

## **ESTIMAÇÃO DA VOLATILIDADE PERCEBIDA FUTURA PELOS MODELOS ARFIMA E NEAREST NEIGHBOR**

**ALCIDES CARLOS DE ARAUJO**  
FEA-SP  
alcides.carlos@yahoo.com.br

**ALESSANDRA DE AVILA MONTINI**  
USP - Universidade de São Paulo  
amontini@usp.br

## ESTIMAÇÃO DA VOLATILIDADE PERCEBIDA FUTURA PELOS MODELOS ARFIMA E NEAREST NEIGHBOR

**Resumo:** A mensuração da volatilidade recebeu relevância nos últimos anos, uma das medidas mais importantes foi a Volatilidade Percebida (*Realized Volatility - RV*), proposta por Andersen e Bollerslev (1998). De acordo com Danielsson (2011), esta medida utiliza dados passados com frequência intradiária, sendo uma medida puramente direcionada pelos dados. Andrada-Felix et al. (2013) propõe a estimação e combinação dos modelos *Autoregressive Fractionally Integrated Moving Average* (ARFIMA) e *Nearest Neighbor* (NN) para projeção da Volatilidade Percebida. Mediante o desafio de selecionar o melhor método de estimação, deveria o tomador de decisão selecionar a "melhor" projeção ou projetar o futuro por meio da combinação das múltiplas projeções? O objetivo do presente artigo é estimar e projetar a Volatilidade Percebida pelos métodos ARFIMA e *Nearest Neighbor* (NN). A combinação das projeções por pesos iguais e *Mean Squared Forecast Error* (MSFE) foram propostas para analisar a eficiência em combinar os modelos ARFIMA e NN. Na comparação entre as projeções individuais e os métodos de combinação, a combinação pelo método MSFE apresentou melhor performance. Os resultados encontrados apresentaram concordância com a literatura, apresentando que a combinação de projeções demonstra melhor performance em relação as projeções de modelos individuais.

**Palavras-Chave:** Dados em alta frequência, Volatilidade Percebida, Combinação de projeções.

## FORECASTING REALIZED VOLATILITY USING ARFIMA-RV AND NEAREST NEIGHBOR MODELS

**Abstract:** *Measuring volatility has received some relevance over the last years, one of the most important has been the Realized Volatility, proposed by Andersen and Bollerslev (1998). According to Danielsson (2011), it is a data driven measure, this metric needs intraday past data. Andrada-Felix et al. (2013) propose the estimation and the combination of the models Autoregressive Fractionally Integrated Moving Average (ARFIMA) and Nearest Neighbor (NN) for Realized Volatility forecasting. With the challenge to choose the best estimation method, should the decision maker choose the "best" projection or to forecast using the combination of multiple projections? The objective of this article is to estimate and forecast the Realized Volatility using the methods ARFIMA and Nearest Neighbor (NN). The forecast combination methods Equally Weighted combination and Mean Squared Forecast Error (MSFE) were proposed to analyze the combination performance between ARFIMA and NN. In the results "in the sample", the model ARFIMA presented better performance. The Mean Squared Forecast Error combination method demonstrated better performance when the individual forecasts and the combination methods were compared. The results showed agreement with the literature review, combining forecasts presents better performance than forecasts based on individual models.*

**Keywords:** *High Frequency Data, Realized Volatility, Forecast Combination*

## INTRODUÇÃO

Um problema da área financeira é a análise de grandes bancos de dados de negociações na bolsa de valores. Estas negociações são realizadas em tempo real e, para diversos ativos, em altíssima frequência. Diante deste ambiente dinâmico, os dados obtidos para análise são denominados dados em alta frequência (*High Frequency Data*).

Uma das atividades de finanças que fomenta e prioriza a análise de dados em alta frequência é o HFT (*High Frequency Trading*), tipo de negociação associada a rápida realocação de capital, em que as transações podem ocorrer em milésimos de segundos. Conforme Aldridge (2013), os sistemas HFT são construídos utilizando algoritmos complexos para otimizar o envio de ordens, analisar dados obtidos e tomar as melhores decisões de compra e venda.

Conforme Portnoy (2011) o crescimento do HFT ocasionou grandes aumentos nos volumes transacionados. No Brasil, Araújo e Montini (2014) apresentam que no período com a presença do HFT houve um crescimento no número de negociações. Este número de negociações influenciou no aumento do volume transacionado, sendo que os aumentos no volume demonstraram associação a maiores níveis de volatilidade.

A volatilidade possui impacto em diversas aplicações financeiras, conforme apresenta Danielsson (2011). As aplicações vão desde gestão de carteiras, alocação de recursos, apreçamento de derivativos, gestão de riscos até tomadas de decisão das instituições a respeito da estabilidade financeira.

Em essência, a volatilidade é uma variável latente; diante disto, diversos métodos foram propostos na literatura para mensurá-la. Uma das formas de mensuração que recebeu relevância nos últimos anos foi a medida de Volatilidade Percebida (*Realized Volatility - RV*), proposta por Andersen e Bollerslev (1998).

Alguns trabalhos no Brasil como os de Pontes (2014) e Val et al. (2014) traduziram o termo *Realized Volatility* como Volatilidade Realizada; no presente artigo, os termos Volatilidade Percebida e Volatilidade Realizada serão sinônimas ao correspondente termo em inglês *Realized Volatility*.

De acordo com Danielsson (2011), esta medida utiliza dados passados com frequência intradiária (podendo ser de 5 minutos ou em alta frequência); deste modo, torna-se uma medida puramente direcionada pelos dados, ou seja, pelo que de fato ocorreu no mercado.

Quanto a projeção da volatilidade, Zivot (2005) e Andersen et al. (2007) apresentam uma melhor performance para os modelos que consideram medidas baseadas em dados de alta frequência. Andrada-Felix et al. (2013) propõe um artigo inovativo ao comparar os modelos *Autoregressive Fractionally Integrated Moving Average* (ARFIMA) e *Nearest Neighbor* (NN), além de propor a combinação de ambos para projeção da Volatilidade Percebida.

Combinar as projeções dos modelos ARFIMA e *Nearest Neighbor* (NN) possui justificativa interessante, dado que o primeiro modelo é baseado em pressupostos e suposições da teoria econômica, enquanto que o segundo está baseado em modelos determinísticos não lineares advindos da física mecânica. Diante disto, combinar o conhecimento de duas áreas importantes da ciência em finanças pode melhorar o desempenho das projeções.

Estes dois modelos são exemplos da amplitude de métodos de projeção desenvolvidos para dados em alta frequência, porém por meio desta amplitude surge um novo problema. Torna-se complicado para um gestor de algoritmos de negociação selecionar uma medida de risco para controlar as perdas de suas operações ou mensurar o quanto poderá perder caso ocorra o pior cenário.

Diante disto, como objetivo geral o artigo propõe a combinação das projeções de volatilidade percebida para a melhoria da capacidade de gestão das operações. Conforme Timmermann (2006) a combinação de projeções demonstrou sucesso empírico em diversas aplicações econômicas, como projeção do PIB, preço do petróleo e retorno de ações. O objetivo específico do presente artigo é combinar as volatilidades estimadas pelos modelos ARFIMA e *Nearest Neighbor* - NN.

Os estudos a respeito de combinação de projeções possuem significativa aplicação prática, um exemplo é apresentado por Genre et al. (2013). Conforme os autores, desde 1999 o Banco Central Europeu apresenta uma pesquisa a respeito das projeções realizadas pelos profissionais de mercado (*Survey of Professional Forecasters* - SPF).

Nesta pesquisa, as projeções sobre inflação, PIB e desemprego são apresentadas pela média simples das projeções de cada profissional. Diante disto, Genre et al. (2013) discutem sobre apresentar as projeções dos profissionais utilizando outras medidas de combinação para enriquecer o conteúdo do relatório.

## 2 REVISÃO DA LITERATURA

### 2.1 Medidas de volatilidade com dados em alta frequência

Um tema de pesquisa relativamente recente, conforme Zivot (2005) é o estudo das medidas de volatilidade para dados em alta frequência. Esta área de estudos possui importância dado que produz estimadores que apresentam melhor exatidão para a volatilidade e correlação condicional quando comparados com estimadores que utilizam dados diários ou mensais.

Os estimadores de volatilidade para dados em alta frequência são derivados a partir do conceito da Variância Integrada (*Integrated Variance* – IV), conforme descrevem Andersen et al. (2001), Barndorff-Nielsen e Shephard (2004) e Bauwens (2012). Conforme os autores, dado a velocidade e liquidez nos mercados financeiros, as operações tendem a ocorrer de forma muito próxima do tempo contínuo.

Seja  $Y_t$  um vetor de preços  $N \times 1$ , em que  $Y_t = \ln p_t$ , sendo  $p_t$  o preço negociado, a difusão dos preços é expresso na equação 1,

$$dY_t = \mu_t dt + \sigma_t dW_t, t \geq 0, \quad (1)$$

em que,  $dY_t$  é o incremento do preço em logaritmo,  $\mu_t$  é a direção (*drift*) do processo contínuo localmente delimitado,  $\sigma_t$  é a volatilidade instantânea estritamente positiva e estacionária e  $W_t$  é o Movimento Browniano Padrão.

Pode-se assumir que  $t$  é um dia de operações, sendo  $i - s$  o espaçamento entre duas observações, com  $s = 1$  e  $i = 1 \dots M$  períodos intradiários, o retorno  $r_{t,i}$  é expresso na equação 2.

$$r_{t,i} \equiv Y_{t,i} - Y_{t,i-1} = \int_{t,i-1}^{t,M} \mu_{t,s} ds + \int_{t,i-1}^{t,M} \sigma_{t,s} dW_{t,s}, i = 1 \dots M. \quad (2)$$

Por meio da expressão 2, a distribuição de probabilidade dos retornos é definida na expressão 3,

$$r_{t,i} \sim N \left( \int_{t,i-1}^{t,M} \mu_{t,s} ds, IV_{t,M} \right), \quad (3)$$

em que,  $IV_{t,s}$  é dada pela expressão 4,

$$IV_{t,s} \equiv \int_{t,i-1}^{t,M} \sigma_{t,s}^2 ds, \quad (4)$$

em que, IV é a Variância Integrada (*Integrated Variance – IV*).

Como se observa, os valores da variável IV precisam ser estimados, uma vez que é uma variável latente. No caso, IV pode ser estimada utilizando dados intradiários, Andersen e Bollerslev (1998) propõe o estimador da Variância Percebida (*Realized Variance - RV*), expressão 5,

$$RV_t \equiv \sum_{i=1}^M r_{t,i} r'_{t,i}. \quad (5)$$

De acordo com Andersen *et al.* (2003), o estimador foi apresentado com o objetivo de integrar dados de negociação em alta frequência e medidas de estimação de volatilidades. Conforme os autores, o estimador é fácil de ser implementado e melhora o desempenho dos modelos preditivos.

## 2.2 Projeção com ARFIMA e NN

Modelos de projeção da volatilidade utilizando dados em alta frequência são amplamente aceitos na academia como melhores preditores da variância futura, conforme apresenta Boudt et al. (2013). Neste tópico são apresentados dois modelos: *Autoregressive Fractionally Integrated Moving Average* (ARFIMA) e *Nearest Neighbor* (NN).

O modelo ARFIMA foi proposto nos artigos de Granger e Joyeux (1980) e Hosking (1981) sendo apresentadas aplicações para hidrologia e economia. Conforme Hansen e Lunde (2012) a idéia para ajustar séries de volatilidades percebidas foi proposta em 1999 e publicada no artigo de Andersen et al. (2003). Após a publicação, as projeções por ARFIMA são amplamente utilizadas para ajustes de séries de volatilidades percebidas, conforme citam Andrada-Felix et al. (2013).

O modelo ARFIMA( $p, d, q$ ) é uma generalização dos modelos ARIMA( $p, d, q$ ), sendo  $p$  e  $q$  as ordens dos componentes de autoregressão e médias móveis respectivamente e  $d$  a ordem de integração fracionada entre 0 e 1. Diante disto, seja  $y_t$  uma série temporal,  $\mu$  uma constante e  $\varepsilon_t$  um vetor de ruídos no tempo com média zero e variância  $\sigma^2$ ; um modelo ARFIMA( $p, d, q$ ) é exibido na expressão 6,

$$\Phi(L)(1 - L)^d(y_t - \mu) = \Theta(L)\varepsilon_t, \quad (6)$$

em que,  $\Phi$  representa o conjunto de parâmetros associados aos componentes autoregressivos,  $\Theta$  representa o conjunto de parâmetros associados aos componentes de médias móveis,  $L$  é o operador defasagem e  $d$  é o operador de diferenças fracionárias.

Nos trabalhos de Hansen e Lunde (2012) e Andrada-Felix et al. (2013), os autores adaptaram  $y_t$  como uma medida de volatilidade percebida (RV). Diante disto, o modelo ARFIMA para estimar a volatilidade para dados em alta frequência é expresso em 7,

$$\Phi(L)(1 - L)^d(RV_t - \mu) = \Theta(L)\varepsilon_t. \quad (7)$$

A técnica denominada *Nearest Neighbor* (NN) teve origem nas áreas de física e engenharia, sendo bastante aplicada em análise de reconhecimento de padrões. Os trabalhos de Farmer e Sidorowich (1987) e Yakowitz (1987) foram os primeiros a propor a utilização da técnica NN para projeção de séries temporais.

Conforme Andrada-Felix et al. (2013), a técnica é baseada em modelos determinísticos não lineares, situada no campo de estudos das séries que apresentam movimentos caóticos. As principais vantagens da técnica são não necessitar assumir alguma função particular para os processos de média, variância condicional e inovações (choques ou ruídos) e a flexibilidade.

O modelo possui função objetivo com minimização local, a ferramenta é adaptável mesmo na existência de quebras estruturais, mudanças de regime e não estacionariedade da série. Diante disto, o modelo considera as variações do tempo nas estimações.

Para realizar as projeções deve-se, primeiramente, subamostrar a série de volatilidade percebida numa base  $x_t^m$ , em que  $m$  é o número de padrões contidos em  $x_t$ , como é exibido na expressão 8,

$$\mathbf{x}_t^m = (x_t, x_{t-1}, \dots, x_{t-(m-1)}) \in \mathbb{R}^m, m \leq t \leq T - 1, \quad (8)$$

em que,  $T$  é o tamanho da série.

Por meio da base de padrões formada, busca encontrar os  $k$  vizinhos mais próximos pela minimização da função objetivo 9, neste caso, os  $k$  valores mínimos da função de distância Euclidiana 10,

$$\|x_{t_i}^m - x_T^m\|, t_i = m, m + 1, \dots, T - 1, \quad (9)$$

$$\sqrt{\sum_{i=1}^T (x_t^m - x_{t+1}^m)^2}, \quad (10)$$

em que,  $x_T^m$  representa o grupo de padrões mais recentes da série.

A projeção  $\hat{x}_{T+1}$  é obtida pela combinação das observações correspondentes a cada  $k$  sequência vizinha e ajustada por uma função  $F(\cdot)$ , conforme é exibido na expressão 11,

$$\hat{x}_{T+1} = F(x_{t_1+1}, x_{t_2+1}, \dots, x_{t_k+1}), \quad (11)$$

em que,  $F(\cdot)$  pode ser uma função da média, mediana, kernel ou autoregressão.

No artigo de Hansen e Lunde (2012) os autores apresentam resultados de diversas outras pesquisas, destacaram-se os artigos de Oomen (2001) e Andersen et al. (2003). Em ambos os trabalhos, os autores compararam a performance nas projeções da Volatilidade Percebida entre os modelos ARFIMA e GARCH, os resultados demonstraram melhor desempenho para o ARFIMA.

No trabalho de Andrada-Felix et al. (2013) os autores comparam os modelos ARFIMA e *Nearest Neighbor* (NN) e de propõem a combinação de ambos para projeção da Volatilidade Percebida. Quanto as projeções, os autores demonstraram que a pior performance foi do modelo ARFIMA, porém as diferenças não foram significantes.

## 2.2 Combinação de projeções

No presente tópico é apresentada a abordagem da combinação de projeções. Neste caso, as projeções de modelos bem especificados serão combinadas buscando melhor assertividade.

Esta abordagem foi inspirada no problema dos tomadores de decisão possuírem múltiplas projeções para a mesma variável. De acordo com Timmermann (2006), mediante diversas projeções em que cada uma tende a demonstrar diferenças quanto a informações inseridas ou método de estimação; deveria o tomador de decisão selecionar a "melhor" projeção ou projetar o futuro por meio da combinação das múltiplas projeções?

Diante disto, Timmermann (2006) apresenta que selecionar uma única projeção possui a vantagem de que o modelo utilizado será aquele que demonstrou menor erro de previsão. Contudo, combinar as projeções possibilita a utilização do "melhor" que cada modelo pode oferecer, minimizando os erros de previsão.

A combinação pela média simples (pesos iguais) dos valores previstos pelas previsões individuais, segundo Makridakis e Winkler (1983), Makridakis et al. (1982), Clemen (1989), Chan, Stock e Watson (1999), Timmerman (2006), Smith e Wallis (2009), Genre et al. (2013) e Graefe et al. (2014), apresenta resultados difíceis de serem superados. Portanto, frequentemente é utilizado como *benchmark*.

Por este motivo, utilizou-se neste trabalho a combinação a ser realizada com pesos iguais. Na expressão 12 é apresentada a ponderação por pesos iguais,

$$\omega_{i,t}^{ig} = \frac{1}{n}, \quad (12)$$

em que,  $n$  é o número de modelos a serem combinados.

Outro método de combinação muito comentado na literatura é a utilização da Média do Erro de Previsão (*Mean Squared Forecast Error - MSFE*). Conforme é apresentado em Stock e

Watson (2001) e Stock e Watson (2004) o método associa maiores ponderações as previsões individuais que demonstrem menor MSFE na amostra utilizada para estimação do valor de  $m$ . O método pela MSFE é apresentado nas expressões 11 e 12,

$$\omega_{i,t}^{msfe} = \frac{m_{i,t}^{-1}}{\sum_{i=1}^n m_{i,t}^{-1}}, \quad (13)$$

$$m_{i,t} = \sum_{i=T_0}^t (y_{s+h}^h - \hat{y}_{i,s+h}^h)^2. \quad (14)$$

Neste método, conforme apresenta Stock e Watson (2004) é necessário estimar o valor de  $m$  (expressão 14) por meio de uma amostra de projeções, para depois calcular os pesos (expressão 13).

Como trabalhos empíricos, pode-se apresentar os artigos de Cavaleri (2008), Huang e Lee (2013) e Baumeister et al. (2014). Em relação a combinação de modelos de projeção de volatilidade, Cavaleri (2008) apresentou que as combinações demonstraram melhores resultados em relação as projeções individuais.

No caso dos artigos de Huang e Lee (2013) e Baumeister et al. (2014), os autores realizaram a estimação de seus modelos utilizando dados em alta frequência. Após as estimações, testaram a eficiência da combinação das projeções, os resultados demonstraram que a combinação melhorou as projeções. No Quadro 1 são apresentados diversos artigos que utilizaram a combinação de projeções.

**Quadro 1 - Resumo utilização de combinação de projeções**

<b>Autor</b>	<b>Dados</b>	<b>Combinação</b>	<b>Resultado</b>
Makridakis e Winkler (1983)	Séries macro e micro econômicas	Pesos iguais e MSFE	Pesos iguais melhor performance
Chan, Stock e Watson (1999)	Séries macro e micro econômicas do mercado americano	Pesos iguais, MSFE, Regressão, PCA, Mediana, PLS	Combinação melhor performance, pesos iguais melhor combinação
Stock e Watson (2004)	73 projeções de dados econômicos de 7 países da OECD	MSFE, <i>shrinkage</i> , pesos iguais, PCA, Filtro de Kalman	Combinação pesos iguais melhor performance
Cavaleri (2008)	Ibovespa, Dow Jones e IGP-M	Pesos iguais, <i>shrinkage</i> e Filtro de Kalman	Combinação por <i>shrinkage</i> melhor performance
Smith e Wallis (2009)	GDP e Índice de Produção Industrial	Pesos Iguais e MSFE	Resultados similares entre os métodos
Genre et al. (2013)	Inflação, GDP e Taxa de desemprego na Europa	Pesos iguais, PCA, Regressão, <i>shrinkage</i> , MSFE	Resultados similares entre os métodos
Huang e Lee (2013)	Série em alta frequência do índice S&P500	Pesos iguais, mediana das projeções e PCA	Combinação melhor performance
Andrada-Felix et al. (2013)	Índice S&P 100	Pesos iguais	Resultados similares com projeções individuais
Baumeister et al. (2014)	Série Petróleo	Pesos Iguais	Combinação melhor performance

**Fonte:** elaborado pelos autores.

### 3 METODOLOGIA

#### 3.1 Dados e métodos

O objetivo do presente artigo é propor a combinação das projeções de Volatilidade Percebida obtidas pelos modelos ARFIMA e *Nearest Neighbor* (NN). Para realizar as projeções, combinações e testes foi utilizada a série de variância percebida do índice Bovespa no período entre 03/01/2000 a 01/10/2014, total de 2714 observações.

A pesquisa possuiu 4 etapas: obtenção da série de volatilidade percebida (RV), estimação dos modelos ARFIMA (expressão 7) e NN (expressões 8 a 11), combinação das projeções (expressões 12 a 14) e avaliação de performance.

Para obtenção da série de volatilidade percebida foram utilizados os retornos continuamente compostos em intervalos de cinco minutos. No artigo de Wink Júnior e Pereira (2011) os autores testaram a utilização das frequências de 1, 5 e 15 minutos; os resultados demonstraram que a frequência de 5 minutos apresentou o menor intervalo de confiança.

Os retornos obtidos nos períodos de abertura e fechamento foram retirados; estes períodos são marcados por forte volatilidade que pode causar viés em toda a análise, conforme cita Falkenberry (2001). Este procedimento foi proposto em Rydberg e Shephard (2003), conforme os autores, o objetivo é retirar os efeitos dos leilões de abertura e fechamento das séries.

Na segunda etapa foram estimados os modelos ARFIMA (expressão 7) e *Nearest Neighbor* – NN (expressões 8 a 11), foi utilizada 60% da amostra total, sendo ao todo 1628 observações, compreendendo o período entre 03/01/2000 a 08/07/2008.

Ainda na etapa 2 foi possível realizar a comparação dos modelos dentro da amostra. As medidas utilizadas para avaliação foram Erro Médio Quadrático (*Mean Squared Error* - MSE) e Raiz do Erro Médio Quadrático (*Root Mean Squared Error* - RMSE), expressões 15 e 16 respectivamente.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2, \quad (15)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2}. \quad (16)$$

Estas medidas de performance foram escolhidas, primeiramente, seguindo os trabalhos de Patton (2006) e Forsberg e Ghysels (2007) em que os autores apresentam que a medida MSE é robusta independente da medida de volatilidade utilizada para projeção. Como segunda razão, em trabalhos realizados no Brasil, autores também utilizaram a medida, como em Wink Júnior e Pereira (2011) e Val et al. (2014).

Na terceira etapa foram obtidas as projeções dos modelos ARFIMA e NN, as previsões da volatilidade foram realizadas seguindo a sequência de projeções um passo a frente. Deste

modo, utilizaram-se os valores atuais para representar a variável dependente defasada, conforme procedimento seguido por Val et al. (2014). Estes valores foram obtidos da amostra restante, sendo ao todo 1086 observações; compreendendo o período de 10/07/2008 a 01/10/2014.

Deste modo, para o restante das análises foram utilizadas as 1086 observações de Volatilidades Percebidas do índice Bovespa e as 1086 observações de projeções obtidas dos modelos ARFIMA e NN.

Para realizar as combinações, utilizou-se o procedimento apresentado em Stock e Watson (2004), em que a amostra é dividida em 3 períodos. O primeiro período foi utilizado para realizar as projeções (60% da amostra total - 1628 observações), o segundo e terceiro período são utilizados para estimação dos pesos e avaliação de performance respectivamente.

Para o caso da combinação pelo método dos pesos iguais, o segundo período seria desnecessário; contudo para o caso do método MSFE, é necessário estimar o valor de  $m$ . Deste modo, o segundo período foi utilizado para estimar o valor de  $m$  e, conseqüentemente, obter os pesos pelo método MSFE.

Diante disto, das 1086 observações de valores reais e projetados, 594 observações (10/07/2008 a 09/11/2011) foram utilizadas para determinar os pesos e 492 observações (10/11/2011 a 01/10/2014) foram utilizadas para avaliar a performance das projeções. Estes períodos foram escolhidos para manter uma proporção maior para a amostra de estimação em relação a amostra de avaliação, no caso, valores de aproximadamente 55% para definir os pesos e 45% para avaliar a performance das projeções.

Na quarta etapa foi realizada a avaliação de performance, as medidas de comparação foram o MSE (expressão 15) e RMSE (expressão 16). As análises foram viabilizadas pelo software *R Package*.

## **4 ANÁLISE DOS DADOS**

### **4.1 Estimação e projeção dos modelos**

Os resultados dos modelos ARFIMA estimados são apresentados na Tabela 1. Os modelos 1 e 2 foram os selecionados para a etapa de projeções.

O modelo 1 foi obtido pelo método da busca do melhor modelo pelo critério AIC (*Akaike Information Criterion*), conforme descrito na metodologia proposta em Hyndman e Khandakar (2008). Após obtenção do modelo com melhor ajuste, buscou-se realizar possíveis ajustes para corrigir possível autocorrelação dos resíduos.

O modelo 1 final é um ARFIMA(18,d,4) incompleto. Conforme pode ser observado na Tabela 1, os coeficientes significantes para  $p$  e  $q$  foram ar4, ar18, ma1 e ma4. O coeficiente de integração fracionária foi significativo com valor de 0,2878. Os gráficos de autocorrelação (ACF) e autocorrelação parcial (PACF) dos resíduos são apresentados na Figura 1(a).

O modelo 2 foi obtido buscando-se obter resíduos não autocorrelacionados, conforme a clássica metodologia de identificação proposta por Box e Jenkins (1970). Primeiramente,

observaram-se os correlogramas da série RV, após a identificação dos possíveis  $p$  e  $q$ , estimaram-se os modelos e verificaram-se os correlogramas dos resíduos.

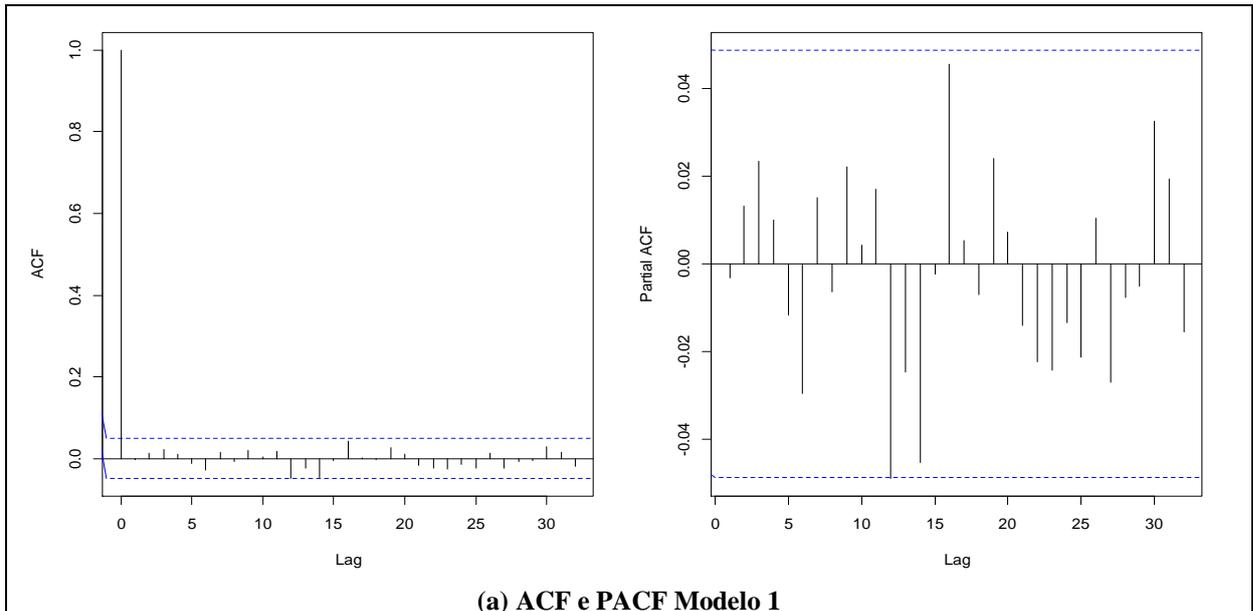
O modelo 2 final é um ARFIMA(18,d,9) incompleto. Por meio da Tabela 1, observam-se os coeficientes significantes: ar1, ar5, ar12, ar18, ma9. O coeficiente de integração fracionária foi significativo com valor de 0,2582. Os gráficos de autocorrelação (ACF) e autocorrelação parcial (PACF) dos resíduos são apresentados na Figura 1(b).

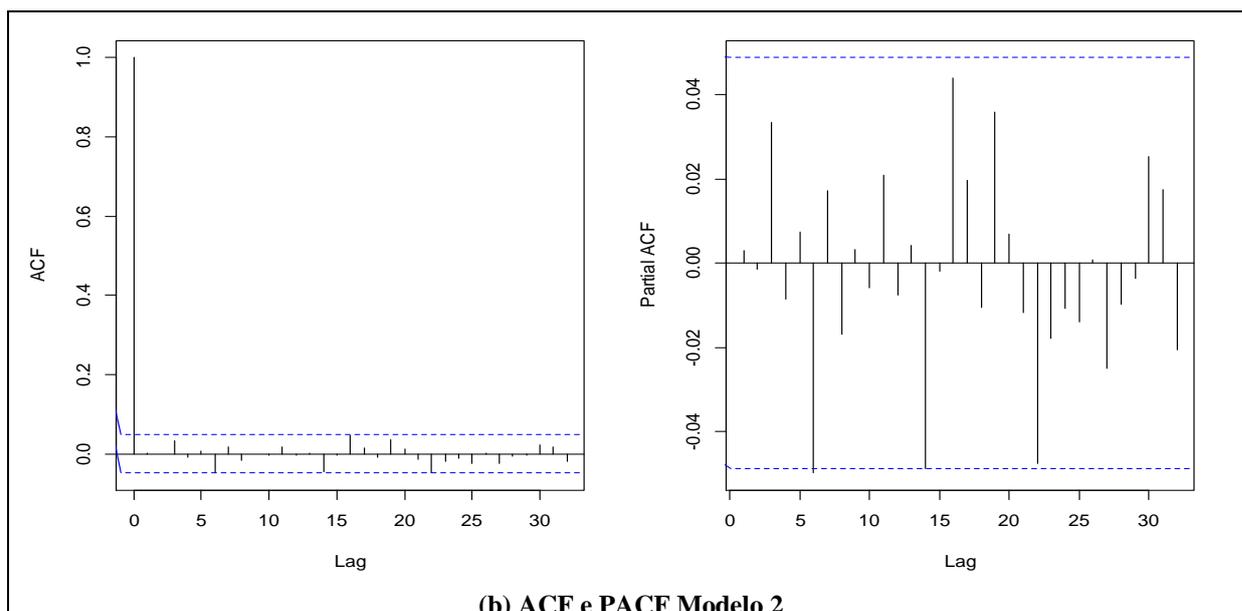
**Tabela 1 - Resultados estimações modelos ARFIMA**

Parâmetro	Modelo 1	Modelo 2
$\mu$	0,0002***	0,0002***
$\phi_1$ (ar1)	-	0,1834***
$\phi_4$ (ar4)	0,4546***	-
$\phi_5$ (ar5)	-	0,0651**
$\phi_{12}$ (ar12)	-	-0,0474**
$\phi_{18}$ (ar18)	-0,0537**	-0,0452**
$\theta_1$ (ma1)	0,1632***	-
$\theta_4$ (ma4)	-0,4841***	-
$\theta_9$ (ma9)	-	0,0596**
$d$	0,2878***	0,2582**
<b>RMSE dentro da amostra</b>	<b>0,0166998</b>	<b>0,0167032</b>

\*\*Significante a 5%. \*\*\*Significante a 1%.

Fonte: dados da pesquisa.





(b) ACF e PACF Modelo 2  
**Figura 1 - Autocorrelações dos resíduos.**

Conforme metodologia para projeção por *Nearest Neighbor* (NN) deve-se, primeiramente, obter os valores de  $m$ , o número de padrões contidos na série, e  $k$ , os vizinhos mais próximos. Para obter estes parâmetros, definiu-se a função de Distância Euclidiana como medida de similaridade e as medidas de média e mediana para ajustar a função  $F(\cdot)$ .

Após definir a função de distância e as funções de ajuste, buscou-se seguir a metodologia utilizada em Andrada-Felix et al. (2013), em que, os valores de  $m$  e  $k$  para projeção são definidos pela minimização dos erros quadrados (MSE) na amostra de estimação.

Na Tabela 2 são apresentados os cinco menores valores de MSE para a respectiva combinação de  $(k, m)$  para as funções de ajuste pela média e mediana. Como pode ser observado, os menores valores de MSE são apresentados pela média, nas combinações (20, 9) e (16, 18). Diante disto, estas duas combinações de  $(k, m)$  e a função de ajuste pela média foram utilizadas para realizar as projeções.

**Tabela 2 - Resultados das estimações modelos NN**

Posição	$(k, m)$	Média	$(k, m)$	Mediana
1	(20, 9)	0,0198985	(10, 12)	0,0204868
2	(16, 18)	0,0199577	(10, 11)	0,0205189
3	(10, 15)	0,0199633	(5, 17)	0,0205860
4	(13, 12)	0,0199756	(10, 14)	0,0205895
5	(20, 13)	0,0199771	(10, 13)	0,0206041

Fonte: dados da pesquisa.

#### 4.2 Combinação e Comparação de performance

Após as estimações dos modelos, foram realizadas as projeções e as combinações das projeções utilizando os métodos de pesos iguais e MSFE.

Conforme descrito na metodologia, primeiramente, buscou-se comparar a performance dos modelos ARFIMA e NN dentro da amostra. Nesta etapa foram utilizadas as 1628 observações que serviram de base para estimação dos parâmetros dos modelos, os valores dos RMSE (*in*

*the sample*) para os modelos ARFIMA 1 e 2 são apresentados na Tabela 1 e os valores RMSE (*in the sample*) para os modelos NN podem ser são apresentados na Tabela 2.

Os modelos ARFIMA 1 e 2 demonstraram melhor desempenho dentro da amostra de estimação com valores de RMSE iguais a 0,0166 e 0,0167 respectivamente.

Após as comparações dentro da amostra, buscou-se comparar os modelos na amostra de avaliação, além de analisar a possibilidade de combinar as projeções pelos métodos de pesos iguais e MSFE. Nesta etapa, foram utilizadas as 492 (10/11/2011 a 01/10/2014) observações separadas para a amostra de avaliação.

Em relação aos métodos de combinação, os pesos para o método MSFE foram estimados utilizando a amostra de 594 observações, período de 10/07/2008 a 09/11/2011. Como resultados, foram alocados pesos de 35,27% para o modelo ARFIMA 1, 35,68% para o modelo ARFIMA 2, 14,23% para o modelo NN 1 e 14,82% para o modelo NN 2. Quanto aos pesos iguais, foram estabelecidos pesos de 25% para cada modelo.

Na Tabela 3 são apresentados os resultados de MSE e RMSE para cada modelo e método de combinação na amostra de avaliação. Individualmente, as projeções realizadas pelos modelos ARFIMA demonstram melhor desempenho na amostra de avaliação. A combinação por pesos iguais demonstrou performance similar ao modelo ARFIMA 2. A combinação pelo método MSFE demonstrou o melhor desempenho entre todos os métodos.

Os resultados encontrados apresentaram concordância com a literatura. Diante disto, a combinação de projeções demonstra melhor performance em relação as projeções de modelos individuais, conforme é citado nos artigos de Newbold e Granger (1974), Winkler e Makridakis (1983) e Baumeister et al. (2014).

Na Figura 2 é apresentada a série real e a série projetada pela combinação MSFE no período entre 10/11/2011 a 01/10/2014.

**Tabela 3 - Resultados projeções**

<b>Projeções</b>	<b>MSE</b>	<b>RMSE</b>
ARFIMA 1	0,000119528	0,01093289
ARFIMA 2	0,000117756	0,01085153
NN 1	0,000161340	0,01270196
NN 2	0,000138941	0,01178733
Combinação Pesos Iguais	0,000117829	0,01085489
<b>Combinação MSFE</b>	<b>0,000115328</b>	<b>0,01073908</b>

Fonte: dados da pesquisa.

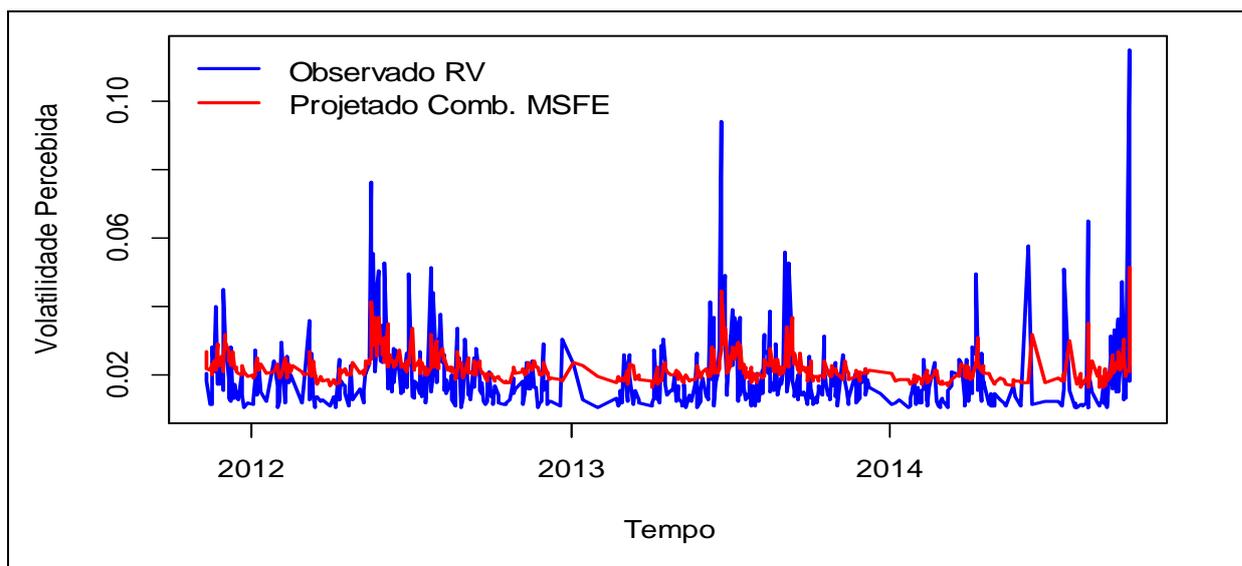


Figura 2 – Série observada e projetada

## 5 CONSIDERAÇÕES FINAIS

O objetivo do presente artigo foi propor a combinação das projeções de Volatilidade Percebida obtidas pelos modelos ARFIMA e *Nearest Neighbor* (NN). Conforme Timmermann (2006) e diversos outros autores, a combinação de projeções demonstrou sucesso empírico em diversas aplicações econômicas, como projeção do PIB, preço do petróleo e retorno de ações.

Combinar ambas as técnicas para projeção da Volatilidade Percebida é inovativo, dado que duas áreas de conhecimento diferentes são utilizadas para melhorar o desempenho de projeções. A primeira é baseada em pressupostos da economia, enquanto que a segunda teve origem na área de física e engenharia.

Para analisar os resultados foi utilizada uma amostra da série de Volatilidade Percebida diária do Ibovespa, no período entre 03/01/2000 a 01/10/2014, com um total de 2714 observações. Nos resultados fora da amostra, observou-se que os modelos ARFIMA demonstraram melhor performance em relação ao NN. Na comparação geral, a combinação de projeções pelo método de pesos MSFE apresentou menores valores de MSE e RMSE.

Como limitações, o presente artigo analisou somente a série de Volatilidade Percebida do Ibovespa, sendo necessário verificar os resultados em mais séries. O estimador da Volatilidade Percebida utilizado foi o proposto por Andersen e Bollerslev (1998), sendo que medidas mais robustas a saltos e ruídos de microestrutura foram propostas nos últimos anos.

Vale ressaltar que os modelos estimados utilizaram a série pura da Volatilidade Percebida, sendo que em diversos trabalhos da área, os autores utilizaram séries em  $\log RV$ . Diante disto, em trabalhos futuros, os modelos deverão ser ajustados com séries no formato citado, além de mais séries devem ser utilizadas.

Em trabalhos futuros, mais métodos de combinação de projeções devem ser estudados, sendo que no presente trabalho limitou-se ao estudo dos métodos de pesos iguais e MSFE. Ainda em relação a combinação de projeções, existe uma necessidade de trabalhos na área, dado que

ainda não existe um consenso entre os pesquisadores sobre qual peso ótimo deveria ser utilizado.

Em trabalhos futuros, devem-se propor métodos alternativos que considerem a minimização do risco de uma previsão ineficiente para proteção das perdas. Buscando-se posicionar mais pesos para as projeções que apresentem as menores chances de evidenciar riscos abaixo de algum nível esperado.

## REFERÊNCIAS

ALDRIDGE, I. *High-Frequency trading: a practical guide to algorithmic strategies and trading systems*. 2. ed. New Jersey: John Wiley & Sons Ltd, 2013.

ANDERSEN, T. G.; BOLLERSLEV, T. *Answering the skeptics: Yes, standard volatility models do provide accurate forecasts*. *International Economic Review*, v. 39, n. 4, p. 885-905, 1998.

ANDERSEN, T. G. et al. *The distribution of realized stock return volatility*. *Journal of Financial Economics*, v. 61, p. 43-76, 2001.

ANDERSEN, T. G. et al. *Modeling and forecasting realized volatility*. *Econometrica*, v. 71, n. 2, p. 579-625, 2003.

ANDERSEN, T. G.; BOLLERSLEV, T.; DIEBOLD, F. X. *Roughing it up: including jump components in the measurement, modelling and forecasting of return volatility*. *The Review of Economics and Statistics*, v. 89, p. 701-720, 2007.

ANDRADA-FELIX, J.; FERNANDEZ-RODRIGUEZ, F.; FUERTES, A. *Combining nearest neighbor predictions and model-based predictions of realized variance: Does it pay?* *Working Paper*, 2013. Disponível em: <http://ssrn.com/abstract=2258457>. Acesso em: 20/01/2014.

ARAÚJO, A. C.; MONTINI, A. A. *High Frequency Trading: Preço, Volume e Volatilidade em uma Nova Microestrutura de Mercado*. In: SEMINÁRIOS EM ADMINISTRAÇÃO - SEMEAD, 17, 2014, São Paulo. *Anais...* São Paulo: SEMEAD, 2014.

BARNDORFF-NIELSEN, O. E.; SHEPHARD, N. *Power and bipower variation with stochastic volatility and jumps*. *Journal of Financial Econometrics*, Oxford Univ. Press, v. 2, n. 1, p. 1-37, 2004.

BAUMEISTER, C.; GUERIN, P.; KILIAN, L. *Do high-frequency financial data help forecast oil prices? the MIDAS touch at work*. *Working Paper*, 2014.

BAUWENS, L.; HAFNER, C.; LAURENT, S. *Volatility models and their applications*. New Jersey: John Wiley & Sons, 2012.

BOUDT, K.; CORNELISSEN, J.; PAYSEUR, S. *Highfrequency: Toolkit for the analysis of highfrequency financial data in R*. *Working Paper*, 2013.

BOX, G. E. P.; JENKINS, G. *Time Series Analysis, Forecasting and Control*. HoldenDay: San Francisco, 1970.

CAVALERI, R. **Combinações de previsões aplicada à volatilidade**. Dissertação (Mestrado). Programa de Pós-Graduação da Faculdade de Ciências Econômicas da Universidade Federal do Rio Grande do Sul - UFRGS, 2008.

CHAN, Y.; STOCK, J.; WATSON, M. *A Dynamic Factor Model Framework for Forecast Combination*. *Spanish Economic Review*, v. 1, p. 91–121, 1999.

CLEMEN, R. T. *Combining forecasts: a review and annotated bibliography*. *International Journal of Forecasting*, v. 5, p. 559-583, 1989.

DANIÉLSSON, J. *Financial risk forecasting: The theory and practice of forecasting market risk with implementation in R and Matlab*. Chichester: John Wiley & Sons, 2011.

FARMER, J. D.; SIDOROWICH, J. J. *Predicting chaotic time series*. *Physical Review Letters*, v. 59, n. 8, p. 845-848, 1987.

FORSBERG, L.; GHYSELS, E. *Why do absolute returns predict volatility so well?* *Journal of Financial Econometrics*, v. 5, n. 1, p. 31-67, 2007.

GENRE, V. *et al.* *Combining expert forecasts: Can anything beat the simple average?* *International Journal of Forecasting*, v. 29, n. 1, p. 108-121, 2013.

GRAEFE, A. *et al.* *Combining forecasts: An application to elections*. *International Journal of Forecasting*, v. 30, n. 1, p. 43-54, 2014.

GRANGER, C.; JOYEUX, R. *An introduction to long memory times series and fractional differencing*. *Journal of Time Series Analysis*, v. 1, p. 15-29, 1980.

HANSEN, P. R.; LUNDE, A. *Forecasting volatility using high-frequency data*. In: CLEMENTS, M. P.; HENDRY, D. F. *The Oxford Handbook of Economic Forecasting*. Oxford, UK: Oxford University Press, 2012.

HYNDMAN, R. J.; KHANDAKAR, Y. *Automatic Time Series Forecasting: The forecast Package for R*. *Journal of Statistical Software*, v. 27, n. 3, p. 1-22, 2008.

HOSKING, J. *Fractional differencing*. *Biometrika*, v. 68, n. 1, p. 165-176, 1981.

HUANG, H.; LEE, T. *Forecasting Value-at-Risk using high-frequency information*. *Econometrics*, v. 1, p. 127-140, 2013.

OOMEN, R. C. A. *Using high-frequency data stock market index data calculate, model and forecast realized return variance*. *Economics Discussion Paper No. 2001/6*, 2001. Disponível em: [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=267498](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=267498). Acesso em: 16/03/2014.

PATTON, A. J. *Volatility forecast comparison using imperfect volatilities proxies*. *Journal of Econometrics*, v. 160, p. 246-256, 2011.

PONTES, T. T. S. Previsão da volatilidade realizada: o impacto dos saltos na série do Ibovespa. In: ENCONTRO DA ANPAD - ENANPAD, 38, 2014, Rio de Janeiro. **Anais...** Rio de Janeiro: ENANPAD, 2014.

PORTNOY, K. *High frequency trading and the stock market: a look at the effects of trade volume on stock price changes.* **The Park Place Economist**, v. 19, n. 1, p. 68-75, 2011.

MAKRIDAKIS, S. *et al.* *The accuracy of extrapolation (time series) methods: Results of a forecasting competition.* **Journal of forecasting**, v. 1, n. 2, p. 111-153, 1982.

MAKRIDAKIS, S.; WINKLER, R. L. *Averages of forecasts: Some empirical results.* **Management Science**, v. 29, n. 9, p. 987-996, 1983.

NEWBOLD, P.; GRANGER, C. W. J. *Experience with forecasting univariate time series and the combination of forecasts.* **Journal of the Royal Statistical Society**, v. 137, n. 2, p. 131-165, 1974.

RYDBERG, T. H.; SHEPHARD, N. *Dynamics of trade-by-trade price movements: decomposition and models.* **Journal of Financial Econometrics**, v. 1, n. 1, p. 2-25, 2003.

SMITH, J.; WALLIS, K. F. *A simple explanation of the forecast combination puzzle.* **Oxford Bulletin of Economics and Statistics**, v. 71, n. 3, p. 331-355, 2009.

STOCK, J. H.; WATSON, M. W. *A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series.* **Working Paper**, 2001.

STOCK, J. H.; WATSON, M. W. *Combination forecasts of output growth in a seven country data set.* **Journal of Forecasting**, v. 23, p. 405-430, 2004.

TIMMERMANN, A. *Forecast combinations.* **Handbook of economic forecasting**, v. 1, p. 135-196, 2006.

VAL, F. F.; PINTO, A. C. F.; KLOTZLE, M. C. Volatilidade e previsão de retorno com modelos de alta frequência e GARCH: evidências para o mercado brasileiro. **Revista Contabilidade & Finanças**, v. 25, n. 65, p. 189-201, 2014.

WINK JÚNIOR, M. V.; PEREIRA, P. L. V. *Modeling and Forecasting of Realized Volatility: Evidence from Brazil.* **Brazilian Review of Econometrics**, v. 31, n. 2, p. 315-337, 2011.

WINKLER, Robert L.; MAKRIDAKIS, Spyros. *The Combination of Forecasts.* **Journal of the Royal Statistical Society**. v. 146, n. 2, p.150-157, 1983.

YAKOWITZ, S. *Nearest-Neighbour methods for time series analysis.* **Journal of Time Series Analysis**, v. 8, n. 2, p. 235-247, 1987.

ZIVOT, E. *Analysis of high frequency financial data: models, methods and software. Part II: Modeling and forecasting realized variance measures.* **Technical Report UWEC-2005-03**, 2005.