de valores extremos em relação à distribuição normal; assimetria, da relação entre a volatilidade e retorno defasado; e agrupamentos de volatilidade no tempo com períodos de alta volatilidade alternando-se com períodos de baixa volatilidade (TSAY, 2005; MORETTIN, TOLOI, 2004; RYDBERG, 2000).

As propriedades de não linearidade nas séries financeiras se manifestam na sua volatilidade, o qual é o momento passível de ser modelado devido à presença de pouca correlação nos retornos originais. As não linearidades presentes na volatilidade são

- (a) *Não linearidade*, a volatilidade responde de maneira diferente a retornos grandes e pequenos, tendendo a ser menor frente a variações das cotações pequenas e se torna mais pronunciada com alterações nas cotações altas;
- (b) *Efeito alavancagem*, a volatilidade tem comportamento distinto frente a retornos positivos e negativos de mesma magnitude. Variações das cotações para redução de seu valor são seguidas de maior volatilidade do que alterações de cotações no sentido de alta com mesmo valor em módulo;
- (c) *Agrupamentos de volatilidade*, os mercados financeiros em alguns momentos são calmos e em outros instáveis. Nos momentos de estabilidade os retornos possuem baixa variação, com menor variância. Nos momentos de alta volatilidade os retornos possuem maior variação, grande variância. Desta forma podemos modelar a série de retornos por uma mistura de distribuições, e volatilidade que se altera de um regime para o outro.

Embora a não linearidade da volatilidade na forma de resposta não linear aos retornos e efeito “*alavancagem*” sejam aspectos empíricos comuns a muitas séries financeiras os agrupamentos de volatilidade assumem a forma de mudança de regime em casos mais particulares. Verifica-se mudança de regime com grande frequência para estudos de séries com horizontes de análise estendidos (LAMOUREUX, LASTRAPES, 1990).

3.2 Modelos de Volatilidade

A volatilidade descreve a variação dos preços/retornos. A volatilidade condicional de um ativo é uma variável não observável, o que temos são observações das cotações de fechamento dos ativos que usamos para extrair a volatilidade. Modelamos a heterocedasticidade condicional dos retornos financeiros que se altera ao longo das observações dos retornos.

Sejam X_t os retornos não correlacionados ou resíduos de um modelo de série temporal para os retornos, $\{X_t\}_{t=0}^{+\infty}$. A volatilidade é modelada pela variância condicional dos retornos dos ativos, conforme a seguir:

$$X_t = \sqrt{h_t} * v_t$$

$$h_t = \alpha_0 + H(X_{t-1}, \dots, X_{t-r}, h_{t-1}, \dots, h_{t-s}; \theta, S_t), \quad r \text{ e } s \in \mathbb{N} \text{ e } S_t \in [1, \dots, N],$$

onde v_t é uma variável aleatória não correlacionada com h_t com média zero, $E[v_t] = 0$, e variância igual a unidade, $VAR[v_t] = 1$, comumente com distribuição normal, $N(0,1)$, ou t de Student, t_n , com n graus de liberdade; θ , vetor de parâmetros; S_t , os estados passíveis de serem assumidos pela série financeira, $S_t \in [1, \dots, N]$.

Temos diversos modelos para a volatilidade que geralmente expressam a volatilidade contemporânea como função dos retornos e/ou volatilidade de períodos anteriores, $H(X_{t-1}, \dots, X_{t-r}, h_{t-1}, \dots, h_{t-s}; \theta, S_t)$. A equação para a variância condicional é em geral uma função não linear, $f(\cdot)$, e esta ainda pode variar de um estado para o outro (no caso de múltiplos regimes) caracterizando um modelo não linear por partes, constituído a partir de equações que se alteram nas sub-amostras da série pertencentes a regimes distintos. As especificações possíveis para a variância condicional são

$$(a) H(X_{t-1}, \dots, X_{t-r}, h_{t-1}, \dots, h_{t-s}; \theta, S_t) = f(X_{t-1}, \dots, X_{t-r}, h_{t-1}, \dots, h_{t-s}; \theta, S_t = 1);$$

$$(b) H(X_{t-1}, \dots, X_{t-r}, h_{t-1}, \dots, h_{t-s}; \theta, S_t) = \begin{cases} f(X_{t-1}, \dots, X_{t-r}, h_{t-1}, \dots, h_{t-s}; \theta, S_t = 1) \\ f(X_{t-1}, \dots, X_{t-r}, h_{t-1}, \dots, h_{t-s}; \theta, S_t = 2) \\ \vdots \\ f(X_{t-1}, \dots, X_{t-r}, h_{t-1}, \dots, h_{t-s}; \theta, S_t = N) \end{cases},$$

para $S_t \in [1, \dots, N]$.

Os modelos da família GARCH tratam a volatilidade como um modelo não linear dos retornos e volatilidade passada, $H(X_{t-1}, \dots, X_{t-r}, h_{t-1}, \dots, h_{t-s}; \theta, S_t)$. As variações dos modelos GARCH tentam descrever de forma mais precisa as características dos retornos financeiros.

O modelo ARCH (*Autoregressive Conditional Heteroskedasticity*) ou modelo de Heterocedasticidade Condicional Autoregressiva (ENGLE, 1982) descreve a volatilidade como função somente dos quadrados dos retornos passados. O modelo ARCH de ordem r , ARCH(r), é dado por:

$$X_t = \sqrt{h_t} * v_t \quad (3.1)$$

$$h_t = \alpha_0 + \sum_{i=1}^r \alpha_i X_{t-i}^2, \alpha_0 \geq 0, \alpha_i \geq 0, i = 1, \dots, r. \quad (3.2)$$

As condições $\alpha_0 \geq 0, \alpha_i \geq 0, i = 1, \dots, r$, asseguram a não negatividade da variância. O modelo ARCH delinea os *clusters* de volatilidade por um comportamento autoregressivo para os quadrados dos retornos, maiores retornos em módulo têm mais chances de serem seguidos por maiores retornos.

O modelo GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*), ou modelo Generalizado de Heterocedasticidade Condicional Autorregressiva, proposto por Engle e Bollerslev (1986), é uma generalização do modelo ARCH. O modelo GARCH apresenta a volatilidade contemporânea como dependendo dos quadrados dos retornos passados e também de volatilidades condicionais passadas. O modelo GARCH(r,s), é especificado como:

$$X_t = \sqrt{h_t} * v_t \quad (3.3)$$

$$h_t = \alpha_0 + \sum_{i=1}^r \alpha_i X_{t-i}^2 + \sum_{j=1}^s \beta_j h_{t-j}, \quad (3.4)$$

onde $\alpha_0 \geq 0, \alpha_i \geq 0, \beta_j \geq 0, i = 1, \dots, r$ e $j = 1, \dots, s$.

A não negatividade dos parâmetros garante o valor não negativo da variância. Se a ordem do termo GARCH é zero, $s = 0$, o mesmo se reduz ao modelo ARCH(r). O modelo GARCH descreve a volatilidade como associada aos retornos e volatilidade prévia. Um modelo GARCH permite uma representação como um processo autoregressivo e de médias móveis para os quadrados dos retornos.

Os modelos ARCH e GARCH são modelos simétricos, enquanto em certos casos a volatilidade pode responder com magnitude distinta frente a retornos positivos e negativos de igual valor em modulo, o que é conhecido como efeito *alavancagem*. Os modelos ARCH e GARCH não incorporam o efeito dos sinais dos retornos sobre a volatilidade, pois na equação da variância só aparecem seus quadrados. A assimetria na trajetória da volatilidade é considerada pelos modelos TGARCH e EGARCH.

O modelo TGARCH (*Threshold GARCH*), modelo Generalizado de Heterocedasticidade Condicional Autoregressiva com Limiar, incorpora a assimetria da volatilidade ante os choques negativos através de uma variável *dummy*. Seja a série de retornos, X_t , a variância condicional depende do quadrado dos erros de previsão e volatilidade passados adicionada ao efeito limiar. O modelo TGARCH(r,s) (Equações 3.5 - 3.7) é expresso por

$$X_t = \sqrt{h_t} * v_t \quad (3.5)$$

$$h_t = \alpha_0 + \sum_{i=1}^r \alpha_i X_{t-i}^2 + \sum_{j=1}^s \beta_j h_{t-j} + \sum_{i=1}^r \gamma D_{t-i} X_{t-i}^2 \quad (3.6)$$

$$\begin{cases} D_{t-i} = 1, \text{ se } X_{t-i} < 0 \\ D_{t-i} = 0, \text{ caso contrário,} \end{cases} \quad (3.7)$$

onde $\alpha_0 \geq 0, \alpha_i \geq 0, \beta_j \geq 0 \ i = 1, \dots, r \ e \ j = 1, \dots, s$.

Um coeficiente da variável *dummy* significativo e positivo denota o efeito *alavancagem*, com volatilidade mais pronunciada frente a retornos negativos em relação a positivos.

O modelo EGARCH (*Exponencial GARCH*), modelo Generalizado de Heterocedasticidade Condicional Autoregressiva Exponencial considera a assimetria sobre a volatilidade pelo impacto dos retornos padronizados. Além disso, devido à especificação logarítmica da volatilidade possui menos restrições sobre os parâmetros tornando desnecessárias restrições de não negatividade. Seja a série de retornos, X_t , então o modelo EGARCH é especificado como

$$X_t = \sqrt{h_t} * v_t \quad (3.8)$$

$$\ln(h_t) = \alpha_0 + \sum_{i=1}^r \alpha_i \frac{X_{t-i}}{\sqrt{h_{t-i}}} + \sum_{i=1}^r \lambda_i \left| \frac{X_{t-i}}{\sqrt{h_{t-i}}} \right| + \sum_{j=1}^s \beta_j \ln(h_{t-j}). \quad (3.9)$$

No modelo E-GARCH se um choque é positivo o seu efeito sobre o logaritmo da variância condicional é $\sum_{i=1}^r \alpha_i + \sum_{i=1}^r \lambda_i$. Para apenas um retorno negativo, X_{t-i} , seu

efeito sobre o logaritmo da variância condicional é $-\alpha_i + \lambda_i$; espera-se que o coeficiente α_i seja negativo, de modo que a volatilidade seja maior frente a choques negativos em relação a choques positivos.

O efeito alavancagem explorado pelos modelos assimétricos traz o conceito de choques sobre os retornos. Os agentes econômicos no mercado financeiro compõem sua carteira de investimento de acordo com a expectativa de valoração em relação a um ativo. Os agentes formulam suas expectativas sobre as realizações futuras do mercado usando da maneira mais eficiente a informação do período corrente, estando interessados no retorno esperado condicional à informação anterior $E[y_t|F_{t-1}]$, sua expectativa de retorno, e na variância condicional a informação passada, $VAR[y_t|F_{t-1}]$, como risco esperado (ENGLE, NG, 1993). F_{t-1} denota o conjunto de informação do período anterior, composto das realizações de todas as variáveis relevantes até o momento passado. Como o resultado do mercado é incerto os agentes incorrem em erros sobre suas previsões $X_t = y_t - E[y_t|F_{t-1}]$. Estes erros de previsão constituem em nova informação que será incorporada nos subseqüentes movimentos do mercado. Os choques sobre os retornos podem ser positivos, $X_t > 0$, ou negativos, $X_t < 0$. A volatilidade tem diferente resposta frente a choques de distinto sinal e magnitude.

Os modelos de volatilidade são modelos empíricos para descrever a variabilidades dos retornos dos ativos. Embora não possamos afirmar que os retornos passados e seus erros de previsão sejam as causas da volatilidade, neles estão contidas as informações que resultam na volatilidade. A literatura de finanças empíricas atribui a variabilidade dos retornos e seus agrupamentos à chegada de nova informação. Como a chegada de novas notícias está agrupada no tempo, como resultado tem-se os *clusters* de volatilidade; ademais a volatilidade de certos ativos tem efeitos sobre outros mercados (ENGLE, 2004). Outros argumentos encontrados na literatura para a volatilidade e sua observação em *clusters*, justificam este fenômeno atribuindo a deformação temporal onde o tempo econômico seria distinto de sua medida pelo tempo calendário (BOLLERSLEV, CHOU, KRONER, 1992; ZIEGELMANN, PEREIRA 1997). Frente à informação que se condensa para um dia atípico de transação observamos variações dos retornos em um dia de magnitude igual a que se verifica sob sua variação acumulada em mais de um período, estes últimos sob fluxo de informação para ambientes normais.

A assimetria dos choques dos retornos sobre a volatilidade, incorporando a resposta da volatilidade à nova informação, é medida pela curva de chegada de informação. A curva de chegada de informação (*News Impact Curve - NIC*) é uma medida do efeito dos retornos não previsíveis do período prévio, X_{t-1} , sobre a volatilidade corrente, mantendo constante toda a informação anterior ao retorno passado. Para o modelo GARCH a curva de chegada de informação é simétrica, os modelos T-GARCH e E-GARCH possuem uma expressão assimétrica. Para os modelos GARCH(1,1), e T-GARCH(1,1) as curvas de chegada de informação são, respectivamente,

$$i) NIC_{GARCH}(X_{t-1}|h_{t-1} = \sigma^2) = A + \alpha_1 * X_{t-1}^2; \quad (3.10)$$

$$ii) NIC_{TGARCH}(X_{t-1}|h_{t-1} = \sigma^2) = \begin{cases} A + \alpha_1 * X_{t-1}^2, & \text{se } X_{t-1} > 0, \\ A + (\alpha_1 + \gamma) * X_{t-1}^2, & \text{se } X_{t-1} < 0, \end{cases} \quad (3.11)$$

para $A = \alpha_0 + \beta * \sigma^2$.⁵ O modelo E-GARCH apresenta curva de chegada de informação similar ao T-GARCH, porém na forma exponencial para desfazer a transformação logarítmica pela qual é estimado.

Outro conceito importante na análise dos modelos GARCH é a persistência. A persistência no modelo de volatilidade é uma medida da permanência de impacto de um choque sobre a volatilidade. No modelo GARCH a persistência é dada pela soma dos coeficientes ARCH e GARCH, $\lambda = \sum_{i=1}^r \alpha_i + \sum_{j=1}^s \beta_j$, correspondendo ao fator de decaimento da solução da equação de diferenças (Equação 3.12) para sua previsão,

$$E[h_t|F_{t-1}] = \alpha_0 + \alpha_1 E[X_{t-1}^2|F_{t-1}] + \beta_1 E[h_{t-1}|F_{t-1}]. \quad (3.12)$$

A estimação dos modelos de volatilidade da família GARCH, é baseada na função densidade dos retornos condicionada ao conjunto de informação anterior $Y_t|F_{t-1} \sim IID(\mu_t, h_t; \theta)$, empregando métodos de máxima verossimilhança, quase máxima verossimilhança ou método generalizado dos momentos (SHEPHARD, 1996). A função densidade conjunta condicionada resulta na função de verossimilhança que

⁵ A curva de chegada de informação mede o efeito do retorno prévio não previsto sobre a volatilidade corrente mantendo constante todo o conjunto de informação anterior ao período passado, ou seja, a esperança da volatilidade contemporânea condicional ao conjunto de informação t-2, F_{t-2} . Desta forma, para o modelo GARCH(1,1) a curva de chegada de informação é $E[h_t|F_{t-2}] = \alpha_0 + \alpha_1 E[X_{t-1}^2|F_{t-2}] + \beta_1 E[h_{t-1}|F_{t-2}]$, assumindo estacionaridade em covariância $E[h_t|F_{t-1}] = E[h_{t-1}|F_{t-2}] = \dots = E[h_{t-\tau}|F_{t-\tau-1}] = \sigma^2, \forall \tau$. Assim, a curva de chegada de informação para um processo GARCH(1,1) se expressa como $NIC_{GARCH} = A + \alpha_1 * X_{t-1}^2$, onde $A = \alpha_0 + \beta * \sigma^2$. Aplicando o mesmo princípio chegamos a curva de chegada de informação para o modelo T-GARCH(1,1).

maximizamos em relação aos parâmetros do modelo. Supondo distribuição normal, a função de verossimilhança é dada por

$$f(y_t, y_{t-1}, \dots, y_{t-T}; \theta, F_{t-1}) = \sum_{t=1}^T \ln f(y_t | y_{t-1}; \theta) = -\frac{1}{2} \sum_{t=1}^T \ln h_t - \frac{1}{2} \sum_{t=1}^T \frac{(y_t - \mu_t)^2}{h_t}. \quad (3.13)$$

Outras distribuições podem ser consideradas incorporando as caudas pesadas dos retornos, tais como a distribuição t de Student.

A volatilidade pode ter um padrão de comportamento que se altera ao longo do tempo, em determinados períodos a volatilidade é maior que outros. Estudos de simulação mostram que os modelos GARCH têm persistência muito elevada quando há quebra estrutural na variância incondicional (LAMOREOUX, LASTRAPES, 1990). A volatilidade muda seu comportamento no tempo e se não modelarmos esta alteração de comportamento, como mudança da variância incondicional que se reflete na variância condicional, os parâmetros estimados acabam enviesados com superestimação da persistência da volatilidade.

A alta persistência na volatilidade pode levar a uma menor capacidade de previsão para os modelos GARCH, visto que um choque sobre os retornos tem efeito que se propaga sobre a volatilidade após grande período de tempo. Caso haja uma quebra de regime ou presença de outlier, isto pode levar a previsões limitadas, pois a persistência estimada resulta numa estimação da volatilidade que é geralmente maior que a real.

Os modelos de volatilidade com mudança de regime especificam a variância condicional $H(X_{t-1}, \dots, X_{t-r}, h_{t-1}, \dots, h_{t-s}; \theta, S_t)$ com equações distintas para cada regime descrevendo a mudança de regime entre os estados possíveis seguindo uma cadeia de Markov. Aqui é apresentado o modelo SWARCH de Hamilton e Smell (1994); para uma discussão ampla de modelos de volatilidade com mudança de regime ver Dueker (1997).

No modelo SWARCH temos um processo ARCH com mudança de regime que altera a escala da volatilidade, com transição entre N regimes possíveis, $s_t = \{1, 2, \dots, N\}$ governada por uma cadeia de Markov de primeira ordem. O modelo SWARCH(N, r) (Equações 3.14 – 3.18), sob dois regimes, $N = 2$, com um processo ARCH(r), é formulado como

$$u_t = \sqrt{g_{st}} * \tilde{u}_t, \quad (3.14)$$

$$\tilde{u}_t = h_t * v_t, \quad (3.15)$$

$$h_t^2 = \alpha_0 + \sum_{i=1}^r \tilde{u}_{t-i}^2, \quad (3.16)$$

$$P\{S_{t+1} = j | S_t = i\} = p_{i,j}, \quad (3.17)$$

$$\text{onde } \sum_{j=1}^N p_{i,j} = 1, i = 1, 2, \dots \text{ e } p_{i,j} \geq 0, i, j \in [1, \dots, N]. \quad (3.18)$$

O fator que multiplica o processo ARCH, $\sqrt{g_{st}}$, para o primeiro regime, é $\sqrt{g_1}$; e $\sqrt{g_2}$, para o regime 2. O fator para o primeiro regime é normalizado na unidade, $g_1 = 1$, tal que os demais regimes exibem maior volatilidade, $g_{st} \geq 1$, para $s_t = \{1, 2, \dots, N\}$, com $p_{i,j}$ representando a probabilidade de transição entre os regime i e j .

Na estimação do modelo ARCH com mudança de regime, para uma dada especificação de distribuição para os retornos, normal ou t-Student, a variância depende do regime em t . O procedimento de estimação segue a especificação da distribuição marginal das observações e a função da verossimilhança (HAMILTON, SUSMEL, 1994). O modelo SWARCH é um modelo de volatilidade com mudança de regime parcimonioso, uma especificação GARCH introduziria a dependência sobre todos os estados passados, pelo componente GARCH⁶, e torna a estimação dificultada e somente é possível por métodos de aproximação (DUEKER, 1997).

⁶ Um modelo GARCH após sucessivas recursões resulta num modelo ARCH de ordem infinita.

3.3 Estimação da Volatilidade por Redes Neurais e Modelos GARCH: Um Estudo Empírico

Aqui consideram-se duas séries financeiras: o índice S&P500 e as cotações do petróleo Brent. A aplicação de redes neurais para estimar a volatilidade é comparada com os modelos GARCH em termos de estimativas de curva de chegada de informação, desempenho preditivo e intervalos de confiança. Assim estudamos os aspectos quantitativos e qualitativos acerca das estimativas da volatilidade proporcionada pelos diferentes modelos.

O mercado financeiro tem importância fundamental para a economia, permitindo a capitalização de empresas no mercado acionário, além de ser indicador do desempenho da atividade econômica. O índice S&P500, além de ser uma medida do estado da economia americana, proporciona também uma sinalização sobre eventos na economia mundial. Com a importância dos EUA para a economia internacional, seu estado econômico tem efeito sobre inúmeras economias, com canal de transmissão pelo mercado internacional de bens, serviços e financeiro.

Tão importante quanto a bolsa de valores, os recursos energéticos são fatores estratégicos para a economia. O crescimento econômico precisa de fontes energéticas; a sua escassez constitui em forte restrição para o aumento da atividade econômica e até mesmo provoca recessão. A energia é um insumo de produção básico para toda economia, envolvendo todos os setores e seus preços. O petróleo em especial, o qual é a principal fonte de energia primária mundial, tem impactos sobre toda economia - variações extremas das cotações de seu preço levam a choques inflacionários que se transmitem ao longo de toda a cadeia produtiva, originando impactos macroeconômicos.

Neste trabalho são analisadas as séries de retornos do índice S&P500 e as cotações do petróleo Brent no período entre 20/05/1987 a 31/12/2009 (Gráficos 1 e 2). Diferenças no número de observações ocorrem devido a calendários de transações distintos. Na estimação dos modelos de volatilidade, consideram-se os retornos dos ativos, especificamente os log-retornos multiplicados por 100, conforme abaixo:

$$r_t = 100 * [\ln(P_t) - \ln(P_{t-1})],$$

onde P_t é a cotação no período t .

O índice S&P 500 (Composite Index Standard and Poor 500) corresponde a uma carteira de papéis das 500 principais empresas americanas com ações na bolsa de valores. O índice da bolsa é uma carteira de portfólio representativa das empresas com maior volume de negociação e valorização, constituída para descrever o risco e retorno do mercado, e em certa medida sua liquidez. O índice da bolsa tem o propósito de ser uma referência para a composição de uma carteira de portfólio ótima.

A evolução do índice S&P 500 caracteriza dois momentos no mercado financeiro internacional, um inicial de estável expansão da economia mundial e internacionalização do mercado financeiro crescente, 1987 a 1995; outro de ambiente de crises internacionais e incerteza, 1996 a 2008. Entre 1987 a 1995 o mercado financeiro mostra-se próspero com acentuado crescimento, fora crises iniciais e no final do período que provocam choques negativos no índice, como a crise de 1987 e crise mexicana em 1995, esta última sem repercussão prolongada. O período de 1996 a 2008, apresenta grandes oscilações das cotações, com alternância de movimentos bem definidos de valorização e quedas do índice, e povoados por crises como asiática a partir do segundo semestre de 1997, moratória Russa em 1998 a queda da Nasdaq em 2000, e entre 2008-2009 a prolongada crise Sub-prime.

Na trajetória do índice S&P500 os períodos de maiores variações das cotações são seguidos de aglomerados dos retornos, indicando a presença de *clusters* de volatilidade. Nota-se que os maiores aglomerados de volatilidade coincidem com as quedas do índice, em especial movimentos abruptos.

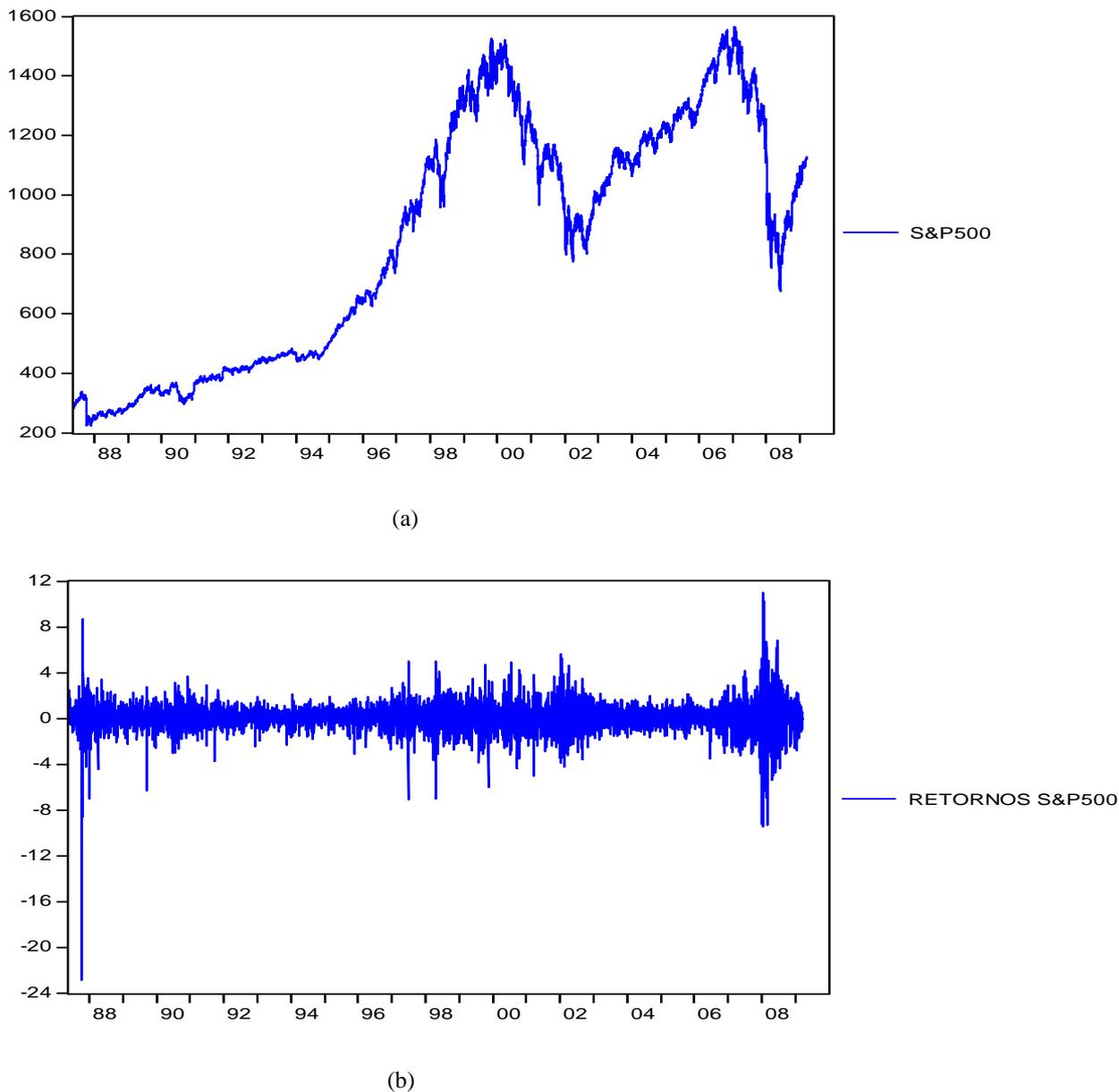
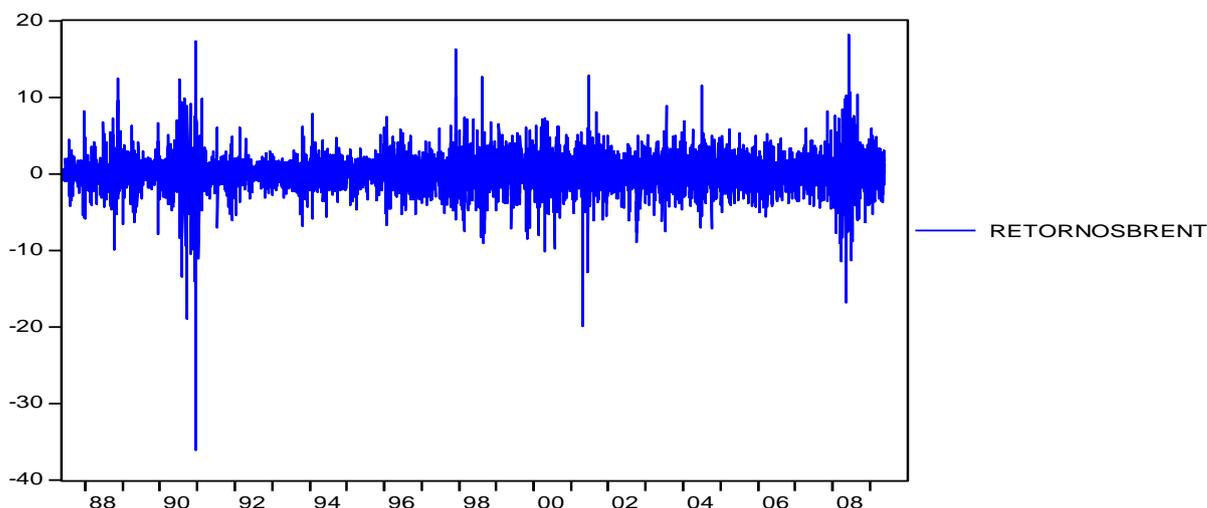


Gráfico 3 – Índice e Retornos: S&P 500.

As cotações de referência para o petróleo são feitas a partir dos chamados petróleos marcadores: WTI, Brent e Dubai. O petróleo tipo Brent é uma cotação de referência para o Mercado europeu, todavia com movimentos muitos similares às cotações dos demais petróleos marcadores. O petróleo é a principal fonte de energia primária mundial, e a concentração das reservas petrolíferas em alguns países, sobretudo no Oriente Médio, tem caracterizado a importância de fatores geopolíticos para as cotações. Ademais, o Cartel da OPEP (Organização dos Países Exportadores de Petróleo) tem importante margem de manobra sobre os preços.



(a)



(b)

Gráfico 4 – Índice e Retornos: Cotações Petróleo Brent.

As cotações do petróleo Brent (Gráfico 4) tem um comportamento instável com a influência de fatores geopolíticos, além dos movimentos usuais de oferta e demanda. Sua evolução é marcada por um movimento de cotação mais baixa, inicialmente cerca de 20 US\$/Barril, que é interrompido com um salto em 1990 com a Guerra do Golfo após invasão do Kuwait pelo Iraque. A rápida atuação bélica americana permite uma estabilização dos preços. No período de 1994 a 1998 o mercado se apresenta estável, com quedas dos preços em 1998 com um quadro recessivo no Sudeste Asiático e Rússia. A partir de 1998, a OPEP diminui a produção para elevar os preços, e a recuperação econômica contribui para a trajetória de alta dos preços. Em 2001 ocorre

um vale na série. A trajetória de expansão nos preços permanece a partir de 2003, com a economia mundial aquecida e a atuação da China em crescimento acelerado, aumentando a demanda por petróleo. A crise econômica do Sub-prime faz os preços caírem vertiginosamente com a apreensão de redução da demanda, no quadro recessivo da economia mundial, porém os preços voltam a aumentar com a sinalização de recuperação econômica de vários países.

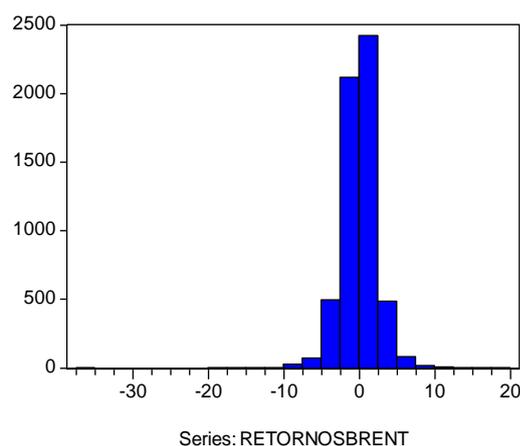
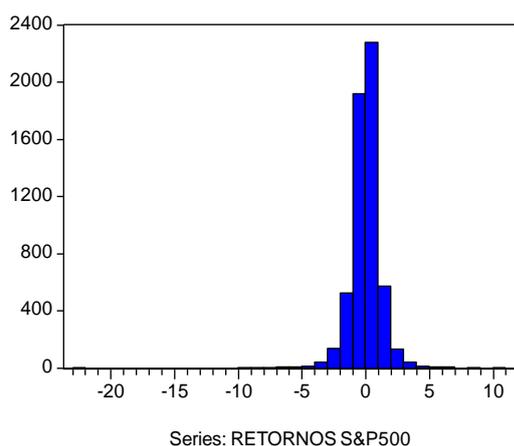
O comportamento instável dos preços com efeito de choques geopolíticos e choques de demanda e oferta faz com que os retornos para as cotações Brent sejam muito instáveis [Gráfico 4 - (b)]. Como resultado, os retornos para o petróleo tipo Brent apresentam “*clusters*” de volatilidade, com agrupamentos de instabilidade e estabilidade se adensando ao longo do tempo, similar as séries financeiras.

Embora os mercados de ações e de petróleo tenham dinâmica distinta, com principais fatores responsáveis pela variação nas cotações sendo específicos às condições prevalentes nestes mercados, determinados momentos de grande instabilidade na economia são comuns a ambas as séries como a crise Sub-prime. Embora sejam mercados particulares temos elementos que afetam as cotações de ambos, não sendo possível entretanto estabelecer-se uma relação de causalidade de maneira bem definida.

As séries de retornos do S&P500 e das cotações Brent exibem muitos dos fatos estilizados dos retornos (Tabela 1). Ambas as séries possuem média próxima de zero, caudas pesadas com excesso de curtose, em valor maior do que o da distribuição normal (que como referência conta com curtose 3), denotando a maior probabilidade de ocorrência de valores extremos dos retornos do que sob a distribuição gaussiana. O maior desvio padrão é para os retornos Brent, embora se tenha caudas mais pesadas para os retornos do índice S&P500. Os histogramas dos retornos (Gráfico 5) ilustram estas características.

Tabela 1 – Estatísticas Descritivas: Retornos S&P 500 e Retornos Brent.

	RETORNOS S&P500	RETORNOSBRENT
Mean	0.024339	0.024608
Median	0.05653	0.016673
Maximum	10.9572	18.12974
Minimum	-22.89972	-36.12144
Std. Dev.	1.215288	2.401766
Skewness	-1.395148	-0.669036
Kurtosis	33.199	17.56154
Jarque-Bera	218597.5	51212
Probability	0	0
Sum	138.8323	141.4476
Sum Sq. Dev.	8422.9	33151.45
Observations	5704	5748

**Gráfico 5 – Histograma: Retornos S&P 500 e Retornos Brent.**

As séries financeiras em alguns momentos são estáveis e em outros instáveis. A dependência no tempo entre os períodos de instabilidade (maior variância) e estabilidade (menor variância) manifesta-se pela correlação dos quadrados dos retornos que corresponde ao efeito ARCH. Ambas as séries de retornos, S&P500 e petróleo Brent, possuem correlação dos quadrados dos retornos ao nível usual de 5% de significância com correlação alta em defasagens pequenas e queda lenta nas demais (Gráfico 6). Isto nos habilita a estimar os modelos de volatilidade para estas séries.

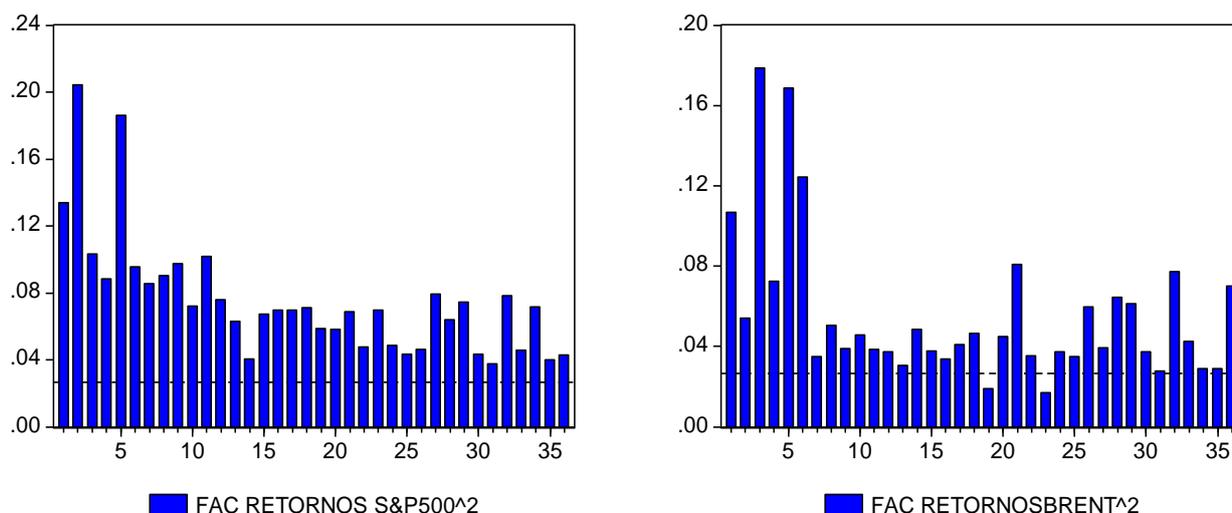


Gráfico 6 – Função de Autocorrelação para Quadrados dos Retornos: S&P 500 e Brent

Nota: -- Nível de significância de 5%.

Na comparação dos modelos GARCH e de redes neurais para a estimação da volatilidade, considera-se o desempenho das estimativas dos modelos levando em conta aspectos quantitativos e qualitativos. Avaliam-se os modelos de acordo com o seu desempenho em termos de critério de erros de previsão em relação a uma medida de volatilidade padrão, e acerca dos intervalos de confiança estimados pelos modelos. Ainda, para investigar a base empírica da forma funcional da relação entre os retornos e a volatilidade, confrontam-se as aproximações de curvas de chegada de informação ajustadas pelos modelos GARCH e RNA.

Na estimação da volatilidade para as séries financeiras utiliza-se um grande intervalo de tempo, com períodos de estabilidade e instabilidade. O ampliado período de análise (22 anos) permite estudar o comportamento dos modelos em ambos os casos, situações ordinárias e aquelas marcadas por dinâmica complexa sob grandes turbulências. Podemos verificar como os modelos incorporam a informação passada. A estimação dos modelos é baseada nas observações até 20/07/2007 cobrindo as 5120 primeiras observações, a amostra restante é para teste dos modelos. Isto nos permite comparar o desempenho dos modelos em momentos antecedentes à crise, 21/07/2007 a 15/05/2008, e momentos após a crise, como aquele que incorpora o evento atípico da crise Sub-prime, 16/05/2008 a 31/12/2009.

A estimação dos modelos de volatilidade da família GARCH é baseada nos modelos GARCH e T-GARCH. Começando pela série do índice S&P500, para o modelo GARCH a ordem mais adequada foi o modelo GARCH(1,1), com todos os coeficientes significativos e menores valores nos critérios AIC/BIC. Como normalmente ocorre em séries financeiras, a persistência é alta, $\alpha_1 + \beta_1 = 0.9904$, com efeito prolongado de um choque passado para a previsão da volatilidade. Por sua vez, o modelo T-GARCH(1,1) evidencia a assimetria da volatilidade frente aos retornos, com o coeficiente da dummy para retornos negativos significativo, com retornos negativos tendo impacto sobre a volatilidade cerca de 10 vezes maior que um retorno positivo. No modelo TGARCH todos os coeficientes são significativos ao nível de significância de 5 % e os critérios de informação indicam melhor ajuste em relação ao modelo GARCH.

Tabela 2 – Modelos GARCH e TGARCH para os Retornos S&P 500.

(1) <i>RETORNOS S&P500</i> =	0.055594	-	0.038133 r_{t-5}
z-statistic (p-value)	4.9780 (0.0000)		-2.6214 (0.0088)
$h_t =$	0.014002 +	0.0858 X_{t-1}^2	+ 0.9046 h_{t-1}
p-value	(0.0000)	(0.0000)	(0.0000)
<i>Estatísticas</i>			
Q(10) = 12.3460 (0.1950), Q(20) = 24.517 (0.1770) ; Q_2(10) = 4.5172 (0.8740) , Q_2(20) = 10.0190(0.9520) ;			
; Adjusted R-squared= -0.003175, Log likelihood= -6732.184, AIC = 2.6343, SBC = 2.6407; DW = 2.0087			
(2) <i>RETORNOS S&P500</i> =	0.031584	-	0.030109 r_{t-5}
z-statistic (p-value)	2.8184 (0.0048)		-2.1321 (0.0330)
$h_t =$	0.018134 +	0.0111 X_{t-1}^2	+ 0.1226 $d_{t-1} * X_{t-1}^2$ + 0.9099 h_{t-1}
p-value	(0.0000)	(0.0364)	(0.0000) (0.0000)
<i>Estatísticas</i>			
Q(10) = 11.5730 (0.2380), Q(20) = 24.2030 (0.1880) ; Q_2(10) = 1.9355 (0.9920) , Q_2(20) = 6.3122 (0.9970) ;			
; Adjusted R-squared= -0.002195, Log likelihood= -6667.787, AIC = 2.6095, SBC = 2.6172; DW = 2.0094			
Nota: Q_k(n) denota a estatística Ljung-Box para a série dos erros k=1, e para a série dos erros ao quadrado k=2; (), p-value.			

A resposta dos retornos sobre a volatilidade é representada pela curva de chegada de informação, que é uma medida do impacto da informação do retorno do período prévio sobre a volatilidade. As estimativas de curva de chegada de informação pelos modelos GARCH para a série S&P500 são apresentadas no Gráfico 7. O modelo GARCH resulta numa curva de chegada de informação quadrática, enquanto o modelo TGARCH representa o efeito alavancagem com maior inclinação para retornos negativos – maior impacto dos retornos negativos sobre a volatilidade em relação a um retorno positivo de mesma magnitude.

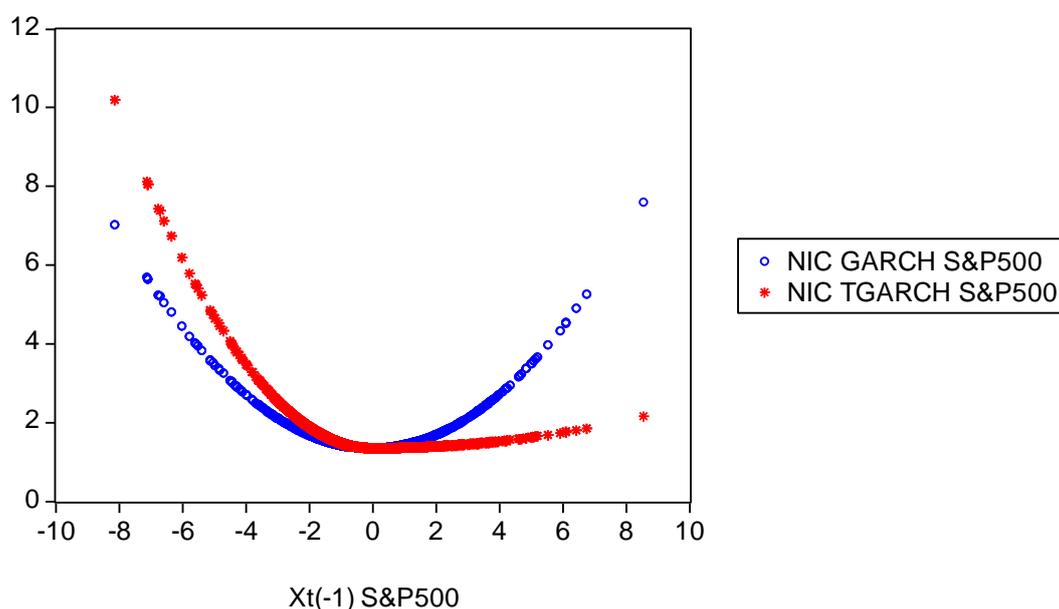


Gráfico 7 – Curva de Chegada de Informação GARCH e TGARCH: Retornos S&P 500

A estimação dos modelos GARCH para a série de retornos Brent encontra-se na Tabela 3. O modelo GARCH com melhor ajuste encontrado foi o modelo GARCH(1,1), maiores ordens apresentaram termos com coeficientes não significativos. No modelo GARCH(1,1) todos os coeficientes são significativos e a persistência elevada, $\alpha_1 + \beta_1 = 0.9928$. A série dos retornos Brent não dá evidências do efeito alavancagem, visto que o modelo TGARCH estimado possui coeficiente sobre os retornos negativos não significativo. A estimativa da curva de chegada de informação para os retornos Brent (Gráfico 8) sugere o efeito simétrico dos retornos sobre a volatilidade.

Tabela 3 – Modelos GARCH e TGARCH para os Retornos Brent.

(1) $RETORNOSBRENT = 0.047731r_{t-1}$
z-statistic (p-value) 3.2201 (0.0013)

$$h_t = 0.05881 + 0.0850X_{t-1}^2 + 0.9078h_{t-1}$$

p-value (0.0000) (0.0000) (0.0000)

Estatísticas

Q(10) = 8.7693 (0.4590), Q(20) = 22.5060 (0.2600) ; Q_2(10) = 20.7440 (0.0140) , Q_2(20) = 30.3510 (0.0470) ;
; Adjusted R-squared= 0.000304, Log likelihood= -10944.71, AIC = 4.2777, SBC = 4.2827; DW = 2.0247

(2) $RETORNOSBRENT = 0.047546r_{t-1}$
z-statistic (p-value) 3.1997 (0.0014)

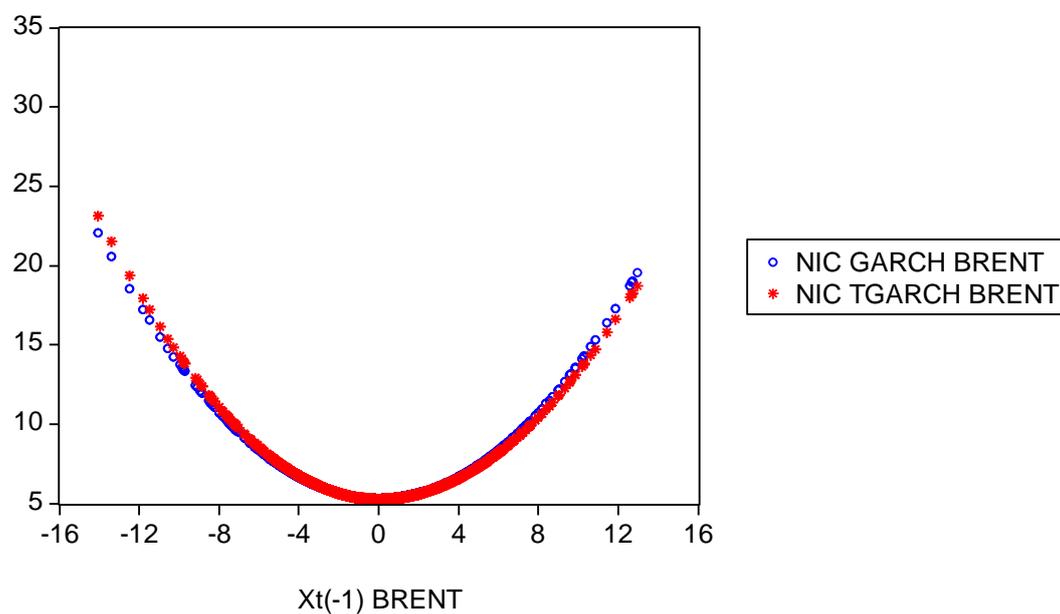
$$h_t = 0.059992 + 0.0803X_{t-1}^2 + 0.0101d_{t-1} * X_{t-1}^2 + 0.9073h_{t-1}$$

p-value (0.0000) (0.0000) (0.1327) (0.0000)

Estatísticas

Q(10) = 8.7616 (0.4600), Q(20) = 22.3380 (0.2680) ; Q_2(10) = 20.0920 (0.0170) , Q_2(20) = 29.4740 (0.0590) ;
; Adjusted R-squared= 0.000114, Log likelihood= -10944.2, AIC = 4.2779, SBC = 4.2843; DW = 2.0243

Nota: Q_k(n) denota a estatística Ljung-Box para a série dos erros k=1, e para a série dos erros ao quadrado k=2; (), p-value.

**Gráfico 8 – Curva de Chegada de Informação GARCH e TGARCH: Retornos Brent**

Para a estimação da curva de chegada de informação usando redes neurais considera-se uma arquitetura de rede neural com apenas um insumo, o retorno passado, e a variável meta como os quadrados dos retornos⁷. A estimação da curva de chegada de informação por redes neurais (Gráfico 9) revela uma forma não linear para a volatilidade: resposta maior da volatilidade para maiores retornos passados; assimetria, retornos negativos tendem a ter maior impacto sobre a volatilidade que retornos positivos; mudança de regime, o efeito dos retornos sobre a volatilidade varia de acordo com a magnitude dos retornos e seu sinal, porém sobre uma relação funcional distinta. Embora a relação da volatilidade estimada por redes neurais ofereça algum suporte empírico ao modelo GARCH, com respeito a não linearidade na forma da resposta não linear e assimetria, este é limitado no que diz respeito a mudança de regime. A mudança de regime para a volatilidade é denotada por redes neurais com uma relação dos retornos sobre a volatilidade que é diferente para sub-amostras distintas, determinados intervalos de retornos apresentam um processo para a volatilidade que seguem uma regra diferente.

Apesar de RNA possuírem capacidade de aproximar relações complexas entre variáveis, sendo um modelo que se aproxima de métodos não paramétricos, precisa de maior número de observações para representação adequada de um processo. Assim, as estimativas em regiões com um menor número de observações – os valores extremos dos retornos – são pobres e seguem padrões que tem representação limitada acerca do processo gerador dos dados.

A aproximação da curva de chegada de informação indica a forma funcional da relação entre os retornos passados e a volatilidade. Os modelos de redes neurais chegam aos mesmos resultados que os modelos GARCH em termos de uma relação para a volatilidade como modelo não linear e assimétrico. Os modelos de redes neurais diferem dos modelos GARCH ao sugerir um modelo com mudança de regime para a volatilidade, para estes dados e extensão da amostra de análise.

⁷ Na estimação da curva de chegada de informação por redes neurais estima-se a relação da volatilidade com apenas um retorno defasado como insumo. O objetivo é isolar o efeito do retorno passado sobre a volatilidade corrente. Porém diferente do modelo GARCH, onde podemos separar o efeito do choque passado sobre a volatilidade contemporânea de sua dependência temporal sobre a volatilidade anterior, nesta curva de chegada de informação usando RNA o retorno com um *lag de* defasagem acaba por reproduzir toda a relação da volatilidade.

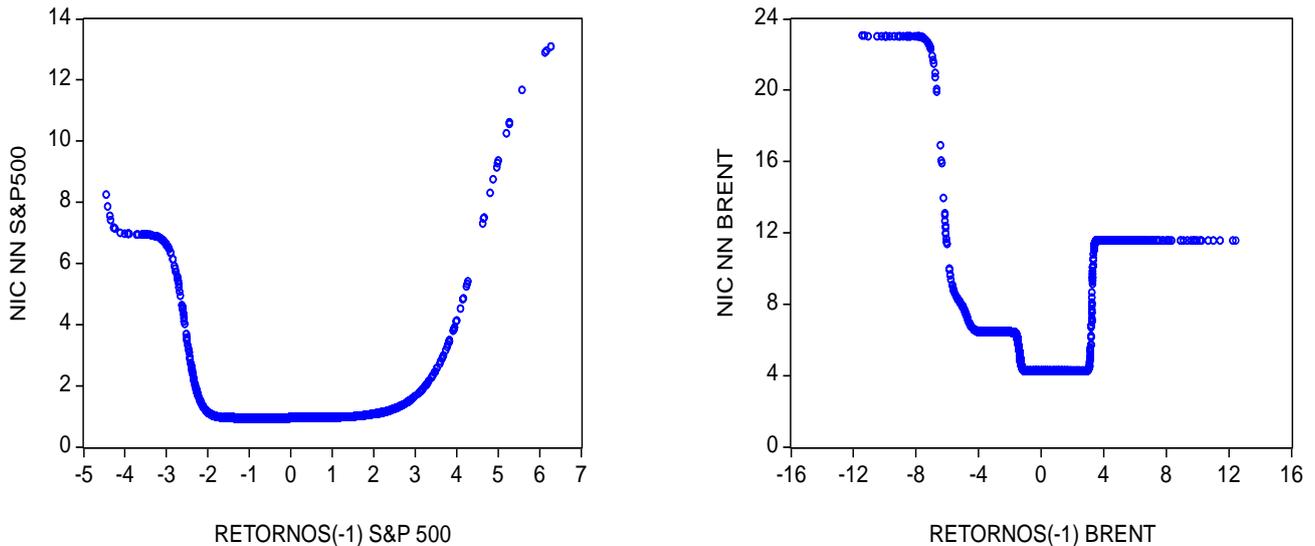


Gráfico 9 – Curva de Chegada de Informação Aproximada por Redes Neurais: Retornos S&P500 e Retornos Brent

Nota: Aproximação de curva de chegada de informação por redes neurais baseada em arquitetura: NN(1,3,1) para os retornos S&P500, e NN(1,6,1) para retornos Brent. Os modelos de redes neurais foram selecionados a partir de distintas arquiteturas e o modelo final selecionado de acordo com melhor desempenho sobre uma amostra de teste.

Continuando a comparação dos modelos GARCH e redes neurais para a estimação da volatilidade, confrontam-se seus desempenhos preditivos para a previsão um passo a frente em termos de uma medida de volatilidade padrão. Como medida de volatilidade para base de comparação do desempenho preditivo dos modelos empregam-se os quadrados dos retornos. Seja a série de retornos, X_t , com $E[X_t|F_{t-1}] = 0$. Note que o modelo de volatilidade com heterocedasticidade condicional, onde $X_t = \sqrt{h_t} * v_t$ com $v_t \sim N(0,1)$, implica $VAR(X_t|F_{t-1}) = E[X_t^2|F_{t-1}] = h_t$.

Os modelos RNA não têm a capacidade de definir relação causal, e usamos informação prévia teórica ou empírica para definir os insumos da rede neural. Para estimação da volatilidade por redes neurais utiliza-se como variável meta os quadrados dos retornos e como insumos os retornos passados, seguindo a estrutura dos modelos GARCH. Uma forma de definir os insumos para análise de séries temporais por redes neurais é a partir de sua função de autocorrelação.

A definição da arquitetura de uma rede neural não possui base em resultado analítico e define-se esta por mineração dos dados, comparando arquiteturas alternativas de neurônios na camada escondida e insumos de acordo com seu desempenho sobre uma amostra de teste. Alternativas como *early-stopping*, determinando o desempenho da rede neural treinada sobre um conjunto de treinamento com término do treinamento

após aumento do erro sobre uma amostra de validação é uma alternativa útil, porém nada garante que o critério de erro não poderia novamente diminuir após maior número de iterações (KAASTRA, BOYD, 1996).

Um modelo RNA é definido basicamente pelos insumos e metas; arquitetura da rede, número de camadas escondidas, unidades na camada escondida e dimensão do vetor de insumos; algoritmo de treinamento; e números de épocas de treinamento, número de iterações durante o processo de aprendizado. Após a comparação de distintas arquiteturas para as redes neurais, os modelos RNA definidos para estimar a volatilidade da série de retornos S&P 500 e retornos Brent, como redes neurais de Perceptron de múltiplas camadas com uma camada escondida alimentada para frente treinados por retro-propagação, possuem parâmetros de acordo com a Tabela 4. A arquitetura definida para os retornos S&P500 apresenta 5 defasagens dos retornos como insumos e seis neurônios na camada escondida; enquanto os retornos Brent contam com uma arquitetura com 8 neurônios na camada escondida e 6 lags de retornos para o vetor de insumos. As demais especificidades seguiram os critérios previamente definidos pelo simulador de redes neurais do MATLAB (*Neural Network Toolbox*) na sua versão 6.5.

Tabela 4 – Especificações dos Modelos de Redes Neurais para Estimação da Volatilidade: S&P500 e Brent

Rede neural	Série	
	Retornos S&P 500	Retornos Brent
Tipo	Multilayer Perceptron	Multilayer Perceptron
Arquitetura	NN(5,6,1)	NN(6,8,1)
Função de Ativação	- Logística na camada escondida - Função identidade na camada de saída	- Logística na camada escondida - Função identidade na camada de saída
Método de Aprendizado	Retro-propagação	Retro-propagação
Método de Operação	Treinamento por Batches	Treinamento por Batches
Algoritmo de Treinamento	Levenberg-Marquardt	Levenberg-Marquardt
Nº de épocas de Treinamento	1000	1000

A análise do desempenho de predição dos modelos considera a diferença entre os valores previstos e os realizados utilizando uma função perda. O melhor ajuste corresponde ao menor critério de erro. As medidas de previsão que são comparadas são a Raiz do Erro Quadrático Médio (*Root Mean Squared Error – RMSE*) e o Erro Percentual Absoluto Médio (*Mean Absolute Percent Error - MAPE*), definidas como

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{Y}_t - Y_t)^2}; \quad (3.19)$$

$$MAPE = \frac{1}{T} \sum_{t=1}^T \left| \frac{\hat{Y}_t - Y_t}{Y_t} \right|. \quad (3.20)$$

O RMSE aproxima o erro médio de previsão, o erro de previsão esperado. O MAPE avalia o erro médio de previsão, porém pondera a magnitude do erro de previsão em relação ao valor observado em termos percentuais. O RMSE penaliza mais os erros sobre valores observados maiores, enquanto o MAPE impõe menor peso a um erro sobre uma observação de maior magnitude – um mesmo erro de dada magnitude tem maior custo no RMSE em relação ao MAPE sobre maiores realizações da série comparativamente a pequenos valores.

As medidas de critério de ajuste das previsões um passo a frente da volatilidade são avaliadas considerando os modelos estimados sobre as 5120 primeiras observações, período de 20/05/1987 a 20/07/2007; o restante da amostra é usado para teste. A parte da amostra reservada ao teste dos modelos, de 21/07/2007 a 31/12/2009, cobre um período de grande turbulência no mercado internacional com o evento atípico da crise Sub-prime. Isto nos habilita a fazer um teste de *stress* dos modelos para seu desempenho sobre situações extremas. Comparam-se os modelos com divisão da amostra de teste: um primeiro intervalo de 03/07/2007 a 15/05/2008, caracterizando um período de normalidade, denominado período normal; e um segundo intervalo 16/05/2008 a 31/12/2009, caracterizando um momento atípico que engloba a prolongada crise Sub-prime, denominado período de crise.

Ainda faz-se uma divisão entre amostra de treinamento e simulação. A amostra de treinamento foi usada para estimar os modelos GARCH e treinar a rede neural. A amostra de simulação consiste nas previsões GARCH e dos modelos neurais. A previsão da volatilidade um passo a frente por RNA utiliza a rede neural treinada e simula sua resposta sobre as observações de insumo fora da amostra de treinamento. Esta divisão é justificada, pois os modelos neurais estão sujeitos ao fenômeno de *over-*

fitting, em que super ajustam a amostra de treinamento com respostas inadequadas para insumos fora da amostra de treinamento. Em geral os modelos de rede neural apresentam ajuste muito bom sobre a amostra de treinamento com critérios de ajuste com desempenho elevado, como coeficiente de determinação próximo da unidade; porém sua generalização é limitada.

Os critérios de desempenho preditivo para previsão da volatilidade um passo a frente (Tabela 5) indicam performance similar entre os modelos GARCH e redes neurais para aproximar a volatilidade. Pelo RMSE na amostra de treinamento, como esperado, o desempenho dos modelos RNA é superior aos modelos GARCH. A análise do RMSE sobre a amostra de simulação indica medidas muito próximas entre os modelos GARCH e RNA, para ambas as séries, com resultados ligeiramente melhores para os modelos GARCH (RNA apresentaram menos desvantagem nas estimativas para série de retornos Brent). Os resultados não se alteram muito sobre o período de normalidade ou de crise. Por sua vez, o MAPE indica melhor desempenho para os modelos RNA, novamente os resultados se mantêm similares sobre o período de crise e de normalidade em ambas as séries.

Tabela 5 - Critérios de Desempenho Preditivo GARCH(1,1) e Redes Neurais para as Séries de Retornos S&P 500 e Brent.

Modelo/Serie	RMSE treinamento	RMSE simulação	RMSE		MAPE treinamento	MAPE simulação	MAPE	
			normal	crise			normal	crise
GARCH(1.1) - S&P500	7.8282	10.125	2.6548	12.6514	2193.9	276.6554	395.305	203.742
NN(5.6.1) - S&P500	6.8449	11.0777	2.8165	13.8527	2760.6	142.4213	202.195	105.689
GARCH(1.1) - Brent	23.1174	22.1958	4.9111	27.2887	73.9152	435.2141	697.441	215.861
NN(6.8.1) - Brent	21.8805	22.8854	5.4086	28.1021	79.4929	369.7532	639.504	144.106

Nota: O calculo do MAPE exclui valores nulos para retornos observados, que tornariam esta medida infinita.

Um aspecto interessante é que os modelos de redes neurais combinam menor critério RMSE com maior MAPE na amostra de treinamento, relativamente ao modelo GARCH. Isto sugere uma espécie de *tradeoff*, melhor adaptação sobre os dados com ajuste de forma flexível sobre as observações dentro da amostra de treinamento implica em menor erro médio, porém à custa de maior MAPE.

O resultado dos critérios de ajuste indica que os modelos GARCH, confrontados com RNA, apresentam desempenho ligeiramente melhor em termos de RMSE e pior considerando o MAPE. As diferenças dos modelos GARCH e neurais seriam mais qualitativas que quantitativas, os modelos GARCH respondem a volatilidade mais precisamente em média, sobre os retornos grandes e ordinários, enquanto redes neurais proporcionam melhores estimativas da volatilidade sobre os resultados normais dos retornos.

Para completar a comparação dos modelos GARCH e redes neurais são considerados os intervalos de confiança gerados com base nas estimativas de volatilidade destes modelos. As estimativas de intervalos de confiança para séries com volatilidade são importantes instrumentos para apoio a decisão e devidamente adaptadas resultam no valor em risco.

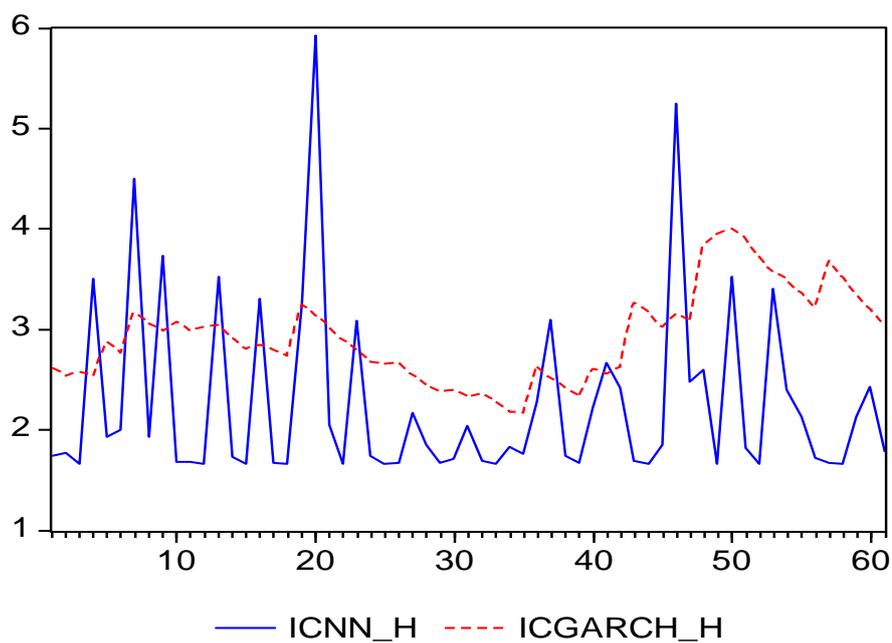
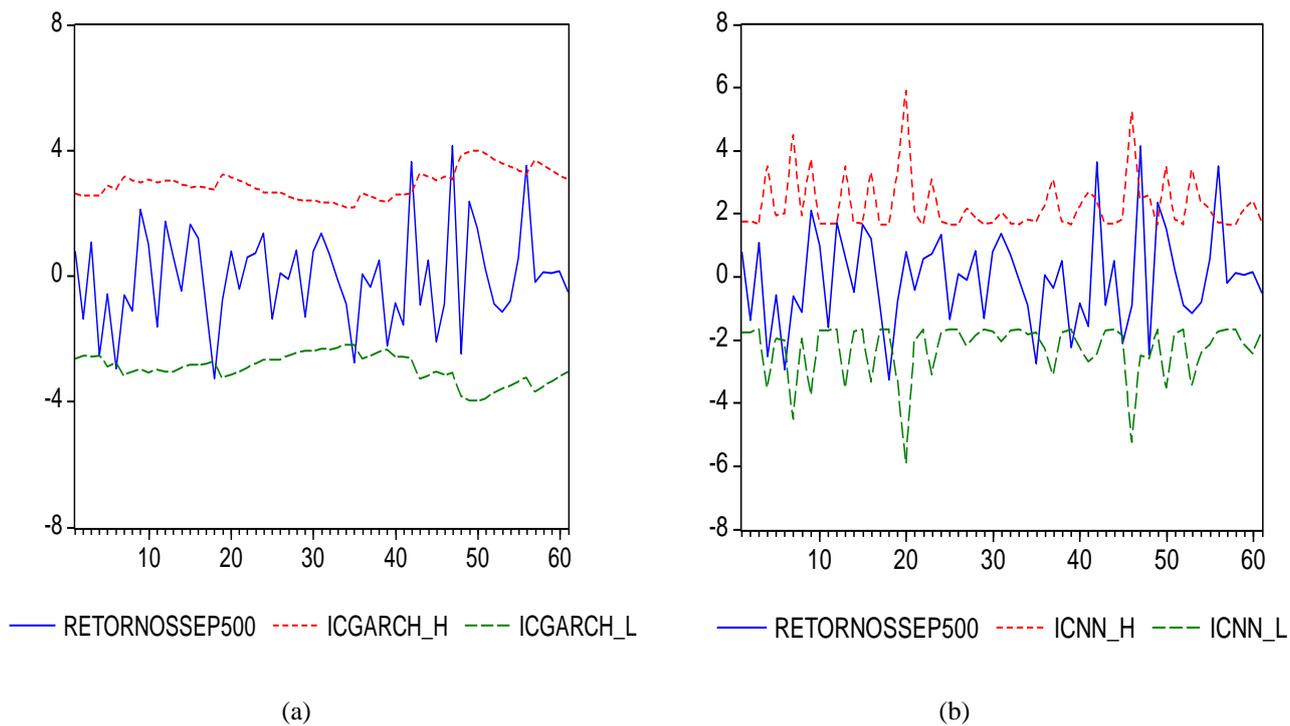
Para a série de retornos S&P 500 (Gráfico 10), apresento os intervalos de confiança em um período de retornos de menor magnitude, sobre o período de 01/2008 a 4/2008. Os retornos diários sobre esta amostra estão entre 4% a -4%, com valores compatíveis a um período comum no mercado financeiro. Os modelos GARCH proporcionam intervalos de confiança suavizados sobre as observações dos retornos, enquanto redes neurais têm comportamento mais abrupto no acompanhamento da trajetória dos retornos. Os intervalos de confiança por redes neurais respondem mais fortemente a um retorno que aumenta, sendo caracterizado por picos, e é mais estreito que a estimativa dos modelos GARCH.

Para a série de retornos Brent (Gráfico 11) os intervalos de confiança são apresentados sobre um período de alta instabilidade no auge da crise Sub-Prime, período 12/2008 a 02/2009, com retornos diários entre 20% a -20%. Os intervalos de confiança GARCH exibem comportamento suavizado e as redes neurais matem picos. Os retornos *outliers* permitem verificar a principal diferença entre as estimativas de volatilidade pelos modelos GARCH em relação aos modelos de redes neurais. As estimativas de volatilidade pelo modelo GARCH com a elevada persistência, propagam por longo período um choque de retornos elevados, enquanto a rede neural segue a trajetória dos retornos logo após poucos períodos de um retorno outlier. Por outro lado, no momento imediatamente seguinte a um grande choque a estimativa GARCH é suavizada incorporando a informação passada, enquanto a resposta por RNA é abrupta.

Os resultados dos intervalos de confiança para os modelos GARCH e redes neurais corroboram os critérios de previsão. Redes neurais artificiais proporcionam

estimativas de volatilidade com repostas abruptas a *outliers*, porém que acompanham a série em suas observações ordinárias. Os modelos GARCH possuem estimativas mais suavizadas para a volatilidade, resultando em boas respostas em momentos imediatamente posteriores a grandes choques, porém prolongam o efeito destes maiores choques prejudicando as estimativas de volatilidade dos retornos alguns instantes subsequentes a observações extremas.

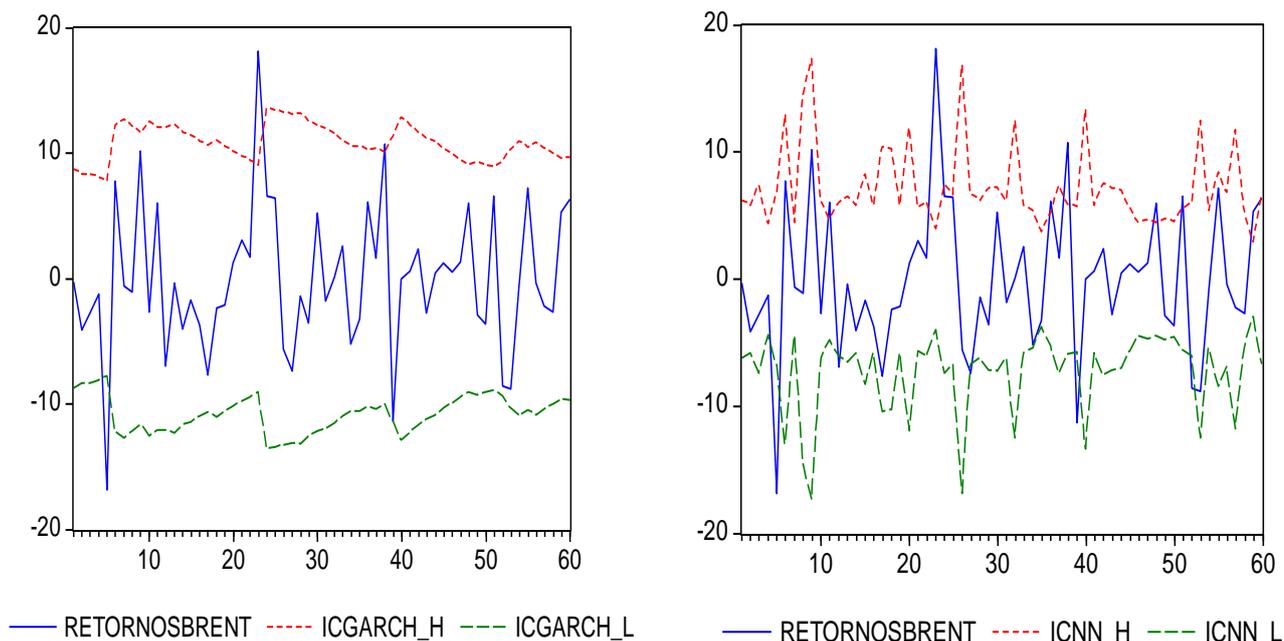
As estimativas da volatilidade são usadas em prática financeira para precificação de opções e ativos, estimativas inadequadas da volatilidade podem implicar em precificação equivocada de ativos. Estimativas da volatilidade com modelos GARCH seriam potencialmente enviesadas, nos casos de persistência enviesada, com sobre predições em situações que sucedem alguns instantes após retornos extremos. Os modelos neurais permitiriam estimativas da volatilidade mais adequadas para situações de normalidade no mercado financeiro, possuindo desempenho limitado após os retornos aberrantes, mas poucos períodos após as variações abruptas levariam a estimativas adequadas.



(c)
Gráfico 10 – Intervalos de Confiança GARCH(1,1) e Redes Neurais para Retornos S&P 500 (Período de 10/01/2008 a 08/04/2008).

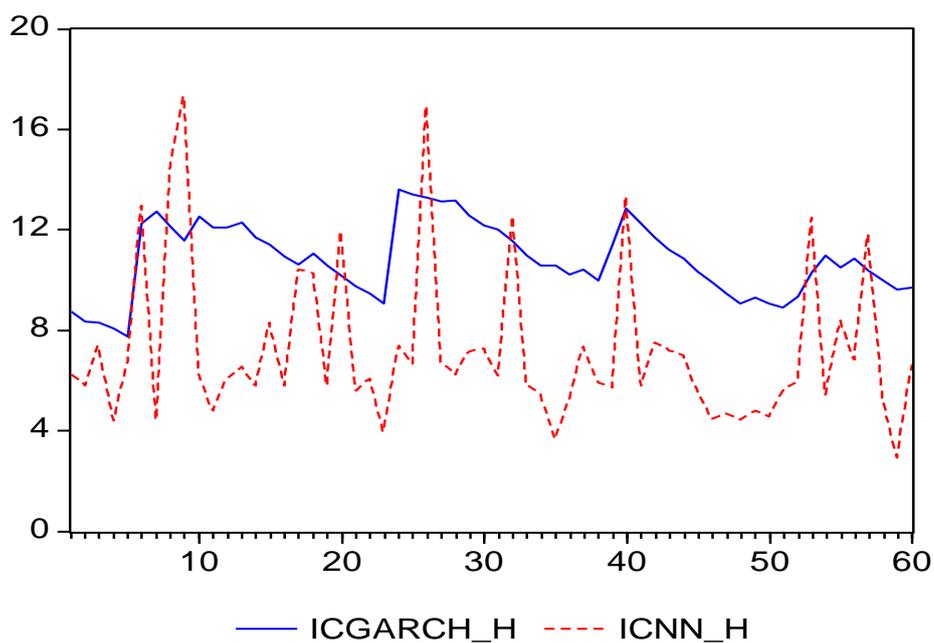
Nota: (1) Intervalos de confiança ao nível de confiança de 95%, $2 * \sigma_{t|F_{t-1}}$.

(2) O representação Ic_H denota intervalo de confiança superior; enquanto Ic_L denota intervalo de confiança inferior.



(a)

(b)



(c)

Gráfico 11 – Intervalos de Confiança GARCH(1,1) e Redes Neurais para Retornos Brent (Período de 1/12/2008 a 26/02/2009).

Nota: (1) Intervalos de confiança ao nível de confiança de 95%, $2 * \sigma_{t|F_{t-1}}$.

(2) O representação Ic_H denota intervalo de confiança superior; enquanto Ic_L denota intervalo de confiança inferior.

4 Modelo Híbrido GARCH Aumentado por Redes Neurais

Os modelos de volatilidade da família GARCH possuem bons resultados na estimação de volatilidade, no entanto muitas vezes não reproduzem toda a não linearidade presente na volatilidade. Dada a capacidade dos modelos de redes neurais em aproximação de relações complexas, uma abordagem atraente é expandir os modelos de volatilidade por redes neurais para estimar a volatilidade.

Neste capítulo é proposto um modelo híbrido de volatilidade aumentado por redes neurais, baseando-se em modificações do modelo NN-GARCH (DONALDSON, KAMSTRA, 1997), uma das abordagens iniciais para incorporar redes neurais ao modelo GARCH na previsão da volatilidade. Outros trabalhos focados em adaptar os modelos de volatilidade da família GARCH com extensões para incorporar não linearidades presentes na volatilidade via redes neurais são Roh (2007) e Bildirici e Ersin (2009). O modelo de Roh (2007) usa redes neurais tendo como insumos os termos GARCH, com os coeficientes GARCH previamente estimados. Bildirici e Ersin (2009) a partir da especificação do modelo de Donaldson e Kamstra (1997) que se baseia no modelo TGARCH generalizam esta proposta para diferentes modelos da classe GARCH: GARCH, EGARCH, TGARCH, SAGARCH, PGARCH, NPGARCH e APAGARCH.

Considerando uma abordagem que associa o modelo GARCH e RNA para a previsão da volatilidade, neste trabalho propõe-se um modelo original que combina redes neurais e o modelo GARCH. Descreve-se também o procedimento para a sua estimação, e adicionalmente é executado um exercício de aplicação empírica.

Na seqüência deste capítulo a seção 4.1 apresenta e discute o modelo NN-GARCH. A seção 4.2 expõe uma nova proposta de um modelo híbrido GARCH aumentado por redes neurais. A seção 4.3 estima o modelo da seção 4.2 para as séries de retornos S&P500 e Brent.

4.1 Modelo Artificial Neural Network–GARCH

O modelo NN-GARCH – Neural Network GARCH (DONALDSON, KAMSTRA, 1997) tem o propósito de incorporar efeitos de assimetria entre os retornos passados e a volatilidade futura não representados pelos modelos GARCH. O modelo

NN-GARCH baseia-se nas redes neurais, enquanto modelos não lineares, para expandir a forma de assimetria representada nos modelos GARCH assimétricos.

O modelo NN-GARCH incorpora componentes não lineares semi paramétricos baseados nos neurônios de uma rede neural ao modelo TGARCH. O modelo NN-GARCH(p,q,r,s) é uma extensão do TGARCH(p,q,r) por s componentes logísticos. Seja X_t uma série de retornos não correlacionados ou resíduos de um modelo de série de tempo para os retornos. Assim, o modelo é dado conforme o conjunto de Equações 4.1 – 4.6:

$$X_t = \sqrt{h_t} v_t, \quad v_t \sim N(0,1), \quad (4.1)$$

$$h_t = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j} + \sum_{k=1}^r \varphi_k d_{t-k} X_{t-k}^2 + \sum_{h=1}^s \xi_h \Psi(z_t \lambda_h), \quad (4.2)$$

$$d_{t-k} = \begin{cases} 1, & \text{se } X_{t-k} < 0 \\ 0, & \text{se } X_{t-k} > 0 \end{cases} \quad (4.3)$$

$$\Psi(z_t \lambda_h) = [1 + \exp(\lambda_{h,0,0} + \sum_{d=1}^v [\sum_{w=1}^m (\lambda_{h,d,w} z_{t-d}^w)])]^{-1}, \quad (4.4)$$

$$z_{t-d} = [X_{t-d} - \hat{\mu}_X] / \hat{\sigma}_X^2, \quad (4.5)$$

$$\frac{1}{2} \lambda_{h,d,w} \sim U[-1, +1], \quad (4.6)$$

onde (4.2) é a equação da variância, (4.3) define as variáveis *dummy* para até r defasagens de retornos negativos, (4.4) descreve a curva logística avaliando a soma ponderada de v defasagens dos retornos passados padronizados de acordo com a Equação (4.5). A seleção aleatória pela Equação (4.6) determina os ponderadores.

O modelo NN-GARCH possui problema de identificação na estimação do parâmetro ξ_h , já que um mesmo termo logístico (Equação 4.4) aparece s vezes na equação da variância com idêntico argumento (exceto o ponderador aleatório). Para superar isto é incluído um termo aleatório com distribuição uniforme, o qual resulta na variação do argumento e proporciona distintos valores para os s componentes logísticos. Os demais parâmetros são estimados por máxima verossimilhança (DONALDSON, KAMSTRA, 1997). A principal extensão incluída nos modelos NN-GARCH para estimar a volatilidade é considerar componentes logísticos inspirados nos modelos de redes neurais, que avaliam potências dos retornos padronizados passados ponderados por termos aleatórios, com o propósito de estender a não linearidade do modelo TGARCH.

A estimação do modelo NN-GARCH começa com a seleção do ponderador aleatório $\lambda_{h,d,w}$ que é distinto para cada h entre os s termos logísticos, e m somas de v retornos não previstos defasados padronizados todos com potencia $w \in [1,m]$. Então prossegue com a estimação por máxima verossimilhança do modelo com a busca da ordem ótima de p, q, r, s, v e m , cada um $\in [0,s]$. A ordem do modelo NN-GARCH (p,q,r,s) é definida pela estimação com melhor critério de informação; no trabalho de Donaldson e Kamstra (1997) empregou-se o SBC(*Schwartz Bayesian Criterion*).

O modelo NN-GARCH, apesar de acrescentar componentes logísticos com estruturas de neurônios ao modelo GARCH, substitui o processo de aprendizado dos pesos da rede neural por um método arbitrário de determinação dos pesos sinápticos dos insumos de uma seleção aleatória. Neste sentido o modelo denominado NN-GARCH é mais um modelo inspirado em redes neurais do que uma implementação de rede neural combinada com modelos GARCH.

4.2 Modelo Híbrido GARCH Aumentado por Redes Neurais

O modelo NN-GARCH tenta associar redes neurais ao modelo GARCH por um procedimento ad-hoc com seleção aleatória de pesos, que mesmo poderiam ser escolhidos arbitrariamente (DONALDSON, KAMSTRA, 1997). Inspirado neste modelo e na idéia de associar redes neurais aos modelos GARCH para a previsão da volatilidade propõe-se um modelo híbrido com parte GARCH e outra RNA.

Um modelo híbrido GARCH aumentado por redes neurais HNN-GARCH(r,s,h) consiste em um modelo GARCH(r,s) estendido por uma rede neural com uma camada escondida com h neurônios de função de ativação logística, conforme abaixo (Equações 4.7 - 4.9):

$$X_t = \sqrt{h_t} v_t, \quad v_t \sim N(0,1) \quad (4.7)$$

$$h_t = \alpha_0 + \sum_{i=1}^r \alpha_i X_{t-i}^2 + \sum_{j=1}^s \beta_j h_{t-j} + \delta * \left[\beta_0 + \sum_{k=1}^h \beta_j * G \left(\sum_{j=1}^m w_{k,i} X_{t-i} + b_k \right) \right], \quad (4.8)$$

$$G(.) = \frac{1}{1 + \exp[-(.)]}, \quad (4.9)$$

$$\alpha_0 \geq 0, \alpha_i \geq 0, \beta_j \geq 0, \delta \geq 0.$$

Acima temos um modelo GARCH combinado com uma rede neural com uma única camada escondida, com insumos correspondentes as defasagens dos retornos. A parte RNA apresenta insumos na mesma ordem dos termos ARCH, ponderados por pesos sinápticos que conectam cada defasagem de retornos ao neurônio k , entre os h neurônios presentes no modelo.

O modelo NN-GARCH emprega pesos aleatórios a combinação linear de termos logísticos que avaliam somas ponderadas dos retornos passados na especificação da equação da variância. Como diferentes pesos levam a distintas formas funcionais não lineares, nessa especificação a seleção aleatória de pesos traz uma forma funcional arbitrária à equação da variância. O modelo HNN-GARCH expande a equação da variância por uma rede neural tendo como meta o erro de previsão GARCH incorporando-se ao modelo de volatilidade uma relação subjacente dos erros de previsão GARCH e os retornos passados não previstos.

A proposta de um modelo híbrido GARCH com redes neurais é descrever não linearidades presentes na volatilidade não capturadas pelo modelo GARCH. Para representar toda não linearidade presente na volatilidade empregamos uma rede neural treinada com os erros de previsão da volatilidade GARCH, que podem ser entendidos como regressores de aspectos não lineares contidos na volatilidade.

A estimação de um modelo GARCH combinado com redes neurais é impraticável por máxima verossimilhança devido ao problema de identificação. Basicamente, teríamos o mesmo problema em estimar os parâmetros de uma rede neural por máxima verossimilhança, um mesmo retorno passado tem parâmetros distintos para cada neurônio na camada escondida, não há como identificar separadamente os valores dos parâmetros de um mesmo retorno em cada neurônio.

O modelo híbrido GARCH aumentado por redes neurais possui um componente de redes neurais com os pesos sinápticos treinados por retro-propagação. A estimação do modelo híbrido HNN-GARCH proposto é feita por um procedimento de três passos:

- (a) Estima-se o modelo $GARCH(r,s)$ e faz-se a previsão da volatilidade sobre a amostra de treinamento;
- (b) Treina-se a rede neural Perceptron com uma camada escondida, $NN(r,h,1)$. Esta tendo como meta a diferença entre a estimativa da volatilidade e os

quadrados dos retornos contemporâneos e insumos as r defasagens de retornos da especificação do modelo GARCH;

- (c) Re-estimamos por máxima verossimilhança o modelo GARCH, usando a especificação da equação da variância do modelo GARCH estendida pelo componente neural obtido na etapa (b), onde aplicamos um coeficiente a ser estimado sobre a saída da rede neural treinada com os erros de previsão da volatilidade GARCH. Nesta etapa estimamos por máxima verossimilhança os parâmetros $\alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_s$ e δ da Equação 4.8.

A estimação do modelo híbrido HNN-GARCH exige um procedimento em três passos para eliminar o problema de identificação e a multicolinearidade que resultaria da tentativa de incorporar diretamente a previsão da volatilidade de redes neurais ao modelo GARCH.

Apesar do custo computacional advindo da implementação do modelo híbrido HNN-GARCH, temos uma forma que parece adequada de combinar o modelo GARCH com redes neurais para a previsão da volatilidade. Ademais, este modelo proporciona um modo de talvez capturar um erro de especificação do modelo de GARCH original. Durante a estimação, a etapa do treinamento da rede neural, tendo como meta a diferença entre os quadrados dos retornos e a previsão da volatilidade GARCH e insumos os retornos passados, é capaz de indicar se o modelo GARCH deixou de incorporar alguma característica presente nos dados para a estimação da volatilidade, ou não.

A previsão da volatilidade por redes neurais possui conteúdo de informação de natureza distinta em relação ao modelo GARCH na previsão da volatilidade (ARAGONÉS, BLANCO, ESTÉVEZ, 2007). Outra motivação para o modelo HNN-GARCH é utilizar conteúdo de informação adicional das redes neurais ao modelo GARCH, proporcionando um aspecto mais suavizado das estimativas de volatilidade em relação àquelas mais abruptas oriundas de RNA.

4.3 Estimação do Modelo Híbrido HNN-GARCH para a Série de Retornos S&P500 e Brent

Nesta seção é estimado o modelo híbrido GARCH aumentado por redes neurais para a previsão da volatilidade sobre as séries de retornos do índice S&P500 e cotações do petróleo tipo Brent. Como anteriormente a estimação é executada sobre a amostra do período de 20/05/1987 a 20/07/2007, as observações seguintes são para teste dos modelos.

A estimação do modelo HNN-GARCH (Tabela 6) sugere a não linearidade adicional presente na volatilidade além da volatilidade GARCH, com os coeficientes sobre a representação não linear da volatilidade por redes neurais significativos para ambas as séries. Na forma híbrida de redes neurais com modelo GARCH todos os coeficientes são significativos, pelo menos ao nível de significância de 5%. Além disso, os resultados são similares seja para os retornos S&P500 seja para os retornos Brent. Os coeficientes não se alteram muito em relação à contrapartida do modelo GARCH (Tabelas 2 e 3), com maior magnitude do parâmetro ARCH e coeficiente GARCH ligeiramente inferior. Considerando a informação apenas do retorno não previsto passado para a resposta da volatilidade corrente, a sua forma não linear neural tem maior peso – maior coeficiente – comparativamente ao retorno passado.

Tabela 6 – Estimação do Modelo HNN-GARCH para as Séries de Retornos: S&P500 e Brent

(1) <i>RETORNOS S&P500</i> = 0.044565 - 0.036185 r_{t-5}	
t-statistic	(3.9860) (-2.4403)
$h_t =$	0.017187 + 0.1009 X_{t-1}^2 + 0.9017 h_{t-1} + 0.1814* NN(1,4,1)
t-statistic	(5.6166) (8.9539) (91.3991) (4.6646)
<i>Estatísticas</i>	
Log likelihood= - 6706.115, AIC = 2.6219, SBC = 2.6296	
(2) <i>RETORNOSBRENT</i> = 0.042273 r_{t-1}	
t-statistic	(2.8026)
$h_t =$	0.037909 + 0.1193 X_{t-1}^2 + 0.9001 h_{t-1} + 0.5606 *NN(1,5,1)
t-statistic	(2.4497) (10.0485) (106.3043) (4.0744)
<i>Estatísticas</i>	
Log likelihood= - 10949.81, AIC = 4.2792, SBC = 4.2856	

O modelo HNN-HGARCH é um modelo parcimonioso, possuindo apenas um parâmetro adicional para aumentar o conteúdo de informação usado na estimação da volatilidade. Na série de retornos S&P500 o modelo de volatilidade híbrido possui melhores critérios de informação em relação ao modelo GARCH; nos retornos Brent resultam em ajuste muito similar ao modelo GARCH com marginal melhora dos critérios de informação. Em ambos os casos os critérios de informação são similares ao modelo TGARCH, no entanto a natureza da não linearidade adicional é diferente.

O desempenho do modelo de volatilidade híbrido em termos de capacidade de predição (Tabela 7) é muito semelhante ao modelo GARCH (Tabela 5). Os resultados dentro da amostra de estimação, e fora desta, não sugerem diferenças. As diferenças mais marcantes se encontram no período de normalidade, prévio a crise Sub-prime, o RMSE tende a diminuir e o MAPE aumentar; e no período subsequente de maior instabilidade o RMSE tende a aumentar e o MAPE a diminuir.

Tabela 7 – Critérios de Desempenho Preditivo Modelo HNN-GARCH para as Séries de Retornos S&P500 e Brent

Modelo/Serie	RMSE	RMSE	RMSE		MAPE	MAPE	MAPE	
	treinamento	simulação	normal	crise	treinamento	simulação	normal	crise
HNN-GARCH(1.1.4) - S&P500	7.9721	11.7576	2.6213	14.7452	2342.2	266.6063	400.876	184.095
HNN-GARCH(1.1.5) - Brent	22.7748	22.7778	4.8954	28.0181	72.8868	422.957	697.041	193.685

Nota: O calculo do MAPE exclui valores nulos para retornos observados.

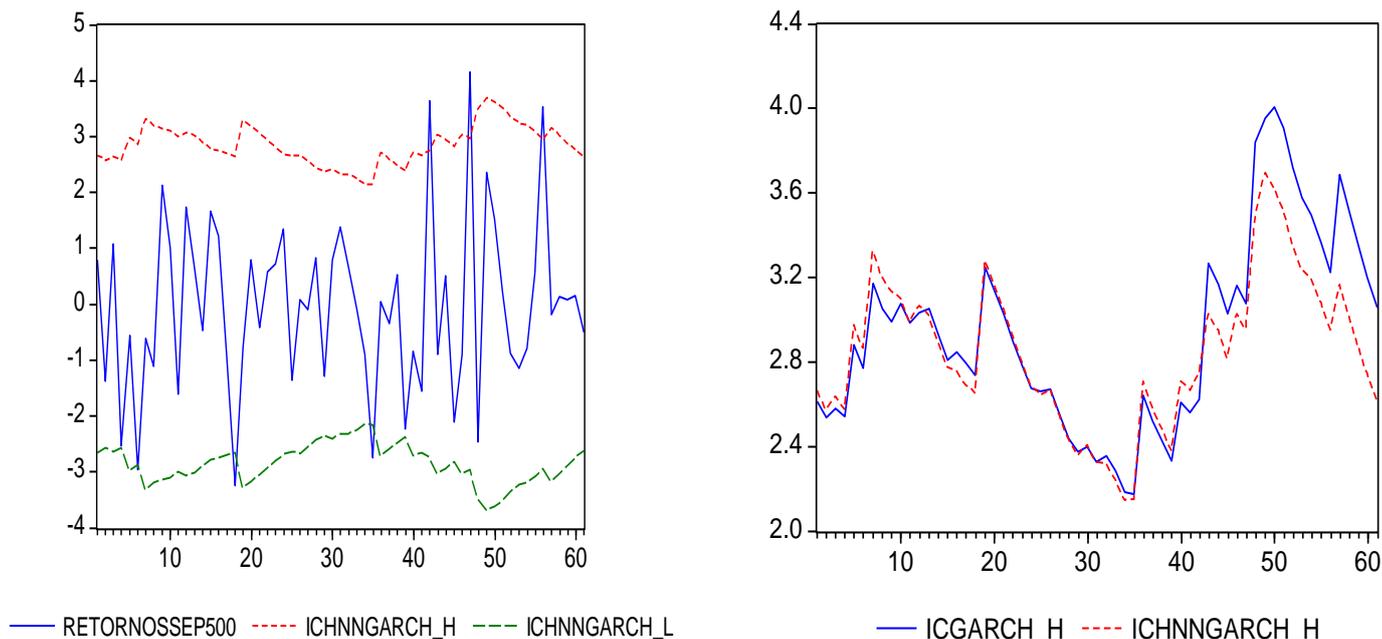


Gráfico 12 – Intervalos de Confiança HNN-GARCH(1,1,4) para Retornos S&P 500 (Período de 10/01/2008 a 08/04/2008).

Nota: (1) Intervalos de confiança ao nível de confiança de 95%, $2*\sigma_{t|F_{t-1}}$.

(2) A representação Ic_H denota intervalo de confiança superior; e Ic_L , intervalo de confiança inferior.

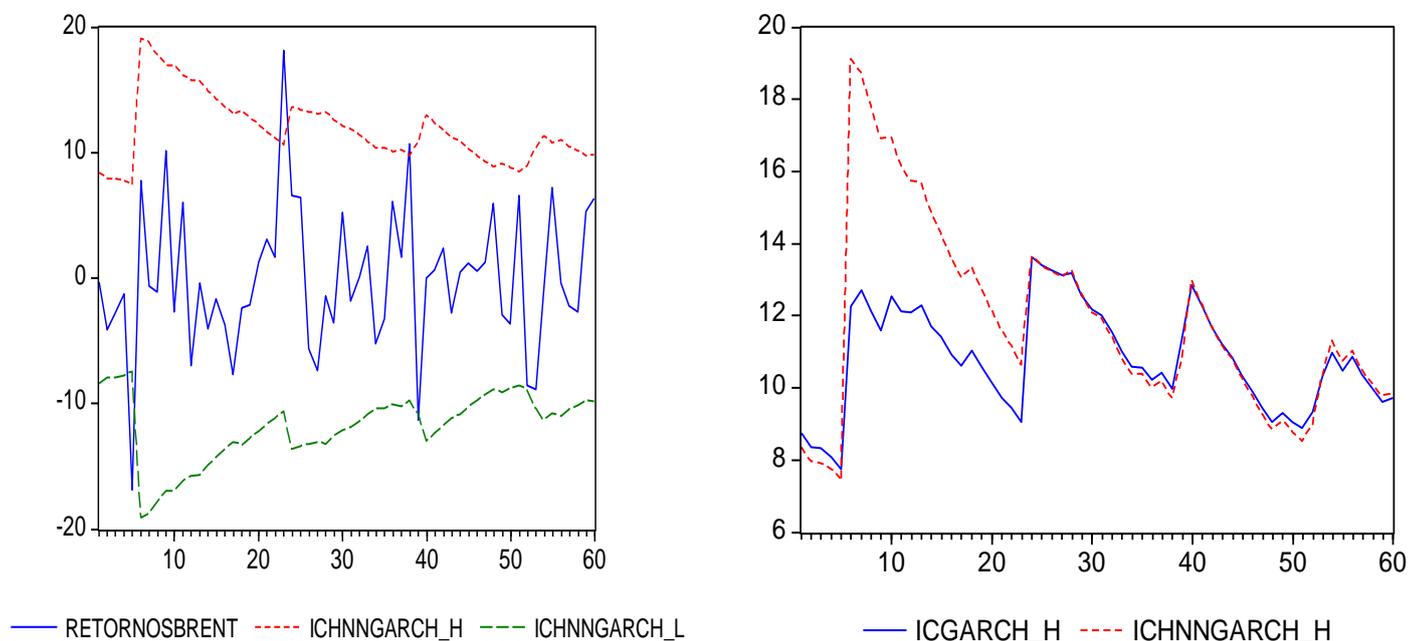


Gráfico 13 – Intervalos de Confiança HNN-GARCH(1,1,5) para Retornos Brent (Período de 1/12/2008 a 26/02/2009).

Nota: (1) Intervalos de confiança ao nível de confiança de 95%, $2*\sigma_{t|F_{t-1}}$.

(2) A representação Ic_H denota intervalo de confiança superior; e Ic_L , intervalo de confiança inferior.

Os resultados de desempenho de previsão são mais bem entendidos pela análise dos intervalos de confiança condicionais - calculados com as estimativas de volatilidade (Gráfico 10 e 11). Em períodos de maior estabilidade e maior turbulência durante a crise, ilustrados sobre a amostra entre 01/2008 a 04/2008 para os retornos S&P500 e entre 12/2008 a 01/2009 para os retornos Brent - respectivamente, o modelo de volatilidade GARCH aumentado por redes neurais produz uma estimativa de volatilidade que tende a estar abaixo do modelo GARCH com menor nível para a volatilidade, mas que responde abruptamente após aumentos dos retornos extremos. Conseqüentemente, se erra mais a meta da volatilidade em períodos de crise, todavia a volatilidade tende a estar mais abaixo do que o modelo GARCH compensando a persistência elevada.

O modelo HNN-GARCH combina as características da estimação da volatilidade por redes neurais com a estimativa GARCH. Por um lado, pelo seu componente GARCH a volatilidade prolonga-se após um retorno extremo e é suavizada; por outro lado, pelo seu componente neural a volatilidade possui resposta acomodando menor nível para a volatilidade, mas é abrupta. O aspecto híbrido para o modelo HNN-GARCH possui características mais adequadas para momentos de estabilidade em relação a momentos de crise.

5 Considerações Finais

A volatilidade apresenta característica não linear: maior variabilidade para alterações das cotações de maior valor em módulo do que sobre magnitudes pequenas; assimetria, variações das cotações diminuindo o valor de um ativo tornam a volatilidade mais pronunciada frente a movimentos de valorização; e *clusters* de volatilidade, com agrupamentos dos retornos caracterizando transição entre alta variância e baixa variância no tempo. Os aglomerados de volatilidade podem assumir a forma de mudança de regime, para séries sobre longo período de análise sob mudança do ambiente econômico, como as séries estudadas neste trabalho. As redes neurais como modelos não lineares e com propriedade de aproximação universal, são modelos que potencialmente permitem descrever mais aspectos da volatilidade que os modelos GARCH.

Na estimação da volatilidade aproximada por redes neurais de camada única - *Feedforward Multilayer Perceptron* com uma camada - comparando-se com a estimação pelos modelos GARCH os critérios de erro de previsão resultaram muito semelhantes entre os modelos GARCH e RNA. Por outro lado, analisando-se os intervalos de confiança a volatilidade GARCH é mais suavizada e prolonga o efeito de um retorno extremo por um longo horizonte de tempo; entretanto a volatilidade estimada por redes neurais acompanha de forma mais próxima a trajetória de retornos, e suas mudanças de comportamento, porém com resposta abrupta a alterações dos retornos de seus níveis verificados no passado mais recente.

A proposta apresentada neste trabalho, um modelo híbrido combinando a estrutura de um modelo GARCH com redes neurais, esta última treinada com o erro de previsão da volatilidade GARCH, contém maior conteúdo de informação para a previsão da volatilidade. O aspecto da volatilidade pelo modelo GARCH aumentado por redes neurais se caracteriza por suavidade, menor nível para a volatilidade e respostas elevadas seguintes a retornos aberrantes, embora mantenha critérios de erros de previsão muito próximos ao modelo GARCH original. No modelo HNN-GARCH o seu componente neural domina a sua estrutura GARCH e aparentemente seria mais adequado para momentos de estabilidade. Neste modelo a resposta abrupta aos maiores retornos prejudica seu desempenho para períodos de maior instabilidade.

Redes neurais podem assumir diversas especificações com pesquisa intensa nos últimos 50 anos. A forma mais tradicional de aplicações de redes neurais, redes alimentadas para frente com camada única (*Feedforward Multilayer Perceptron*), na estimação da volatilidade apresentou aspectos complementares e com não linearidades não descritas pelo modelo GARCH. Melhores desempenhos podem ser alcançados com formas mais complexas de redes neurais, sobretudo que possam descrever a sua dependência dinâmica pela estrutura do modelo, como as redes neurais recorrentes, o que seria uma possível direção para extensão deste trabalho.

Referências

- ARAGONÉS, J. R.; BLANCO, C.; ESTÉVEZ, P. G. Neural Network Volatility Forecasts. *Intelligent Systems in Accounting and Management*, n.15, 107-121, 2007.
- BOLLERSLEV, T. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, v.31, n.3, p.307-327, 1986.
- BOLLERSLEV, T; CHOU, R. Y.; KRONER, K. F. ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence. *Journal of Econometrics*, v.52, p.5-59, 1992.
- BRAGA, A. de P., CARVALHO, A. P. de L. F. de; LUDEMIR, T. B. *Redes Neurais Artificiais: Teoria e Aplicações*. Rio de Janeiro: LTC, 2ª edição, 2007.
- BROOKS, C. Predicting Stock Index Volatility: Can Market Volume Help? *Journal of Forecasting*, v.17, p.59-80, 1998.
- BHAR, R.; HAMORI, S. *Empirical Techniques in Finance*. Germany: Springer, 2005.
- BILDIRICI, M.; ERSIN, O. O. Improving Forecasts of GARCH Family Models with the Artificial Neural Networks: An Application to the Daily Returns in Istanbul Stock Exchange. *Experts System with Applications*, v. 36, p.7355-7362, 2009.
- BISHOP, C. M. *Neural Network for Pattern Recognition*. Oxford: Clarendon Press, 1995.
- CHENG, B.; TITTERINGTON, D. M. Neural Networks: A Review from a Statistical Perspective. *Statistical Science*, v.9, n.1, p.2-54, 1994.
- DORFFNER, G. Neural Network for Time Series Processing. *Neural Network World*, v.6, n.4, p.447-468, 1996.

DONALDSON, R. G.; KAMSTRA, M. An Artificial Neural Network-GARCH Model for International Stock Return Volatility. *Journal of Empirical Finance*, v.4, p.17-46, 1997.

DONALDSON, R. G.; KAMSTRA, M. Forecasting Combining with Neural Networks. *Journal of Forecasting*, v. 15, p.49-61, 1996.

DUEKER, M. J. Markov Switching in GARCH Processes and Mean-reverting Stock Market Volatility. *Journal of Business and Economic Statistics*, v.15, n.1, p.26-34, 1997.

ENDERS, W. *Applied Econometric Time Series*. Second Edition, John Wiley & Sons Inc., 2004.

ENGLE, R. F. Autoregressive Conditional Heteroskedasticity with Estimates of Variance the Variance of U.K. Inflation. *Econometrica*, v.50, n.4, p.987-1008, 1982.

ENGLE, R. F. Riesgo y Volatilidad: Modelos Económicos y Práctica Financiera. *Revista Asturiana de Economía – RAE*, n.31, p.221-252, 2004.

ENGLE, R. F.; LILIEN, D. M.; ROBINS, R. P. Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model. *Econometrica*, v.55, n.2, p.391-407, 1987.

ENGLE, R. F.; NG, V. K. Measuring and Testing the Impact of News on Volatility. *Journal of Finance*, n.5, 1993.

FERNANDES, L. G. L.; NAVAU, P. O. A.; PORTUGAL, M. S. Previsão de Séries de Tempo: Redes Neurais Artificiais e Modelos Estruturais. *Pesquisa e Planejamento Econômico*, Rio de Janeiro, v.26, n.2, p.253-276, ago., 1996.

FRANSES, P. H.; VAN DIJK, D. *Nonlinear Time Series Models in Empirical Finance*. Cambridge University Press, 2000.

FREITAS, S. O.; SOUZA, A. A. Utilização de Redes Neurais na Precificação de Opções. *Resenha BM&F*, n.150, p.63-73, 2001.

GEMAN, S.; BIENENSTOCK, E.; DOURSAT, R. Neural Networks and the Bias Variance Dilemma. *Neural Computation*, v.4, p.1-58, 1992.

HAGAN, M. T.; MENHAJ, M. B. Training Feedforward Networks with the Marquardt Algorithm. *IEEE Transactions on Neural Networks*, v.5, n.6, 1994.

HAMILTON, J. D.; SUSMEL, R. Autoregressive Conditional Heteroskedasticity and Changes in Regime. *Journal of Econometrics*, v.64, p.307-333, 1994.

HAMID, S. A.; IQBAL, Z. Using Neural Networks for Forecasting Volatility of S&P 500 Index Future Prices. *Journal of Business Research*, v.57, p.1116-1125, 2004.

HAYKIN, S. *Neural Networks a Comprehensive Foundation*. Prentice Hall: New Jersey, Second edition, 1999.

HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer Feedforward Networks are Universal Approximators. *Neural Networks*, v.2, p.359-366, 1989.

HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Universal Approximation of an Unknown Mapping and Its Derivatives Using Multilayer Feedforward Networks. *Neural Networks*, v.3, p.551-560, 1990.

KUAN, C.-M; WHITE, H. Artificial Neural Networks: an Econometric Perspective. *Econometric Reviews*, v.13, p.1-143, 1994.

LAMOUREUX, C. G.; LASTRAPES, W. D. Persistence in Variance, Structural Change, and the GARCH Model. *Journal of Business & Economic Statistics*, v.8, n.2, 1990.

LI, H.-X.; LEE, E. S. Interpolation Functions of Feedforward Neural Networks. *An International Journal Computers & Mathematics with Applications*, v.46, p.1861-1874, 2003.

KAASTRA, I.; BOYD, M. Designing a Neural Network for Forecasting Financial and Economic Time Series. *Neurocomputing*, n.10, p. 215-236, 1996.

MEISSNER, G.; KAWANO, N. Capturing the Volatility Smile of Options on High-Tech Stocks – A Combined GARCH-Neural Networks Approach. *Journal of Economics and Finance*, v.25, n.3, 2001.

MORAIS, I. A. C. de; PORTUGAL, M. S. Modelagem e Previsão de Volatilidade Determinística e Estocástica. *Estudos Econômicos*, São Paulo, v.29, n.3, p.303-341, jul.-set., 1999.

MORETTIN, P. A.; TOLOI, C. M. *Análise de Séries Temporais*. São Paulo: Edgard Blücher LTDA., 2004.

MORETTIN, P. A. *Econometria Financeira: Um Curso em Séries Temporais Financeiras*. São Paulo: Edgard Blücher LTDA., 2008.

PAGAN, A. R.; SCHWERT, W. G. Alternative Models for Conditional Stock Volatility. *Journal of Econometrics*, v.45, p.267-290, 1990.

R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. Disponível em: <<http://www.R-project.org>>.

RYDBERG, T. H. Realistic Statistical Modelling of Financial Data. *International Statistical Review*, v.68, n.3, p.233-258, 2000.

ROH, T. H. Forecasting the Volatility of Stock Price Index. *Experts Systems with Applications*, v.33, p.916-922, 2007.

SCHITTENKOPF, C.; DORFFNER, G.; DOCKNER, E. J. *Forecasting Time-dependent Conditional Densities: A Semi-nonparametric Neural Network Approach*. Report Series, n° 36, 1999.

SCHITTENKOPF, C.; DORFFNER, G.; DOCKNER, E. J. Volatility Prediction with Mixture Density Networks. In: *Proceedings of International Conference on Artificial Network*, 1998, Skovde, Sweden.

TSAY, R. S. *Analysis of Financial Time Series*. Second Edition, John Wiley & Sons Inc., 2005.

ZIEGELMANN, F. A.; PEREIRA, P. L. V. Modelos de Volatilidade Estocástica com Deformação Temporal: Um Estudo Empírico para o índice Ibovespa. *Pesquisa e Planejamento Econômico*, v.27, n.2, 1997.

Apêndice A – Algoritmo de Retro-propagação

No processo de treinamento da rede neural são determinados os pesos e bias de acordo com o algoritmo de aprendizagem, que estabelece a regra para a alteração dos pesos. Neste apêndice é demonstrado que o princípio do algoritmo de retro-propagação, com aplicação de métodos numéricos de minimização pelo gradiente descendente, aproxima a saída da rede neural dos valores da meta observada em cada iteração.

Os pesos sinápticos e bias ótimos da rede neural, após operação do algoritmo de aprendizagem, minimizam a função perda (Equação A.1) expressa pela soma do quadrado das diferenças entre a meta desejada, Y_d , e a saída da rede neural, Y . Para distintos parâmetros da rede neural observamos diferentes saídas da rede resultando em variados valores para a função perda, dados os insumos. Para arbitrários pesos sinápticos e bias determina-se a superfície de erro.

$$L(w) = \frac{1}{2} \sum_{j=1}^T (Y_{d,j} - Y_j)^2 \quad (\text{A.1})$$

No treinamento da rede neural queremos os pesos sinápticos e termos de bias que minimiza a superfície de erro. O algoritmo de retropropagação se baseia em métodos de otimização numérica pelo gradiente descendente num processo iterativo. Seja os pesos sinápticos e bias representados por W , a partir de um valor inicial, W_0 , segue-se um procedimento iterativo que reduz a função perda a cada iteração, $L(W_{n+1}) < L(W_n)$. O treinamento da rede neural é interrompido quando obtemos os pesos W^* que proporcionem o mínimo da função perda ou um valor de erro tolerável.

O algoritmo de retro-propagação considera uma regra para alteração dos pesos na direção oposta do gradiente da superfície de erro. Para uma rede neural com uma única camada e neurônio, contando com uma única variável meta e vetor de insumos X , o procedimento para determinação dos pesos é a apresentado na Equação A.2, neste caso chamada de regra delta (Equação A.3),

$$\Delta W_n = -\eta \frac{\partial L}{\partial W_n}, \quad \text{isto é,} \quad (\text{A.2})$$

$$W_{n+1} = W_n + \eta L(W_j) X_i, \quad (\text{A.3})$$

onde $L(W_j)$ é o j -ésimo sinal de erro $L(W_j) = Y_{d,j} - Y_j$ e X_i o i -ésimo insumo do vetor X .

Seja uma aproximação linear da função perda com base numa expansão de Taylor de primeira ordem,

$$L(W_{n+1}) = L(W_n) + \frac{\partial L}{\partial W_n} \Delta W. \quad (A.4)$$

Para demonstrarmos que a operação do algoritmo usando um método de otimização baseado no gradiente descendente proporciona uma redução da função perda após cada iteração substituímos a regra de alteração dos pesos (Equação A.2) na expansão de Taylor de primeira ordem da função perda (Equação A.4), tal que,

$$\begin{aligned} L(W_{n+1}) &= L(W_n) + \frac{\partial L}{\partial W_n} \left(-\eta \frac{\partial L}{\partial W_n} \right) \rightarrow L(W_{n+1}) = L(W_n) - \eta \left(\frac{\partial L}{\partial W_n} \right)^2 \rightarrow \\ L(W_{n+1}) - L(W_n) &= -\eta \left(\frac{\partial L}{\partial W_n} \right)^2 \rightarrow L(W_{n+1}) - L(W_n) < 0. \end{aligned}$$

Cujo resultado é válido para η assumindo valores suficientemente pequenos.

O algoritmo de retro-propagação (Equação 1.4) é uma generalização da regra delta (Equação A.3) realizando a busca dos pesos ótimos de acordo com a Equação A.2. Onde temos aproximações do gradiente com aplicações de regra da cadeia e procedimentos iterativos, dada a natureza paralela e seqüencialmente conectada da rede neural. No caso de uma rede neural com múltiplas camadas alimentada para frente com uma única camada escondida com arquitetura NN($m, q, 1$) o algoritmo de retro-propagação resulta na alteração dos pesos do neurônio na camada de saída (Equação A.5) e camada intermediária (Equações A.6), como abaixo:

$$\beta_{j(n+1)} = \beta_{j(n)} + \eta L(w_j) F'(\beta_0 + \sum_{k=1}^q \beta_j y_k) y_k, \quad (A.5)$$

$$w_{k,j(n+1)} = w_{k,j(n)} + \eta L(w_j) F'(\beta_0 + \sum_{k=1}^q \beta_j y_k) \sum_{j=1}^q \beta_j G'(\sum_{j=1}^m w_{k,j} X_j + b_k) w_{k,j(n)} X_j, \quad (A.6)$$

para $L(w_j) = Y_{d,j} - Y_j$ e $y_k = G(\sum_{j=1}^m w_{k,j} * X_j + b_k)$. As demais notações são as mesmas que as usadas na seção 1.1.