

# Um estudo comparativo de Redes Neurais e Modelos GARCH para Previsão da Volatilidade de Séries Temporais Financeiras

André Barbosa Oliveira  
Doutorando em Economia – FGV/SP

Flávio Augusto Ziegelmann  
Instituto de Matemática – IM/UFRGS

## 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. 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. 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.

## 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; FRANCES & VAN DIJK, 2000). 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.

Este trabalho busca estudar a volatilidade com modelos de redes neurais. Após esta introdução, a seção 2 apresenta os modelos de redes neurais e suas características, e a seção 3 apresenta os modelos de volatilidade da família GARCH. A seção 4 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 trabalho termina com as considerações finais na seção 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.

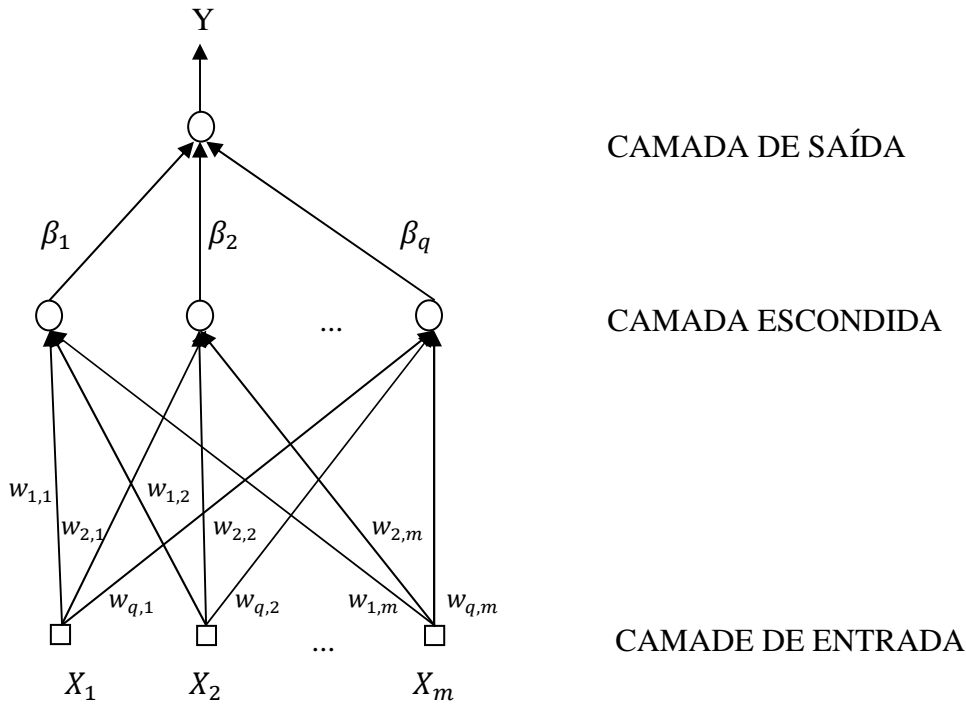
Redes neurais artificiais baseiam-se em um conjunto de dados para aprender uma relação e reproduzi-la. Estas são aplicadas em tarefas de classificação, com a associação de um padrão de entrada a uma classe entre diversas conhecidas; bem como análise de reconhecimento de padrões, como aproximação de funções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

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 Figura 1.

**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,  $\beta_j$ . Ainda adiciona-se o termo de bias,  $\beta_0$ , 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, proporcionando 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.

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

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 regra:

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

onde o índice  $n$  representa a  $n$ -ésima iteração,  $\eta$  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  $\alpha$ .

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).

### 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.

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-diárias<sup>1</sup>.

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 entre

---

<sup>1</sup> 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.

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).

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;  $\theta$ , 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.

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:

$$X_t = \sqrt{h_t} * v_t \tag{3.1}$$

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

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 Autoregressiva, 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  $\alpha_i$  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  $A = \alpha_0 + \beta * \sigma^2$ .<sup>2</sup> 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 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.

#### 4. 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.

<sup>2</sup> 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).

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:

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

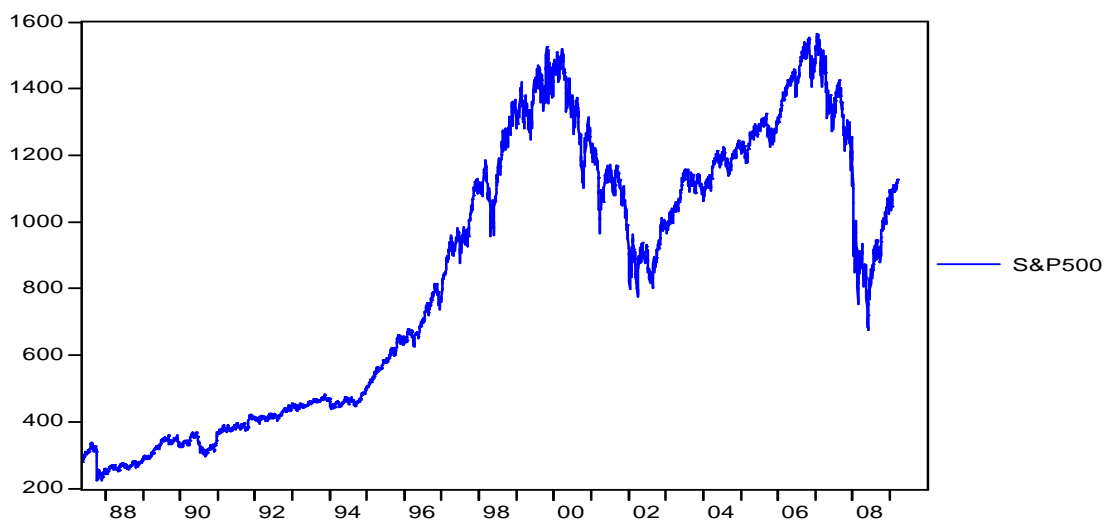
onde  $P_t$  é a cotação no período  $t$ .

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.

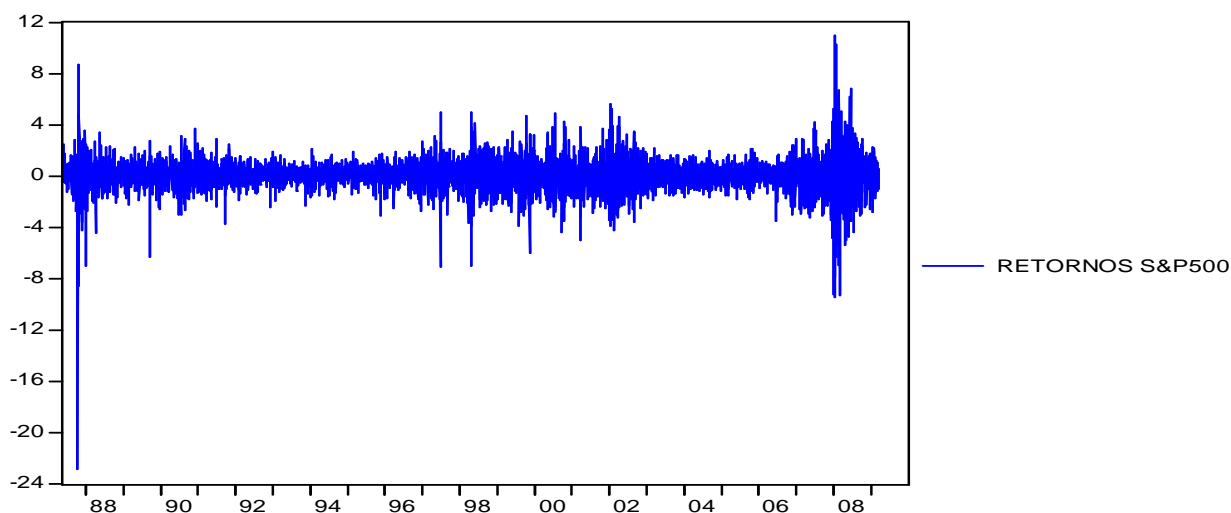
A evolução do índice S&P 500 caracteriza dois momentos no mercado financeiro internacional, um inicial de estável expansão da economia mundial e internacionalização do mercado financeiro crescente, 1987 a 1995; outro de ambiente de crises internacionais e incerteza, 1996 a 2008. Entre 1987 a 1995 o mercado 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 1 – Índice e Retornos: S&P 500.**



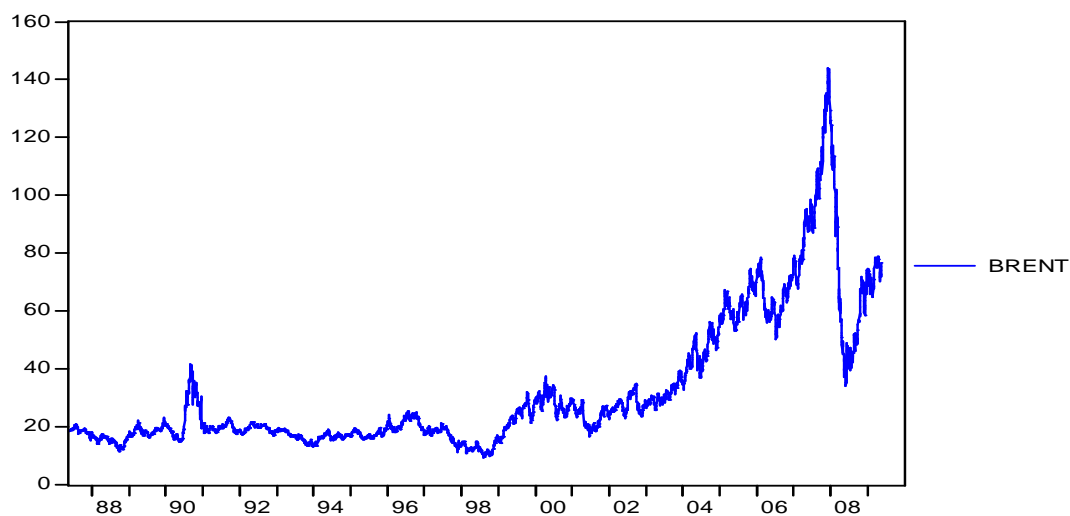
(a)



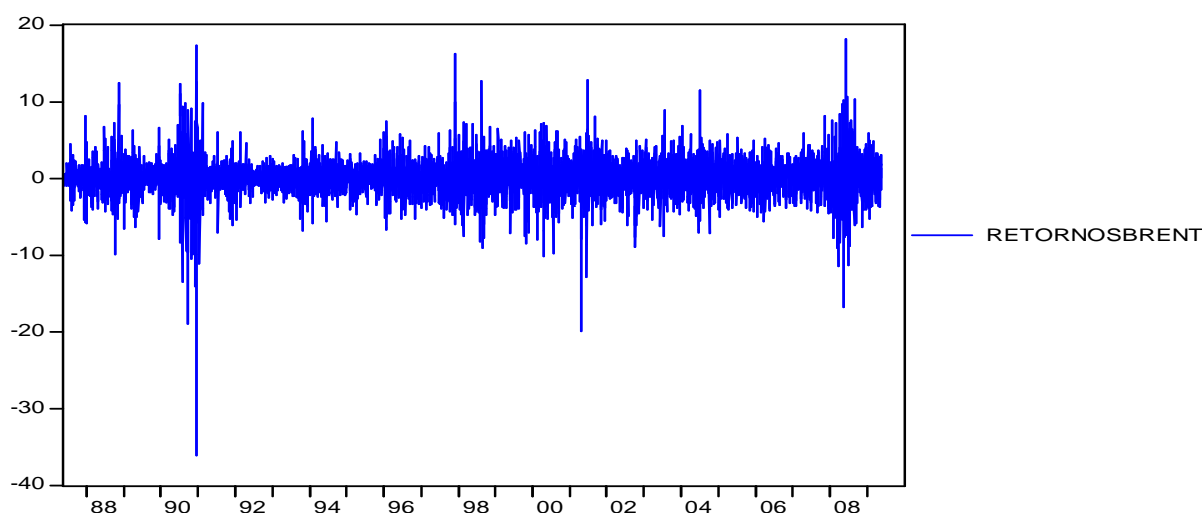
(b)

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 muito 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.

**Gráfico 2 – Índice e Retornos: Cotações Petróleo Brent.**



(a)



(b)

As cotações do petróleo Brent (Gráfico 2) 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 2 - (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.

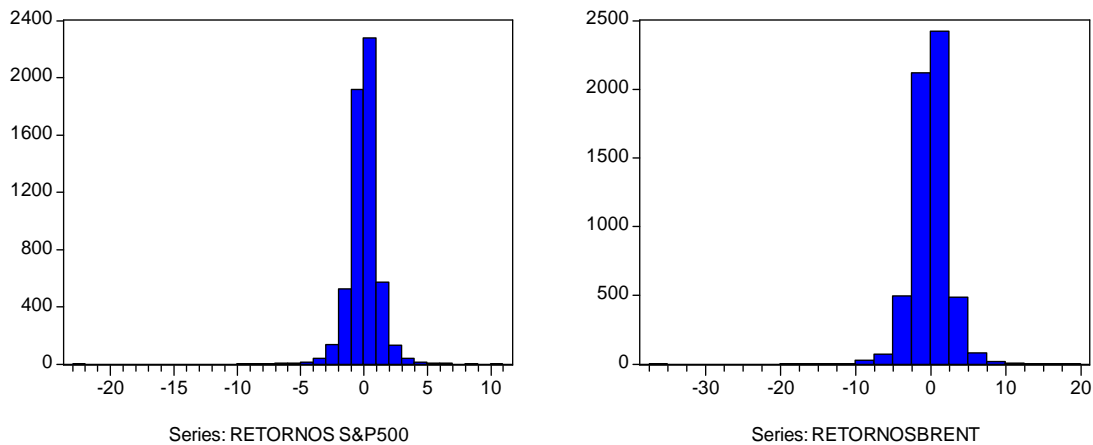
Embora os mercados de ações e de petróleo tenham dinâmica distinta, com principais fatores responsáveis pela variação nas cotações sendo específicos às condições prevalentes nestes mercados, determinados momentos de grande instabilidade na economia são comuns a ambas as séries como a crise Sub-prime. 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 (Tabela1). 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 3) ilustram estas características.

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

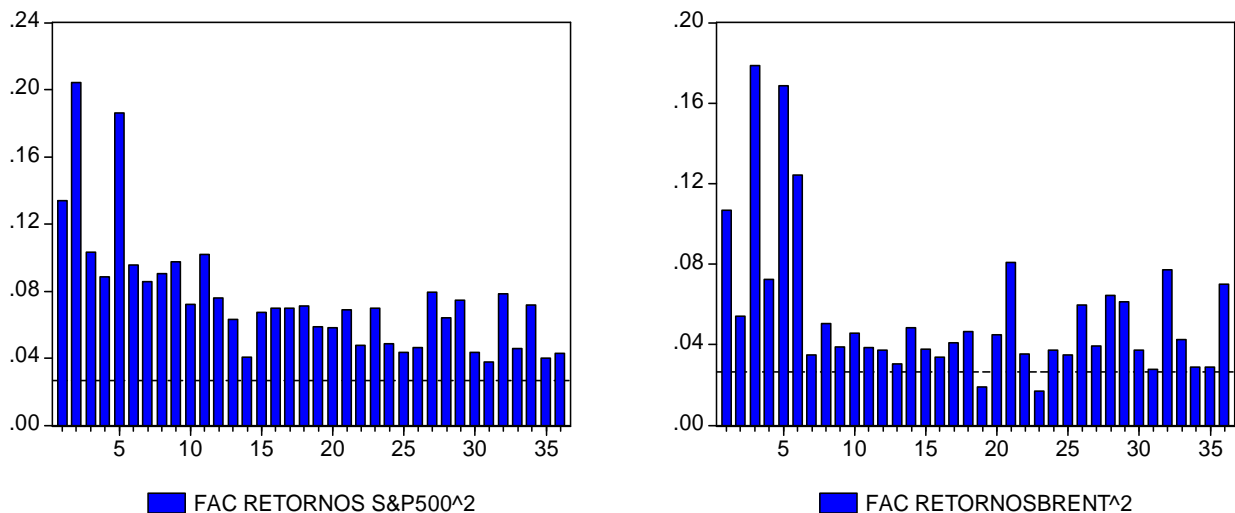
	RETORNOS S&P500	RETORNOSBRENT
Mean	0.024339	0.024608
Median	0.05653	0.016673
Maximum	10.9572	18.12974
Minimum	-22.89972	-36.12144
Std. Dev.	1.215288	2.401766
Skewness	-1.395148	-0.669036
Kurtosis	33.199	17.56154
Jarque-Bera	218597.5	51212
Probability	0	0
Sum	138.8323	141.4476
Sum Sq. Dev.	8422.9	33151.45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

**Gráfico 3 – 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 4). Isto nos habilita a estimar os modelos de volatilidade para estas séries.

**Gráfico 4 – 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.**

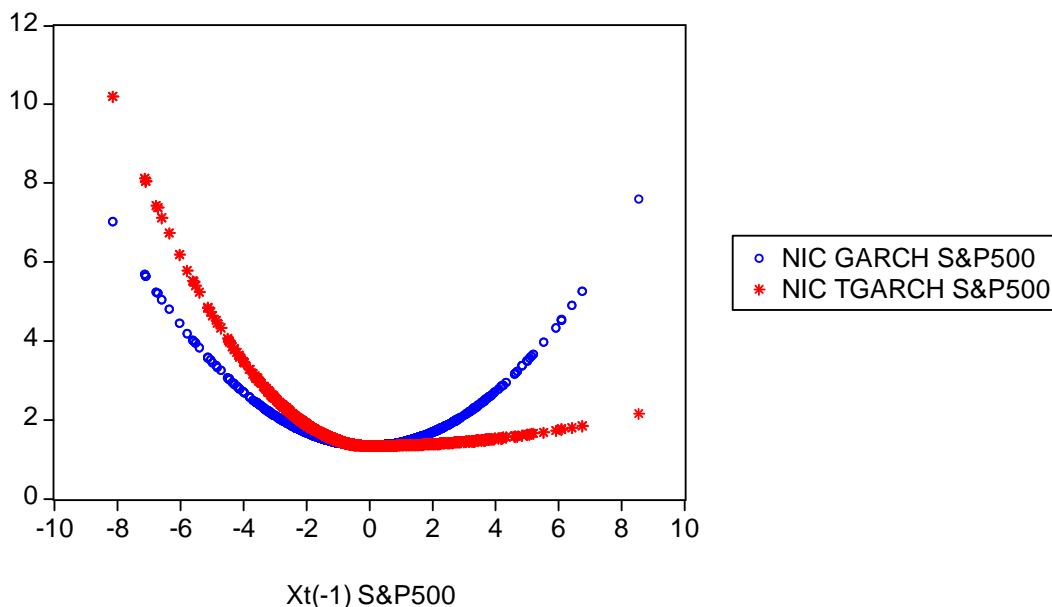
(1) <i>RETORNOS S&amp;P500</i> =	0.055594	-	0.038133	$r_{t-5}$
z-statistic (p-value)	4.9780 (0.0000)		-2.6214	(0.0088)
$h_t =$	0.014002 +	0.0858	$X_{t-1}^2$	+ 0.9046
p-value	(0.0000)	(0.0000)		(0.0000)
<i>Estatísticas</i>				
Q(10) = 12.3460 (0.1950), Q(20) = 24.517 (0.1770) ; Q_2(10) = 4.5172 (0.8740) , 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&amp;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
p-value	(0.0000)	(0.0364)	(0.0000)	$d_{t-1} * X_{t-1}^2$ + 0.9099
				$h_{t-1}$
				(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				
<b>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.</b>				

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 5. 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 5 – 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 6) sugere o efeito simétrico dos retornos sobre a volatilidade.

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

$$(1) \text{RETORNOSBRENT} = 0.047731r_{t-1}$$

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

$$(2) \text{RETORNOSBRENT} = 0.047546r_{t-1}$$

z-statistic (p-value) 3.1997 (0.0014)

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

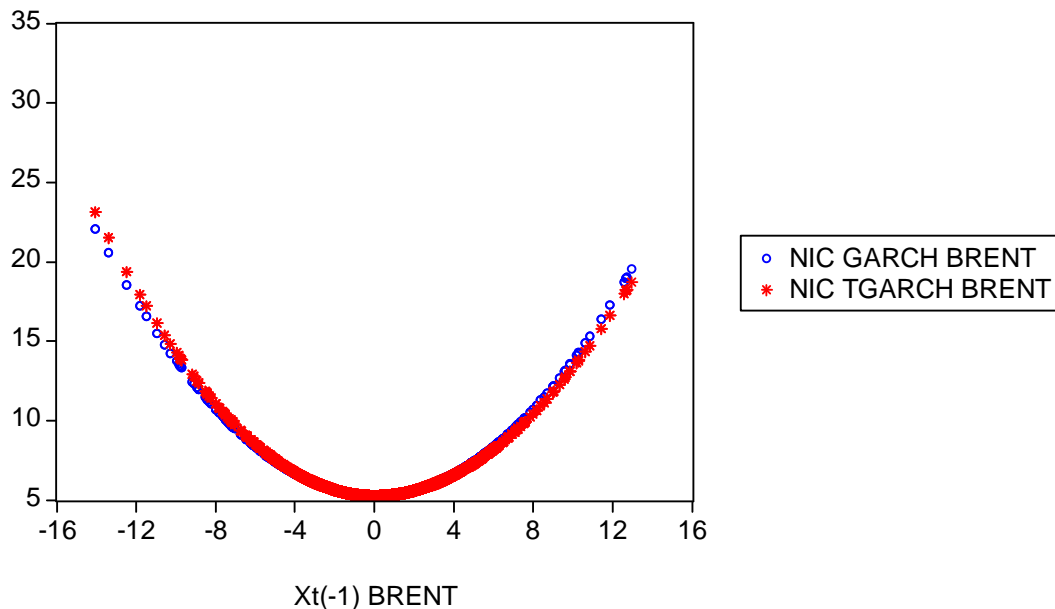
p-value (0.0000) (0.0000) (0.1327) (0.0000)

*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 6 – 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 retornos<sup>3</sup>. A estimação da curva de chegada de informação por redes neurais (Gráfico 7) 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.

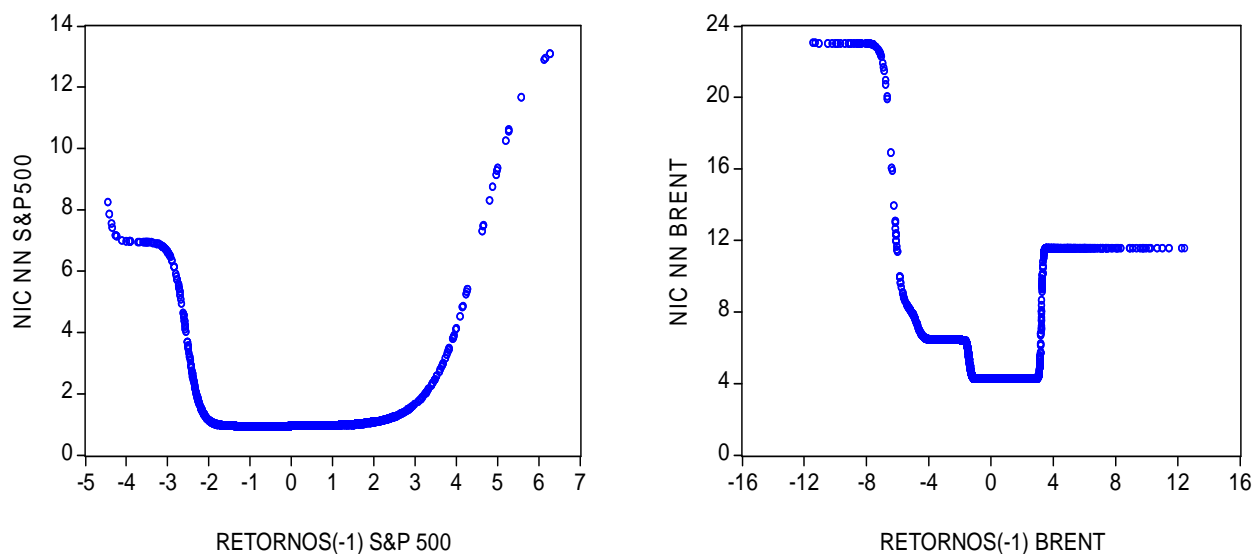
Apesar de RNA possuírem capacidade de aproximar relações complexas entre variáveis, sendo um modelo que se aproxima de métodos não paramétricos, precisa de maior numero de observações para representação adequada de um processo. Assim, as estimativas em regiões com um menor número de observações – os 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.

---

<sup>3</sup> 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 7 – 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

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{Y}_t - Y_t)^2}; \quad (2.19)$$

$$MAPE = \frac{1}{T} \sum_{t=1}^T \left| \frac{\hat{Y}_t - Y_t}{Y_t} \right|. \quad (2.20)$$

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 8), 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.

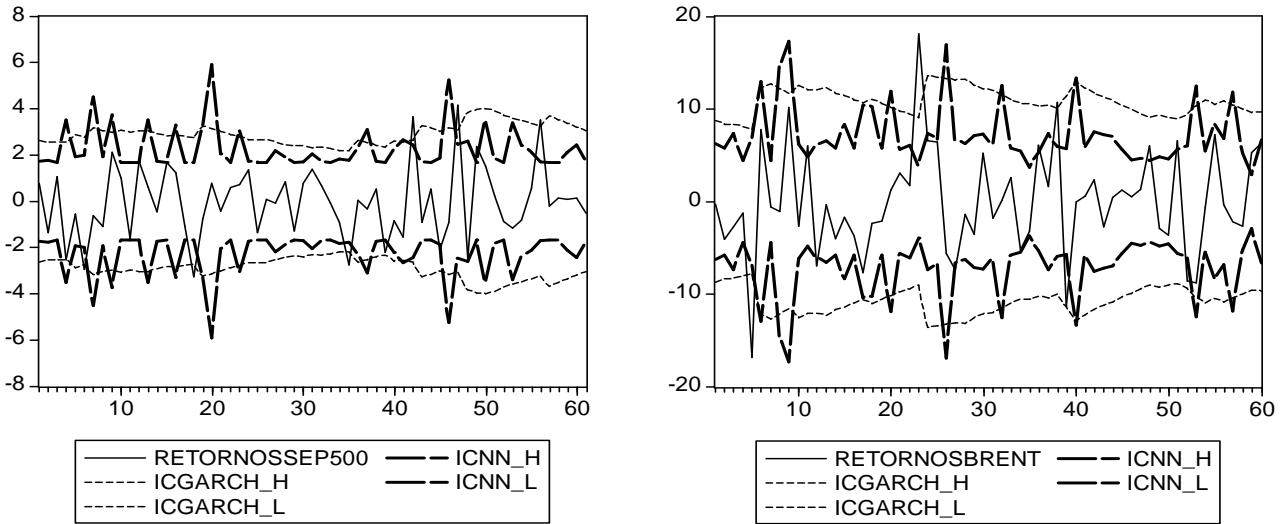
Para a série de retornos Brent (Gráfico 8) os intervalos de confiança são apresentados sobre um período de alta instabilidade no auge da crise Sub-Prime, período 12/2008 a 02/2009, com retornos diários entre 20% a -20%. Os intervalos de confiança GARCH exibem comportamento suavizado e as redes neurais 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.

**Gráfico 8 – Intervalos de Confiança GARCH(1,1) e Redes Neurais: Retornos S&P 500 (Período de 10/01/2008 a 08/04/2008) 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 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; entretantes a volatilidade estimada por redes neurais acompanha de forma mais próxima a trajetória de retornos, e sua mudanças de comportamento, porém com resposta abrupta a alterações dos retornos de seus níveis verificados no passado mais recente.

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 Bibliográficas**

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.

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.

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 Econométricos y Práctica Financiera. *Revista Asturiana de Economía – RAE*, n.31, p.221-252, 2004.

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.

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.

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.