TEXTO PARA DISCUSSÃO

No. 547

The benefits of bagging for forecast models for realized volatility

Eric Hillebrand
Marcelo C. Medeiros

DEPARTAMENTO DE ECONOMIA
www.econ.puc-rio.br

THE BENEFITS OF BAGGING FOR FORECAST MODELS OF REALIZED VOLATILITY

ERIC HILLEBRAND AND MARCELO C. MEDEIROS

ABSTRACT. This paper shows that bagging can improve the forecast accuracy of time series models for realized volatility. We consider 23 stocks from the Dow Jones Industrial Average over the sample period 1995 to 2005 and employ two different forecast models, a log-linear specification in the spirit of the heterogeneous autoregressive model and a nonlinear specification with logistic transitions. Both forecast model types benefit from bagging, in particular in the 1990s part of our sample. The log-linear specification shows larger improvements than the nonlinear model. Bagging the log-linear model yields the highest forecast accuracy on our sample.

1. INTRODUCTION

We consider the problem of forecasting stock market volatility. The aim is to apply bagging (bootstrap aggregation), a recently proposed statistical learning technique, to realized volatility, a recently proposed improved measure of asset price variance. Time series models for realized volatility often comprise lags of realized volatility aggregated over different time horizons, lagged returns cumulated over different time horizons, and other possible exogenous variables. The selection of these variables puts the forecaster in a situation where bagging can help: There are potentially many regressors to choose from and the individual regressors have little forecast power (Inoue and Kilian in press, Inoue and Kilian 2004).

We find that bagging can substantially improve forecast mean squared errors for realized volatility. We consider a log-linear forecast model specification that is commonly used and can approximate long range dependence by linearly combining aggregated past volatility at different time scales (Corsi 2004, Andersen, Bollerslev, and Diebold 2007). In addition, we consider a nonlinear neural network specification that apart from long range dependence captures possible threshold and transition effects. These nonlinear phenomena have been documented in many studies of realized volatility and earlier latent volatility models (e.g. Nelson 1991, Glosten, Jagannathan, and Runkle 1993, Martens,

Date: August 13, 2009.

van Dijk, and de Pooter 2004, Hillebrand and Medeiros to appear, McAleer and Medeiros 2006, Hillebrand and Medeiros 2008). For our sample of 23 stocks from the set of the Dow Jones Industrial Average index and the period from 1995 to 2005, bagging results in improved forecast accuracy for both model types. The log-linear model benefits more from bagging than the nonlinear specification, in particular for forecasts made for the 1990s.

The idea of bagging was introduced in Breiman (1996), studied more rigorously in Bühlmann and Yu (2002), and introduced to econometrics in Inoue and Kilian (2004). Bagging is motivated by the observation that in models where statistical decision rules are applied to choose from a set of predictors, such as significance in pre-tests, the set of selected regressors is data-dependent and random. Bootstrap replications of the raw data are used to re-evaluate the selection of predictors, to generate bootstrap replications of forecasts, and to average over these bootstrapped forecasts. It has been shown in a number of studies that bagging reduces the mean squared error of forecasts considerably by averaging over the randomness of variable selection (Inoue and Kilian in press, Lee and Yang 2006). Applications include, among others, financial volatility (Huang and Lee 2007a), equity premia (Huang and Lee 2007b), and employment data (Rapach and Strauss 2007).

Realized volatility was introduced in Andersen and Bollerslev (1998) and has developed into a large literature that is concerned with finding consistent and robust estimators of realized volatility as well as time series models for realized volatility (e.g. Andersen, Bollerslev, Diebold, and Ebens 2001, Andersen, Bollerslev, Diebold, and Labys 2001, Andersen, Bollerslev, Diebold, and Labys 2003, Barndorff-Nielsen and Shephard 2002, Corsi 2004, Zhang, Mykland, and Aït-Sahalia 2005, Barndorff-Nielsen, Hansen, Lunde, and Shephard in press, Christensen, Oomen, and Podolskij 2008). We present a brief outline of the concept in Section 2. Realized volatility inherits all stylized facts that have been established for volatility in earlier latent variable specifications, most notably long-range dependence (e.g. Engle 1982, Engle and Bollerslev 1986, Bollerslev 1987, Ding, Granger, and Engle 1993, Baillie, Bollerslev, and Mikkelsen 1996).

The paper is organized as follows: Section 2 describes the concept of realized volatility. Section 3 defines the two model classes that are considered for bagging. Section 4 defines the bagging schemes. Section 5 describes the data set and presents the empirical findings. Section 6 concludes.

2. REALIZED VOLATILITY

Suppose that on day t the logarithmic price p of a given asset at time $t+\tau$, $0 \leq \tau \leq 1$, follows a continuous time diffusion:

$$dp(t+\tau) = \mu(t+\tau)d\tau + \sigma(t+\tau)dW(t+\tau),\ 0 \leq \tau \leq 1,\ t = 1, \ldots, T,$$

where $\mu(t+\tau)$ is the drift component, $\sigma(t+\tau)$ is instantaneous volatility that may be deterministic or stochastic, and $W(t+\tau)$ is standard Brownian motion.

Andersen, Bollerslev, Diebold, and Labys (2001), Andersen, Bollerslev, Diebold, and Labys (2003) and Barndorff-Nielsen and Shephard (2002), among others, consider daily compound returns $r(t) = p(t) - p(t-1)$. Set $\mathcal{F}_t = \mathcal{F}\left\{\mu(t+\tau-1), \sigma(t+\tau-1)\right\}_{\tau=0}^{\tau=1}$ as sigma-algebra generated by the sample paths of the drift and diffusion processes $\mu(t+\tau-1)$ and $\sigma(t+\tau-1)$ but *not* by the Brownian motion $W(t+\tau-1)$, $0 \leq \tau \leq 1$. Define

$$IV_t = \int_0^1 \sigma^2(t+\tau-1)d\tau = \text{Var}(r(t)|\mathcal{F}_t), \tag{1}$$

or *integrated variance*, as the object of interest in realized volatility theory.

In practical applications, prices are observed at discrete and irregularly spaced intervals and there are many ways to sample the data. Suppose that on a given day t, we partition the interval [0,1] and define the grid of observation times $\{\tau_0, \ldots, \tau_n\}$, $0 = \tau_0 < \tau_1 < \cdots, \tau_n = 1$. The length of the ith subinterval is given by $\delta_i = \tau_i - \tau_{i-1}$. The most widely used sampling scheme is calendar time sampling, where the intervals are equidistant in calendar time, that is $\delta_i = 1/n$. Let $p_{t,i}$, $i = 1, \ldots, n$, be the ith price observation during day t, such that $r_{t,i} = p_{t,i} - p_{t,i-1}$ is the ith intra-period return of day t. Realized volatility is defined as

$$RV(t) = \sqrt{\sum_{i=2}^{n} r_{t,i}^2}. \tag{2}$$

Under additional regularity conditions including the assumption of uncorrelated intra-day returns, realized variance RV_t^2 is a consistent estimator of integrated variance, such that $RV_t^2 \xrightarrow{p} IV_t$. When returns are serially correlated, however, realized variance is a biased estimator of integrated variance. Serial correlation may be the result of market microstructure effects such as bid-ask bounce and discreteness of prices (Campbell, Lo, and MacKinlay 1997, Madhavan 2000, Biais, Glosten, and Spatt 2005). These effects prevent very fine sampling partitions. Realized volatility is therefore not an error-free measure of

volatility. Barndorff-Nielsen and Shephard (2002) study the properties of the estimation error of realized volatility.

The search for asymptotically unbiased, consistent, and efficient methods for measuring realized volatility in the presence of microstructure noise has been one of the most active research topics in financial econometrics over the last few years. While early references in the literature, such as Andersen, Bollerslev, Diebold, and Ebens (2001), advocated the simple selection of an arbitrary lower frequency (typically 5-15 minutes) to balance accuracy and the dissipation of microstructure bias, a procedure that is known as sparse sampling, recent articles have developed estimators that dominate this procedure. The currently available consistent estimators are the realized kernel estimator of Barndorff-Nielsen, Hansen, Lunde, and Shephard (in press), the modified MA filter of Hansen, Large, and Lunde (2008), the two time scales estimator of Zhang, Mykland, and Aït-Sahalia (2005), and the quantile-based estimator of Christensen, Oomen, and Podolskij (2008). For the purposes of this paper, we choose the realized kernel estimator of Barndorff-Nielsen, Hansen, Lunde, and Shephard (in press), which is robust to microstructure noise.

3. Modeling and Forecasting Realized Volatility

Let RV_t^2 be a consistent and unbiased estimator for integrated variance on day t. In this paper, we consider two different forecasting models for $\log(RV_t)$, a log-linear model and a nonlinear extension. As explanatory variables, we consider lags of realized volatility, day-of-the-week dummies, dummies for days where macroeconomically relevant announcements were made, and past cumulative returns that capture possible leverage effects.

3.1. **The Log-Linear Heterogenous Autoregressive Model with Exogenous Variables.**

3.1.1. *Model Definition.* The starting point for our log-linear model specification is the Heterogeneous Autoregressive (HAR) model proposed by Corsi (2004). Let

$$y_{t,k} = \frac{1}{k}\sum_{i=1}^{k} y_{t+1-i}$$

be the average over the last k observations of a time series and consider the model

$$y_t = \beta_0 + \sum_{\iota_j \in I} \beta_j y_{t-1,\iota_j} + \varepsilon_t, \tag{3}$$

where $I = (\iota_1, \iota_2, \ldots, \iota_p)$ is a set of p indices with $\iota_1 < \iota_2 < \ldots < \iota_p$, $j = 1, \ldots, p$. Throughout this paper, ε_t is a zero-mean process with finite, but not necessarily constant variance (Corsi, Mittnik, Pigorsch, and Pigorsch 2008). By substituting the definition of y_{t,ι_j} into (3), we can write

$$y_t = \beta_0 + \frac{\beta_1}{\iota_1}\sum_{i=1}^{\iota_1} y_{t-i} + \frac{\beta_2}{\iota_2}\sum_{i=1}^{\iota_2} y_{t-i} + \cdots + \frac{\beta_p}{\iota_p}\sum_{i=1}^{\iota_p} y_{t-i} + \varepsilon_t, \tag{4}$$

a representation that shows that we are considering a restricted AR(p) model (Craioveanu and Hillebrand 2008). Corsi (2004) proposes model (3) with $y_t = \log(RV_t)$ and $I = (1, 5, 22)$. His specification builds on the HARCH model proposed by Müller, Dacorogna, Dave, Olsen, Pictet, and von Weizsaecker (1997). It captures long-range dependence by aggregating realized volatility over the different time scales in I.

In this paper, we consider a slightly more general version of the HAR model that includes deterministic predictors, a flexible lag structure, and past cumulative returns. Let r_t be the daily log-return, $y_t = \log(RV_t)$, and define, with some abuse of notation,

$$r_{t,k} = \sum_{i=1}^{k} r_{t+1-i}.$$

Let $I = (\iota_1, \ldots, \iota_p)$ and $K = (\kappa_1, \ldots, \kappa_q)$ be two sets of indices, $\mathbf{x}_t = (1, y_{t,\iota_1}, \ldots, y_{t,\iota_p})$, and $\mathbf{r}_t = \left(r_{t,\kappa_1}, \ldots, r_{t,\kappa_q}\right)$. The general HAR model with exogenous variables considered in this paper may be written as

$$y_t = \boldsymbol{\alpha}'\mathbf{w}_t + \boldsymbol{\beta}'\mathbf{x}_{t-1} + \boldsymbol{\rho}'\mathbf{r}_{t-1} + \varepsilon_t, \tag{5}$$

where $\mathbf{w}_t$ is is a vector of dummy variables for weekdays and macroeconomic announcement days, the vector $\mathbf{x}_{t-1}$ contains the usual HAR predictors with possibly more and/or other lags than $(1, 5, 22)$, and $\mathbf{r}_{t-1}$ includes past cumulative returns over different horizons. The vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\rho}$ contain parameters.

The inclusion of announcement dates is motivated by the possible presence of jumps in the volatility process (Andersen, Bollerslev, and Diebold 2007, Barndorff-Nielsen and Shephard 2006, McAleer and Medeiros 2006, Scharth and Medeiros 2006). Several studies consider days-of-the-week dummies in volatility models (Engle and Ng 1993, Martens, van Dijk, and de Pooter 2004). The results in Scharth and Medeiros (2006) and Fernandes, Medeiros, and Scharth (2007) motivate the inclusion of past cumulative returns as possible predictors. The authors show in different frameworks that cumulative

returns over long horizons up to three months improve the forecasting power of models for realized volatility.

3.1.2. *Specification and Estimation.* Typically, the index set $(1, 5, 22)$ is considered in the HAR model. A flexible choice of the lag structure imposes high computational costs. Two parameters need to be set a priori, a maximum time scale to consider and the number of predictors to be used. For example, if we decide to include two time scales ι_2 and ι_3 in addition to $\iota_1 = 1$, and we allow for a maximum time horizon of 250 days, we have to consider $\binom{250}{2} = 31125$ different specifications and pick the one with the highest likelihood, lowest AIC or BIC, or any other criterion. Even for a single time series, the computational costs are substantial. See Craioveanu and Hillebrand (2008) for a study that explores the optimal lag structure for HAR models. For bagging, where we have to repeat this procedure for every bootstrap sample, implementation is prohibitive at this point. To circumvent this problem, we define a grid of lags to choose from such that the selection problem remains tractable and include these lags as regressors in $\mathbf{x}_t$. In this study, we set $I = (1, 2, \ldots, 22)$ and $K = (1, 2, \ldots, 100)$. Equation (5) can then be estimated in a linear regression and we pre-test for regressor significance. Let $\mathbf{z}_t := \left(\mathbf{w}_t', \mathbf{x}_{t-1}', \mathbf{r}_{t-1}'\right)'$. Following the notation in Inoue and Kilian (in press), we can define the one-step ahead forecast of $y_t = \log(RV_t)$ as

$$\hat{y}_{t|t-1} := \mathbb{E}\left[y_t|\mathcal{F}_{t-1}\right] = \begin{cases} 0 & \text{if } |t_j| < c\,\forall j, \\ \widehat{\boldsymbol{\lambda}}\widetilde{\mathbf{z}}_t & \text{otherwise,} \end{cases} \tag{6}$$

where $\widetilde{\mathbf{z}}_t := \mathbf{S}_t\mathbf{z}_t$, $\mathbf{S}_t$ is a diagonal selection matrix with jth diagonal element given by

$$\mathbb{1}_{\{|t_j|>c\}} = \begin{cases} 1 & \text{if } |t_j| > c, \\ 0 & \text{otherwise,} \end{cases}$$

c is a pre-specified critical value of the test, and $\widehat{\boldsymbol{\lambda}}$ is the ordinary least squares estimator given by

$$\widehat{\boldsymbol{\lambda}} = \left[\sum_{t=1}^{T} \widetilde{\mathbf{z}}_t\widetilde{\mathbf{z}}_t'\right]^{-1} \sum_{t=1}^{T} \widetilde{\mathbf{z}}_t' y_t.$$

3.2. **The Neural Network Heterogenous Autoregressive Model and Bayesian Regularization.**

3.2.1. *Model Definition.* In order to compare the out-of-sample performance of the log-linear model described above we consider a nonlinear extension of the HAR model. Let

$y_t = \log(RV_t)$ be generated by the following stochastic process

$$y_t = T(\mathbf{z}_t) + \varepsilon_t, \tag{7}$$

where $T(\mathbf{z}_t)$ is an unknown nonlinear function of $\mathbf{z}_t = \left(\mathbf{w}_t', \mathbf{x}_{t-1}', \mathbf{r}_{t-1}'\right)'$, and ε_t is a zero-mean process with finite variance.

The goal of modeling techniques based on neural networks is to approximate (7) by the following nonlinear specification

$$T(\mathbf{z}_t) \approx G(\mathbf{z}_t; \boldsymbol{\psi}) := \boldsymbol{\beta}_0' \mathbf{z}_t + \sum_{m=1}^{M} \beta_m f(\mathbf{z}_t; \boldsymbol{\gamma}_m), \tag{8}$$

where $G(\mathbf{z}_t; \boldsymbol{\psi})$ is a nonlinear function of the variables $\mathbf{z}_t$ that is indexed by the vector of parameters $\boldsymbol{\psi}$, and $f(\mathbf{z}_t; \boldsymbol{\gamma}_m)$ is the logistic function

$$f(\mathbf{z}_t; \boldsymbol{\gamma}_m) = \frac{1}{1 + e^{-\boldsymbol{\gamma}_m' \mathbf{z}_t}}$$

with state variable $\mathbf{z}_t$ and slope parameters $\boldsymbol{\gamma}_m$.

As first discussed in Kuan and White (1994), the model defined by equation (8) may alternatively have a parametric or a nonparametric interpretation. In the parametric interpretation, the model can be viewed as a kind of smooth transition regression where the transition variable is an unknown linear combination of the explanatory variables in $\mathbf{z}_t$ (van Dijk, Teräsvirta, and Franses 2002). In this case, there is an optimal, fixed number M of logistic transitions that can be understood as the number of limiting regimes (Trapletti, Leisch, and Hornik 2000, Medeiros and Veiga 2000, Medeiros, Teräsvirta, and Rech 2006). On the other hand, for $M \to \infty$ the neural network model is a representation of any Borel-measurable function over a compact set (Grenander 1981, Hornik, Stinchombe, and White 1989, Hornik, Stinchcombe, White, and Auer 1994, Chen and Shen 1998, Chen and White 1998, Chen, Racine, and Swanson 2001). For large M, this representation suggests a nonparametric interpretation as series expansion, sometimes referred to as sieve-approximator. In this paper, we adopt the nonparametric interpretation of the neural network model and show that it approximates typical nonlinear behavior of realized volatility well. Neural network approaches to volatility forecasting have been successfully taken in Donaldson and Kamstra (1997), Hu and Tsoukalas (1999), and Hamid and Iqbal (2004).

3.2.2. *Specification and Estimation.* Usually, $\boldsymbol{\psi}$ is estimated by nonlinear least-squares

$$\widehat{\boldsymbol{\psi}} = \underset{\boldsymbol{\psi}}{\operatorname{argmin}} \mathcal{Q}_T(\boldsymbol{\psi}) = \underset{\boldsymbol{\psi}}{\operatorname{argmin}} \sum_{t=1}^{T} \left[\log(RV_t) - G(\mathbf{z}_t; \boldsymbol{\psi})\right]^2, \tag{9}$$

and the estimated residuals $\widehat{\varepsilon}_t = \log(RV_t) - G(\mathbf{z}_t; \widehat{\boldsymbol{\psi}})$ are an approximation to the true error term ε_t in (7). In most applications, a simple gradient descent algorithm is used to estimate $\boldsymbol{\psi}$.

Approximating (7) by (8) poses two main problems. First, the true vector of variables $\mathbf{z}_t$ is not known in advance and the modeler has to determine which variables should be included in $\mathbf{z}_t$. The second problem is related to the selection of the number M of logistic functions in (8). Selecting a small number of hidden units leads to a poor approximation of the true data generating process. On the other hand, a model with a large number of hidden units may be overfitted and have little forecast accuracy. In most neural network applications, it is customary to select the variables and the number of hidden units using some rule of thumb. A vast number of models with different combinations of variables and numbers of hidden units are estimated and the one with the best performance according to some criterion is chosen as the final specification. Several alternatives to this rule of thumb have appeared in the literature. The simplest one is the so-called *early stopping*. The key idea is to split the available data into three subsets. The first subset is used to estimate the parameters. The second subset is called the validation set. The error on the validation set is monitored during the estimation process. When the network begins to overfit the data, the error on the validation set typically begins to rise. When the validation error increases for a specified number of iterations, the estimation process is discontinued, and the parameters estimated at the minimum of the validation error serve as final estimates. The third subset called the test set is not used for estimation and is saved for comparing the out-of-sample performance of different models. The model with the best forecasting performance is chosen as the final specification.

Pruning is another popular technique to find the smallest network that fits the data well and produces good forecasts. The main idea is to start with a large network and sequentially reduce its size by removing some network connections (Reed 1993). Anders and Korn (1999) compared a number of different methodologies including a simplified version of the specific-to-general approach of Medeiros, Teräsvirta, and Rech (2006) and information criteria (such as AIC or BIC) pruning.

In this paper, we adopt the Bayesian regularization approach proposed by MacKay (1992). The fundamental idea is to find a balance between the number of parameters and goodness of fit by penalizing large models. The objective function is modified in such a way that the estimation algorithm reduces the network by driving irrelevant parameter estimates to zero during the optimization. The parameter vector $\boldsymbol{\psi}$ is estimated as

$$\widehat{\boldsymbol{\psi}} = \underset{\boldsymbol{\psi}}{\mathsf{argmin}}\, \widetilde{\mathcal{Q}}_T(\boldsymbol{\psi}) = \underset{\boldsymbol{\psi}}{\mathsf{argmin}}\, \left[\eta \mathcal{Q}_T(\boldsymbol{\psi}) + \gamma \mathcal{Q}^*(\boldsymbol{\psi})\right], \tag{10}$$

where $\mathcal{Q}_T(\boldsymbol{\psi}) = \sum_{t=1}^{T} \left[\log(RV_t) - G(\mathbf{z}_t; \boldsymbol{\psi})\right]^2$, $\mathcal{Q}^*(\boldsymbol{\psi})$ is the *regularization* or *penalty* term, and $\eta, \gamma > 0$ are *objective function* or *regularization* parameters. The penalty term is usually chosen to be the sum of squared parameters $\mathcal{Q}^*(\boldsymbol{\psi}) = \boldsymbol{\psi}'\boldsymbol{\psi}$.

The forecast accuracy of the neural network model depends crucially on the values of η and γ, especially in small samples. The relative size of the objective function parameters determines the emphasis of the estimation process. If $\eta >> \gamma$, then the optimization algorithm places more weight on error minimization and the network may still overfit. If $\eta << \gamma$, the optimization emphasizes network size reduction at the expense of error size, thus producing a smoother function of the input variables. The main problem with implementing regularization is setting the correct values for these objective function parameters. One approach to determine the optimal objective function parameters is the Bayesian framework, where the parameters of the network are assumed to be random variables with well-specified distributions. The objective function parameters are related to the unknown variances associated with these distributions. Foresee and Hagan (1997) present a detailed discussion of the use of Bayesian regularization in combination with the Levenberg-Marquardt optimization algorithm. The main advantage of this method is that even if the neural network model is over-parametrized, the irrelevant parameter estimates are likely to be close to zero and the model behaves like a small network. See the appendix for more details.

Bayesian regularization has been shown to be a very flexible modeling approach. Teräsvirta, van Dijk, and Medeiros (2005) show that neural network models specified and estimated with Bayesian regularization outperform models that employ the specific-to-general approach proposed by Medeiros, Teräsvirta, and Rech (2006). Medeiros, Teräsvirta, and Rech (2006) present simulation evidence that AIC and BIC tend to underestimate the number of hidden units; see also Anders and Korn (1999). Another successful application of Bayesian regularization can be found in Medeiros, Veiga, and Pedreira (2001).

As suggested in the literature, we tried different numbers of hidden units in the neural network model, ranging from $M = 1$ to $M = 10$. As the number of predictors is high, increasing the number of hidden units increases the complexity of the model and the computational burden to estimate it substantially without yielding additional benefits. In our application, we found that when more than five hidden units are used, the Bayesian regularization approach always reduces the model complexity, so we start the Bayesian regularization algorithm at $M = 5$.

4. Bagging Predictors

Realized volatility models lend themselves to bagging since they involve selection of lags of logarithmic realized volatility, cumulative returns over different horizons, and dummies for weekdays and announcement dates. The optimal predictor structure is data-dependent in the sense that pre-tests as described in Section 3.1.2 apply indicators of the type $\mathbb{1}_{\{t>c\}}$ to predictor variables. This is the standard situation for the application of bagging as described in Bühlmann and Yu (2002), Section 2. Inoue and Kilian (in press) show for inflation data in an analogous problem that bagging averages over the randomness of predictor selection, resulting in lower forecast mean squared errors. We consider the following bagging schemes.

Proposal 1 (Bagging the HAR model). *The bagging forecast in the extended HAR model is defined as follows:*

(1) *Arrange the set of tuples $(y_t, \mathbf{z}_t')$, $t = 2, \ldots, T$, in the form of a matrix $\mathbf{X}$ of dimension $(T-1) \times k$, where k is the number of regressors in $\mathbf{z}_t$.*
(2) *Construct bootstrap samples of the form $\left\{ \left(y^*_{(i)2}, \mathbf{z}'^*_{(i)2} \right), \ldots, \left(y^*_{(i)T}, \mathbf{z}'^*_{(i)T} \right) \right\}$, $i = 1, \ldots, B$, by drawing blocks of m rows of $\mathbf{X}$ with replacement, where the block size m is chosen to capture possible dependence in the error term of the realized volatility series.*
(3) *Following Inoue and Kilian (in press), compute the ith bootstrap forecast as*

$$\hat{y}^*_{(i)t|t-1} = \begin{cases} 0 & \text{if } |t^*_j| < c \, \forall j, \\ \widehat{\boldsymbol{\lambda}}^*_{(i)} \widetilde{\mathbf{z}}^*_{(i)t} & \text{otherwise}, \end{cases} \tag{11}$$

where $\widetilde{\mathbf{z}}^*_{(i)t} := \mathbf{S}^*_{(i)t} \mathbf{z}^*_{(i)t}$ and $\mathbf{S}^*_{(i)t}$ is a selection matrix as in equation (6) that depends on the bootstrap sample. As is common, the asterisk indicates bootstrap replications.

(4) *Compute the average forecast over the bootstrap samples:*

$$\tilde{y}_{t|t-1} = \frac{1}{B}\sum_{i=1}^{B} \hat{y}^*_{(i)t|t-1},$$

where the tilde indicates the bagging forecast.

PROPOSAL 2 (Bagging the neural network model). *The bagging predictor in the nonlinear model is defined as follows:*

(1) *Repeat steps (1) and (2) in Proposal 1.*
(2) *For each bootstrap sample, first remove insignificant regressors by pre-testing as in step (3) of Proposal 1. Then, estimate the nonlinear model using Bayesian Regularization with $M = 5$ in (8). Note that for each bootstrap sample, the optimal selection of variables and hidden units are different. Compute the ith bootstrap forecast and call it $\hat{y}^*_{(i)t|t-1}$.*
(3) *Compute the average forecast over the bootstrap samples:*

$$\tilde{y}_{t|t-1} = \frac{1}{B}\sum_{i=1}^{B} \hat{y}^*_{(i)t|t-1}.$$

Following Hall, Horowitz, and Jing (1995), we choose a block size of $m = T^{1/3}$ for the moving block bootstrap procedure in both proposals. This allows for dependence in the error term of equation (5) (Corsi, Mittnik, Pigorsch, and Pigorsch 2008). The critical value c of the test statistic is set equal to 1.96, corresponding to a two-sided test at the 95% confidence level. We also tried other critical values. Confidence levels much lower than 95% worsen the bagging performance while levels stricter than 95% do not improve the results further.

5. APPLICATION: BAGGING REALIZED VOLATILITY

5.1. **Data.** We use high-frequency tick-by-tick trades on 23 stocks from the set of the Dow Jones Industrial Average index as listed in Table 1. The data are obtained from the NYSE TAQ (Trade and Quote) database. The sample period covers 3-Jan-1995 to 31-Dec-2005. The selection of 23 out of the 30 stocks of the index is motivated by data availability; these are the stocks for which we can obtain data that cover the entire sample period.

In calculating daily realized volatility, we employ the realized kernel estimator with modified Tukey-Hanning weights of Barndorff-Nielsen, Hansen, Lunde, and Shephard

(in press). We start by cleaning the data for outliers. We consider transactions between 9.30 am through 4.00 pm. Following Barndorff-Nielsen, Hansen, Lunde, and Shephard (in press) we employ the following 60 second activity fixed tick time sampling scheme: $fq_t = 1+60n_t/(\tau_{0t}-\tau_{nt})$, where fq_t is the sampling frequency, n_t represents the number of transactions for day t, and τ_{0t}, τ_{nt} are the times for the first and last trade for day t. This is tick-time sampling chosen such that the same number of observations is obtained each day.

Following the realized volatility literature, we focus on logarithmic realized volatility. Figure 7 shows the daily time series of returns, realized volatility, and logarithmic realized volatility for WMT, a typical stock in the sample. We also consider dummies for the days of the week as in Martens, van Dijk, and de Pooter (2004) and dummies for the following macroeconomic announcements: Federal Open Market Committee meetings, The Employment Situation Report from the Bureau of Labor Statistics, CPI and PPI price indices; see also Scharth and Medeiros (2006) and McAleer and Medeiros (2006).

5.2. **Results.** In this section, we apply bagging Proposals 1 and 2 to the log-linear model and to the neural network specification. We find that bagging improves the forecast performance of the log-linear model, in particular in the late 1990s. The neural network specification also benefits from bagging but there is less improvement. We compare the forecast performances of the log-linear model and the neural network specification and find that bagging eliminates any advantage of the nonlinear model.

We adopt the following strategy to compute one-day-ahead forecasts. The models are initially estimated using data from the years 1995 to 1996 only. The forecast period starts on 3-Jan-1997. The parameter estimates from the period 1995–1996 are then used to compute one-day-ahead forecasts for the whole year 1997. Then, the model is re-estimated on the data set 1995–1997 and the estimates are used to compute one-day ahead forecasts for the year 1998, and so on. In the last iteration, we estimate the model on the span 1995–2004 and use the parameters to generate forecasts for 2005. In the bagging procedure using the block-bootstrap of Künsch (1989), we use $B = 200$ replications and a block size of $m = T^{1/3}$ (Hall, Horowitz, and Jing 1995), such that the block size increases with each year.

Tables 2 and 3 show one-step-ahead forecasting results for the log-linear and for the nonlinear specification, respectively. The reported numbers are the fractions of the root mean squared error (RMSE) of the considered model with bagging over the RMSE of

the considered model without bagging. Therefore, a number less than one indicates that bagging reduced the RMSE for that stock and year.

The numbers in parentheses report the p-value of the test for conditional predictive ability of Giacomini and White (2006). The null hypothesis is that the expected loss of the forecast (here RMSE) is the same for both models; a number less than α indicates that the null hypothesis is rejected at significance level α. Thus, a rejection only tells us that one of the models has better forecast performance than the other. Exactly which one of the two performs better has to be read from the RMSE. It is important to notice that forecasting equality can can be rejected by the test even when the ratio is equal to one in the tables. This is caused by rounding effects. For these cases we put a "-" or a "+" meaning that the ratios are, in fact, bellow or above 1, respectively.

Table 2 shows that bagging leads to substantial reductions in forecast RMSE for the log-linear model. The improvements are most pronounced in the 1990s and decline in 2004 and 2005. For the years 1997, 1998, and on the whole forecast sample 1997–2005 the Giacomini and White test rejects the null hypothesis for all series at all common significance levels. We can conclude that bagging significantly improves forecast accuracy.

Table 3 shows the effect of bagging on forecasts made with the neural network specification. While the evidence from the RMSE fractions is more mixed across years and stocks, the Giacomini and White test indicates significant improvements similar to the log-linear specification. For some stocks, the nonlinear model benefits from bagging across all years, such as AIG, GE, HD, MCD, MO, and MRK. The only stock for which the nonlinear specification does not improve by bagging on the total sample is WMT.

Table 4 compares the RMSE of the forecast made with the neural network specification (numerator) to the one by the log-linear model (denominator). Therefore, a number less than one indicates that the nonlinear model outperforms the log-linear model. The years 1997 through 2000 show forecast improvements over the log-linear model for most of the stocks while the years 2004 and 2005 show no improvement at all. Some stock volatility forecasts benefit from nonlinear modeling fairly consistently, such as AA, BA, HON, and UTX. A possible reason for the better performance of the nonlinear model in the late 1990s is that the models are initially estimated on the years 1995 and 1996, which still fall into the low volatility regime that lasted through the mid-1990s. Then, the first year of the forecast period 1997 experiences with the Asian crisis the onset of a high volatility regime that lasts through 2003. Thus, the early estimates of the log-linear model may

be less prepared to forecast the nonlinear phenomenon of the transition into the high volatility regime.

Table 5 makes the same comparison as Table 4 but in a bagging environment for both models. It is apparent that bagging eliminates any advantage that nonlinear modeling displayed in Table 4. The Giacomini and White test in fact indicates that the neural network specification performs significantly worse than the log-linear model under bagging in many cases. A possible reason why the log-linear model catches up with and even outperforms the nonlinear model once it is bagged may be that bagging expands the shrinkage representation of the conditional expectation that is the forecast (Stock and Watson 2005). The shrinkage representation can be seen as an alternative approximation to Equation (7). In summary, bagging the log-linear specification yields the best forecast results on our sample.

6. CONCLUSION

In this paper, we have studied how bagging can improve the out-of-sample accuracy of forecast models for realized volatility. We consider 23 stocks over the period 1995 through 2005. Specifying a log-linear and a nonlinear forecast model for the realized kernel estimator of integrated volatility, we find that bagging reduces the prediction mean squared error for both model types. The improvement is more pronounced for the log-linear model and for forecasts in the 1990s. Bagging is found to eliminate the advantage of the nonlinear specification over the log-linear specification on our sample. Bagging the log-linear model shows the best forecast performance on our sample.

REFERENCES

ANDERS, U., AND O. KORN (1999): "Model Selection in Neural Networks," *Neural Networks*, 12, 309–323.

ANDERSEN, T., AND T. BOLLERSLEV (1998): "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885–906.

ANDERSEN, T., T. BOLLERSLEV, AND F. DIEBOLD (2007): "Roughing it up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility," *Review of Economics and Statistics*, 89, 701–720.

ANDERSEN, T., T. BOLLERSLEV, F. X. DIEBOLD, AND H. EBENS (2001): "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, 61, 43–76.

ANDERSEN, T., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2001): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.

——— (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

BAILLIE, R., T. BOLLERSLEV, AND H. MIKKELSEN (1996): "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 74, 3–30.

BARNDORFF-NIELSEN, O., P. HANSEN, A. LUNDE, AND N. SHEPHARD (in press): "Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise," *Econometrica*.

BARNDORFF-NIELSEN, O., AND N. SHEPHARD (2002): "Econometric analysis of realised volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society B*, 64, 253–280.

——— (2006): "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation," *Journal of Financial Econometrics*, 4, 1–30.

BÜHLMANN, P., AND B. YU (2002): "Analyzing Bagging," *Annals of Statistics*, 30, 927–961.

BIAIS, B., L. GLOSTEN, AND C. SPATT (2005): "Market Microstructure: A Survey of Microfoundations, Empirical Results, and Policy Implications," *Journal of Financial Markets*, 8, 217–264.

BOLLERSLEV, T. (1987): "A conditionally heteroskedastic time series model for speculative prices and rates of return," *The Review of Economics and Statistics*, 69, 542–547.

BREIMAN, L. (1996): "Bagging Predictors," *Machine Learning*, 36, 105–139.

CAMPBELL, J., A. LO, AND A. MACKINLAY (1997): *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ.

CHEN, X., J. RACINE, AND N. R. SWANSON (2001): "Semiparametric ARX neural-network models with an application to forecasting inflation," *IEEE Transactions on Neural Networks*, 12, 674–683.

CHEN, X., AND X. SHEN (1998): "Sieve extremum estimates for weakly dependent data," *Econometrica*, 66, 289–314.

CHEN, X., AND H. WHITE (1998): "Improved rates and asymptotic normality for nonparametric neural network estimators," *IEEE Transactions on Information Theory*, 18, 17–39.

CHRISTENSEN, K., R. OOMEN, AND M. PODOLSKIJ (2008): "Realised Quantile-Based Estimation of the Integrated Variance," Working paper, University of Aarhus, Denmark.

CORSI, F. (2004): "A Simple Long Memory Model of Realized Volatility," Manuscript, University of Southern Switzerland.

CORSI, F., S. MITTNIK, C. PIGORSCH, AND U. PIGORSCH (2008): "The volatility of realized volatility," *Econometric Reviews*, 27, 46–78.

CRAIOVEANU, M., AND E. HILLEBRAND (2008): "Models for Daily Realized Stock Volatility Time Series," Working paper, Louisiana State University.

DING, Z., C. W. J. GRANGER, AND R. F. ENGLE (1993): "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83–106.

DONALDSON, R. G., AND M. KAMSTRA (1997): "An artificial neural network-GARCH model for international stock return volatility," *Journal of Empirical Finance*, 4(1), 17–46.

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflations," *Econometrica*, 50, 987–1007.

ENGLE, R. F., AND T. BOLLERSLEV (1986): "Modelling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 1–50.

ENGLE, R. F., AND V. K. NG (1993): "Measuring and Testing the Impact of News on Volatility," *Journal of Finance*, 48, 1749–1778.

FERNANDES, M., M. MEDEIROS, AND M. SCHARTH (2007): "Modeling and Predicting the CBOE Market Volatility Index," Discussion paper, Pontifical Catholic University of Rio de Janeiro.

FORESEE, F. D., AND M. T. HAGAN (1997): "Gauss-Newton Approximation to Bayesian Regularization," in *IEEE International Conference on Neural Networks (Vol. 3)*, pp. 1930–1935, New York. IEEE.

GIACOMINI, R., AND H. WHITE (2006): "Tests of Conditional Predictive Ability," *Econometrica*, 74, 1545–1578.

GLOSTEN, L., R. JAGANNATHAN, AND R. RUNKLE (1993): "On The Relationship Between The Expected Value and The Volatility of The Nominal Excess Returns on Stocks," *Journal of Finance*, 48, 1779–1801.

GRENANDER, U. (1981): *Abstract Inference*. Wiley Series, New York.

HALL, P., J. HOROWITZ, AND B.-Y. JING (1995): "On Blocking Rules for the Bootstrap with Dependent Data," *Biometrika*, 82, 561–574.

HAMID, S. A., AND Z. IQBAL (2004): "Using neural networks for forecasting volatility of S&P 500 Index futures prices," *Journal of Business Research*, 57(1), 1116–1125.

HANSEN, P., J. LARGE, AND A. LUNDE (2008): "Moving average-based estimators of integrated variance," *Econometric Reviews*, 27, 79–111.

HILLEBRAND, E., AND M. MEDEIROS (to appear): "Estimating and Forecasting GARCH Models in The Presence of Structural Breaks and Regime Switches," in *Forecasting in The Presence of Structural Breaks and Model Uncertainty*, ed. by M. Wohar, and D. Rapach. Elsevier.

HILLEBRAND, E., AND M. C. MEDEIROS (2008): "Asymmetries, Breaks, and Long-Range Dependence: An Estimation Framework for Time Series of Daily Realized Volatility," Working paper, Louisiana State University and Pontifical Catholic University Rio de Janeiro.

HORNIK, K., M. STINCHCOMBE, H. WHITE, AND P. AUER (1994): "Degree of Approximation Results for Feedforward Networks Approximating Unknown Mappings and Their Derivatives," *Neural Computation*, 6, 1262–1274.

HORNIK, K., M. STINCHOMBE, AND H. WHITE (1989): "Multi-Layer Feedforward Networks are Universal Approximators," *Neural Networks*, 2, 359–366.

HU, M. Y., AND C. TSOUKALAS (1999): "Combining conditional volatility forecasts using neural networks: an application to the EMS exchange rates," *Journal of International Financial Markets, Institutions and Money*, 9(4), 407–422.

HUANG, H., AND T.-H. LEE (2007a): "Forecasting Using High-Frequency Financial Time Series," Working paper, University of California at Riverside.

——— (2007b): "To Combine Forecasts or to Combine Information," Working paper, University of California at Riverside.

INOUE, A., AND L. KILIAN (2004): "Bagging Time Series Models," Discussion paper no. 4333, Centre for Economic Policy Research (CEPR).

——— (in press): "How Useful is Bagging in Forecasting Economic Time Series? A Case Study of U.S. CPI Inflation," *Journal of the American Statistical Association*.

KÜNSCH, H. (1989): "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics*, 17, 1217–1241.

KUAN, C. M., AND H. WHITE (1994): "Artificial Neural Networks: An Econometric Perspective," *Econometric Reviews*, 13, 1–91.

LEE, T.-H., AND Y. YANG (2006): "Bagging Binary and Quantile Predictors for Time Series," *Journal of Econometrics*, 135, 465–497.

MACKAY, D. J. C. (1992): "A practical Bayesian framework for Backpropagation Networks," *Neural Computation*, 4, 448–472.

MADHAVAN, A. (2000): "Market Microstructure: A Survey," *Journal of Financial Markets*, 3, 205–258.

MARTENS, M., D. VAN DIJK, AND M. DE POOTER (2004): "Modeling and Forecasting S&P 500 Volatility: Long Memory, Structural Breaks and Nonlinearity," Discussion Paper 04-067/4, Tinbergen Institute.

MCALEER, M., AND M. MEDEIROS (2006): "A Multiple Regime Smooth Transition Heterogenous Autoregressive Model for Long Memory and Asymmetries," *Journal of Econometrics*, to appear.

MEDEIROS, M., A. VEIGA, AND C. PEDREIRA (2001): "Modelling Exchange Rates: Smooth Transitions, Neural Networks, and Linear Models," *IEEE Transactions on Neural Networks*, 12, 755–764.

MEDEIROS, M. C., T. TERÄSVIRTA, AND G. RECH (2006): "Building Neural Network Models for Time Series: A Statistical Approach," *Journal of Forecasting*, pp. 49–75.

MEDEIROS, M. C., AND A. VEIGA (2000): "A Hybrid Linear-Neural Model for Time Series Forecasting," *IEEE Transactions on Neural Networks*, 11, 1402–1412.

MÜLLER, U., M. DACOROGNA, R. DAVE, R. OLSEN, O. PICTET, AND J. VON WEIZSAECKER (1997): "Volatilities of different time resolutions – Analyzing the dynamics of market components," *Journal of Empirical Finance*, 4, 213–239.

NELSON, D. B. (1991): "Conditinal Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

NGUYEN, D., AND B. WIDROW (1990): "Improving the Learning Speed of 2-Layer Neural Networks by Choosing Initial Values of the Adaptive Weights," in *Proceedings of the International Joint Conference on Neural Networks*, vol. 3, pp. 21–26.

RAPACH, D., AND J. STRAUSS (2007): "Bagging or Combining (or Both)? An Analysis Based on Forecasting U.S. Employment Growth," Working paper, Saint Louis University.

REED, R. (1993): "Prunning Algorithms – A Survey," *IEEE Transactions on Neural Networks*, 4, 740–747.

SCHARTH, M., AND M. MEDEIROS (2006): "Asymmetric Effects and Long Memory in the Volatility of DJIA Stocks," Discussion Paper 532, Pontifical Catholic University of Rio de Janeiro.

STOCK, J., AND M. WATSON (2005): "An Empirical Comparison of Methods for Forecasting Using Many Predictors," Working paper, Harvard University.

TERÄSVIRTA, T., D. VAN DIJK, AND M. MEDEIROS (2005): "Linear Models, Smooth Transition Autoregressions, and Neural Networks for Forecasting Macroeconomic Time Series: A Re-Examination," *International Journal of Forecasting*, 21, 755–774.

TRAPLETTI, A., F. LEISCH, AND K. HORNIK (2000): "Stationary and Integrated Autoregressive Neural Network Processes," *Neural Computation*, 12, 2427–2450.

VAN DIJK, D., T. TERÄSVIRTA, AND P. H. FRANSES (2002): "Smooth Transition Autoregressive Models - A Survey of Recent Developments," *Econometric Reviews*, 21, 1–47.

ZHANG, L., P. MYKLAND, AND Y. AÏT-SAHALIA (2005): "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411.

7. APPENDIX

Let $\mathbf{D} := (\mathbf{y}, \mathbf{Z})$ represent the data set, where $\mathbf{y} = (\log(RV_1), \ldots, \log(RV_T))'$ and $\mathbf{Z} = (\mathbf{z}_1', \ldots, \mathbf{z}_T')'$. $\mathcal{M}$ is a particular neural network model. After the data are obtained, the density function for the parameters is updated according to Bayes' rule

$$\mathbb{P}(\boldsymbol{\psi}|\mathbf{D}, \eta, \gamma, \mathcal{M}) = \frac{\mathbb{P}(\mathbf{D}|\boldsymbol{\psi}, \eta, \mathcal{M})\,\mathbb{P}(\boldsymbol{\psi}|\gamma, \mathcal{M})}{\mathbb{P}(\mathbf{D}|\eta, \gamma, \mathcal{M})}, \tag{12}$$

where $\mathbb{P}(\boldsymbol{\psi}|\gamma, \mathcal{M})$ is the prior density, which represents our knowledge of the parameters before any data is collected, $\mathbb{P}(\mathbf{D}|\boldsymbol{\psi}, \eta, \mathcal{M})$ is the likelihood function, which is the probability of the data occurring given the parameters, and $\mathbb{P}(\mathbf{D}|\eta, \gamma, \mathcal{M})$ is a normalization factor, which guarantees that the total probability is equal to one.

If the distribution of ε_t and the prior distribution for the parameters are both Gaussian, then $\mathbb{P}(\mathbf{D}|\boldsymbol{\psi}, \eta, \mathcal{M})$ and $\mathbb{P}(\boldsymbol{\psi}|\eta, \mathcal{M})$ are written as

$$\mathbb{P}(\mathbf{D}|\boldsymbol{\psi}, \eta, \mathcal{M}) = \left(\frac{\pi}{\eta}\right)^{-\frac{T}{2}} \exp\left[-\eta \mathcal{Q}_T(\boldsymbol{\psi})\right] \tag{13}$$

and

$$\mathbb{P}(\boldsymbol{\psi}|\eta, \mathcal{M}) = \left(\frac{\pi}{\gamma}\right)^{-\frac{L}{2}} \exp\left[-\gamma \mathcal{Q}_T^*(\boldsymbol{\psi})\right], \tag{14}$$

where L is the total number of parameters in the NN model. Substituting (14) and (13) into (12), we get

$$\begin{aligned} \mathbb{P}(\boldsymbol{\psi}|\mathbf{D}, \eta, \gamma, \mathcal{M}) &= \frac{\left(\frac{\pi}{\eta}\right)^{-\frac{T}{2}} \left(\frac{\pi}{\gamma}\right)^{-\frac{L}{2}} \exp\left\{-\left[\eta \mathcal{Q}_T(\boldsymbol{\psi}) + \gamma \mathcal{Q}_T^*(\boldsymbol{\psi})\right]\right\}}{\text{Normalization Factor}} \\ &= Z(\eta, \gamma) \exp\left[-\widetilde{\mathcal{Q}}_T(\boldsymbol{\psi})\right]. \end{aligned} \tag{15}$$

In this Bayesian framework, the optimal parameters should maximize the posterior probability $\mathbb{P}(\boldsymbol{\psi}|\mathbf{D}, \eta, \gamma, \mathcal{M})$, which is equivalent to minimizing the regularized objective function given in (10). The regularization parameters are optimized by applying Bayes' rule

$$\mathbb{P}(\eta, \gamma|\mathbf{D}, \mathcal{M}) = \frac{\mathbb{P}(\mathbf{D}|\eta, \gamma, \mathcal{M})\,\mathbb{P}(\eta, \gamma|\mathcal{M})}{\mathbb{P}(\mathbf{D}|\mathcal{M})}. \tag{16}$$

Assuming a uniform prior density $\mathbb{P}(\eta, \gamma|\mathcal{M})$ for the regularization parameters, maximizing the posterior is achieved by maximizing the likelihood function $\mathbb{P}(\mathbf{D}|\boldsymbol{\psi}, \eta, \mathcal{M})$.

Since all probabilities have a Gaussian form, the normalization factor is expressed as

$$\mathbb{P}\left(\mathbf{D}|\eta,\gamma,\mathcal{M}\right)=\frac{\mathbb{P}\left(\mathbf{D}|\boldsymbol{\psi},\eta,\mathcal{M}\right)\mathbb{P}\left(\boldsymbol{\psi}|\gamma,\mathcal{M}\right)}{\mathbb{P}\left(\boldsymbol{\psi}|\mathbf{D},\eta,\gamma,\mathcal{M}\right)}=\left(\frac{\pi}{\eta}\right)^{-\frac{T}{2}}\left(\frac{\pi}{\gamma}\right)^{-\frac{L}{2}}Z^{-1}(\eta,\gamma). \tag{17}$$

Since the objective function is quadratic in a small area surrounding a minimum point, we can expand $\tilde{Q}_N(\boldsymbol{\psi})$ in a Taylor series around the minimum point of the posterior density, where the gradient is zero. Solving for the normalizing constant yields

$$Z(\eta,\gamma)=(2\pi)^{\frac{L}{2}}\left[\det\left(\mathbf{H}\right)\right]^{-1/2}\exp\left[-\widetilde{\mathcal{Q}}_T(\boldsymbol{\psi})\right], \tag{18}$$

where $\mathbf{H}$ is the Hessian matrix of the objective function. Inserting (17) into (16) we solve for the optimal values for η and γ at the minimum point. We do this by computing the derivatives of the log of (17) with respect to γ and η and setting them equal to zero. This yields

$$\widehat{\gamma}=\frac{\xi}{2\mathcal{Q}^*(\boldsymbol{\psi})} \tag{19}$$

and

$$\widehat{\eta}=\frac{T-\xi}{2\mathcal{Q}_T(\boldsymbol{\psi})}, \tag{20}$$

where $\xi=L-2\gamma\text{trace}(\mathbf{H})^{-1}$ is called the effective number of parameters.

Following Foresee and Hagan (1997), the steps required for Bayesian optimization of the regularization parameters with the Gauss-Newton approximation to the Hessian matrix are:

(1) Initialize η, γ, and the network parameters by the Nguyen-Widrow rule (Nguyen and Widrow 1990). After the first estimation step, the objective function parameters recover from the initial setting.
(2) Take one step of the Levenberg-Marquardt algorithm to minimize the objective function $\tilde{Q}_N(\boldsymbol{\psi})$.
(3) Compute the effective number of parameters $\kappa = L - 2\gamma\,\text{trace}(\mathbf{H})^{-1}$ making use of the Gauss-Newton approximation to the Hessian matrix available in the Levenberg-Marquardt optimization algorithm: $\mathbf{H}=\nabla^2\tilde{Q}_N(\boldsymbol{\psi})\approx 2\gamma\mathbf{J}'\mathbf{J}+2\eta\mathbf{I}_T$, where $\mathbf{J}$ is the Jacobian matrix of the estimation set errors.
(4) Compute new estimates for the objective function parameters.
(5) Now iterate steps 1 through 3 until convergence.

TABLE 1. DATA DESCRIPTION.

The first two columns display the symbols and names of the stocks considered in the empirical investigation. The third column gives the average number of transactions per day. The number of days is 2771.

Symbol	Stock	Transactions per day
AA	Alcoa Inc.	2055
AIG	American International Group Inc.	2979
AXP	American Express Co.	2599
BA	Boeing Co.	3006
CAT	Caterpillar Inc.	3597
DD	Du Pont de Nemours & Co.	2587
DIS	Walt Disney Co.	3839
GE	General Electric Co.	8072
GM	General Motors Corp.	2945
HD	Home Depot Inc.	4758
HON	Honeywell International Inc.	1888
IBM	International Business Machines Corp.	5117
JNJ	Johnson & Johnson	3551
JPM	JPMorgan Chase & Co.	3400
KO	Coca-Cola Co.	3302
MCD	McDonald's Corp.	2720
MMM	3M Co.	2183
MO	Altria Group Inc.	4031
MRK	Merck & Co. Inc.	4353
PFE	Pfizer Inc.	7029
PG	Procter and Gamble Co.	3062
UTX	United Technologies Corp.	1834
WMT	Wal-Mart Stores Inc.	4797

TABLE 2. BAGGING THE LOG-LINEAR MODEL.

The table reports the ratio of the root mean squared errors from the log-linear model for one-day ahead forecasts between 1997 and 2005 with and without bagging. Numbers below one indicate that bagging improves the forecast performance. The numbers in parentheses report the p-value of Giacomini and White's (2006) test of conditional predictive ability.

Series	1997	1998	1999	2000	2001	2002	2003	2004	2005	1997–2005
AA	0.79	0.83	0.91	0.87	0.86	0.96	0.97	0.98	1.00	0.88
	(0.00)	(0.00)	(0.17)	(0.14)	(0.00)	(0.38)	(0.25)	(0.35)	(0.37)	(0.00)
AIG	0.81	0.82	0.83	0.89	0.99	0.96	0.99	0.99	0.98	0.91
	(0.00)	(0.00)	(0.00)	(0.00)	(0.35)	(0.04)	(0.57)	(0.18)	(0.04)	(0.00)
AXP	0.80	0.85	0.85	0.89	0.96	0.98	0.97	1.00	1.01	0.91
	(0.00)	(0.00)	(0.01)	(0.00)	(0.17)	(0.25)	(0.00)	(0.54)	(0.44)	(0.00)
BA	0.86	0.86	0.92	0.91	0.93	0.91	0.93	0.99	0.98	0.91
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.20)	(0.20)	(0.00)
CAT	0.86	0.85	0.89	0.90	0.97	0.94	0.96	0.97	0.99	0.91
	(0.00)	(0.00)	(0.00)	(0.00)	(0.07)	(0.00)	(0.01)	(0.02)	(0.77)	(0.00)
DD	0.73	0.77	0.88	0.92	0.95	0.95	0.97	0.99	0.99	0.87
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.03)	(0.61)	(0.66)	(0.00)
DIS	0.80	0.85	0.91	0.91	0.96	0.92	0.96	0.99	1.00	0.91
	(0.00)	(0.00)	(0.01)	(0.00)	(0.04)	(0.00)	(0.01)	(0.83)	(0.78)	(0.00)
GE	0.90	0.90	0.90	0.92	0.97	0.93	0.97	0.97	0.99	0.94
	(0.00)	(0.00)	(0.00)	(0.00)	(0.65)	(0.00)	(0.01)	(0.01)	(0.17)	(0.00)
GM	0.88	0.88	0.93	0.90	0.92	0.93	0.92	0.99	0.98	0.92
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.37)	(0.06)	(0.00)
HD	0.85	0.89	0.89	0.89	0.92	0.96	0.97	0.99	1.00	0.91
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.12)	(0.30)	(0.11)	(0.00)
HON	0.84	0.85	0.85	0.84	0.90	0.96	0.94	0.99	0.99	0.89
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.05)	(0.00)	(0.60)	(0.03)	(0.00)
IBM	0.86	0.94	0.92	0.93	0.97	0.95	0.96	1.00_	0.99	0.95
	(0.00)	(0.01)	(0.00)	(0.02)	(0.32)	(0.05)	(0.36)	(0.01)	(0.60)	(0.00)
JNJ	0.88	0.88	0.90	0.94	0.97	0.98	0.96	0.99	0.98	0.94
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.38)	(0.00)	(0.58)	(0.79)	(0.00)
JPM	0.79	0.72	0.87	0.89	0.93	0.93	0.96	0.99	0.99	0.87
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.77)	(0.72)	(0.00)
KO	0.80	0.88	0.93	0.91	0.96	0.95	0.98	1.00	0.99	0.92
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.19)	(0.86)	(0.82)	(0.00)
MCD	0.93	0.88	0.88	0.92	0.95	0.93	0.96	0.99	0.98	0.93
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.38)	(0.06)	(0.00)
MMM	0.82	0.88	0.89	0.91	0.91	0.97	0.96	0.98	0.99	0.90
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.15)	(0.00)	(0.02)	(0.34)	(0.00)
MO	0.80	0.92	0.95	0.91	0.92	0.97	0.98	0.98	0.99	0.93
	(0.00)	(0.00)	(0.05)	(0.00)	(0.00)	(0.04)	(0.04)	(0.04)	(0.36)	(0.00)
MRK	0.80	0.91	0.92	0.92	0.97	0.95	0.97	0.99	0.98	0.93
	(0.00)	(0.00)	(0.00)	(0.00)	(0.05)	(0.00)	(0.05)	(0.79)	(0.03)	(0.00)
PFE	0.80	0.90	0.93	0.92	0.98	0.98	0.98	0.98	0.98	0.93
	(0.00)	(0.00)	(0.02)	(0.00)	(0.13)	(0.33)	(0.07)	(0.13)	(0.09)	(0.00)
PG	0.76	0.83	0.92	0.91	0.98	0.94	1.00	0.99	0.98	0.91
	(0.00)	(0.00)	(0.00)	(0.00)	(0.05)	(0.00)	(0.41)	(0.43)	(0.01)	(0.00)
UTX	0.75	0.77	0.88	0.86	0.90	0.92	0.95	0.98	0.96	0.86
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.08)	(0.00)	(0.00)
WMT	0.84	0.82	0.84	0.82	0.93	0.98	0.97	0.99	0.99	0.89
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.05)	(0.11)	(0.67)	(0.03)	(0.00)

TABLE 3. BAGGING THE NEURAL NETWORK MODEL.

The table reports the ratio of the root mean squared errors from the nonlinear model for one-day ahead forecasts between 1997 and 2005 with and without bagging. Numbers below one indicate that bagging improves the forecast performance. The numbers in parentheses report the p-value of Giacomini and White's (2006) test of conditional predictive ability.

Series	1997	1998	1999	2000	2001	2002	2003	2004	2005	1997–2005
AA	1.02	0.90	0.93	0.96	0.89	0.99	0.96	0.99	1.00	0.92
	(0.09)	(0.00)	(0.00)	(0.00)	(0.00)	(0.66)	(0.10)	(0.59)	(0.63)	(0.00)
AIG	0.97	0.87	0.88	0.93	0.98	0.96	0.99	0.98	0.97	0.92
	(0.00)	(0.00)	(0.00)	(0.01)	(0.07)	(0.03)	(0.56)	(0.32)	(0.14)	(0.00)
AXP	0.96	0.93	0.89	0.93	0.95	1.00	0.98	0.99	1.00	0.93
	(0.00)	(0.00)	(0.00)	(0.11)	(0.01)	(0.99)	(0.40)	(0.06)	(0.66)	(0.00)
BA	0.93	0.88	0.93	0.92	0.98	0.91	0.96	1.01	1.00	0.92
	(0.00)	(0.00)	(0.00)	(0.00)	(0.21)	(0.00)	(0.09)	(0.81)	(0.92)	(0.00)
CAT	0.95	0.88	0.94	0.94	0.96	0.95	0.97	0.99	1.00	0.90
	(0.00)	(0.00)	(0.00)	(0.01)	(0.02)	(0.00)	(0.05)	(0.94)	(0.85)	(0.00)
DD	0.98	0.84	0.92	0.94	0.95	0.98	0.99	1.00	0.99	0.93
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.21)	(0.85)	(0.88)	(0.76)	(0.00)
DIS	0.96	0.89	0.94	0.93	0.96	0.95	0.96	1.01	0.99	0.93
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.19)	(0.03)	(0.10)	(0.92)	(0.00)
GE	0.97	0.86	0.93	0.96	0.96	0.96	0.96	0.98	0.99	0.95
	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)	(0.04)	(0.69)	(0.80)	(0.48)	(0.00)
GM	1.00_{-}	0.85	0.98	0.95	0.94	0.96	0.94	0.98	0.98	0.92
	(0.06)	(0.00)	(0.09)	(0.00)	(0.00)	(0.01)	(0.00)	(0.13)	(0.03)	(0.00)
HD	0.94	0.91	0.94	0.95	0.95	0.99	0.97	0.99	0.99	0.95
	(0.00)	(0.00)	(0.01)	(0.15)	(0.02)	(0.06)	(0.07)	(0.93)	(0.54)	(0.00)
HON	1.03	1.00_{-}	0.98	0.90	0.94	0.99	0.96	1.00	1.01	0.92
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.07)	(0.79)	(0.16)	(0.00)
IBM	0.91	1.00	0.92	0.96	0.98	0.98	0.98	1.01	0.99	0.96
	(0.00)	(0.39)	(0.01)	(0.25)	(0.58)	(0.34)	(0.16)	(0.02)	(0.23)	(0.00)
JNJ	1.02	0.91	0.94	0.97	0.99	0.98	0.98	0.99	0.99	0.95
	(0.00)	(0.00)	(0.00)	(0.07)	(0.48)	(0.12)	(0.26)	(0.01)	(0.25)	(0.00)
JPM	0.92	0.86	0.91	0.89	0.91	0.95	0.96	1.00	0.97	0.91
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.56)	(0.08)	(0.34)	(0.43)	(0.00)
KO	0.91	0.93	0.93	0.90	0.98	0.97	0.97	1.00_{+}	0.99	0.94
	(0.00)	(0.00)	(0.00)	(0.00)	(0.56)	(0.01)	(0.52)	(0.06)	(0.76)	(0.00)
MCD	0.92	0.92	0.95	0.96	0.95	0.94	0.99	0.99	0.99	0.94
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.68)	(0.92)	(0.06)	(0.00)
MMM	0.96	0.91	0.95	0.93	0.95	0.95	0.98	0.99	1.00	0.95
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.13)	(0.70)	(0.15)	(0.00)
MO	0.89	0.95	0.98	0.94	0.96	0.97	0.99	0.99	0.99	0.92
	(0.00)	(0.00)	(0.95)	(0.00)	(0.05)	(0.05)	(0.65)	(0.98)	(0.42)	(0.00)
MRK	0.99	0.91	0.96	0.93	0.99	0.96	0.97	0.99	0.99	0.95
	(0.00)	(0.00)	(0.00)	(0.00)	(0.48)	(0.01)	(0.35)	(0.65)	(0.43)	(0.00)
PFE	0.89	0.87	0.97	0.97	0.98	1.03	0.99	1.02	0.99	0.93
	(0.00)	(0.00)	(0.04)	(0.01)	(0.11)	(0.15)	(0.17)	(0.01)	(0.05)	(0.00)
PG	0.94	0.90	0.93	0.91	0.98	0.94	0.97	0.98	1.00	0.93
	(0.00)	(0.00)	(0.00)	(0.00)	(0.14)	(0.00)	(0.28)	(0.13)	(0.97)	(0.00)
UTX	0.95	0.92	0.92	0.95	0.96	0.94	0.96	1.00_{-}	0.97	0.95
	(0.00)	(0.00)	(0.00)	(0.28)	(0.04)	(0.01)	(0.02)	(0.02)	(0.07)	(0.00)
WMT	1.00	1.06	0.94	0.90	0.98	1.01	1.04	1.05	0.99	1.01
	(0.35)	(0.00)	(0.00)	(0.00)	(0.19)	(0.48)	(0.00)	(0.00)	(0.00)	(0.00)

TABLE 4. NONLINEAR VS. LOG-LINEAR MODEL WITHOUT BAGGING.

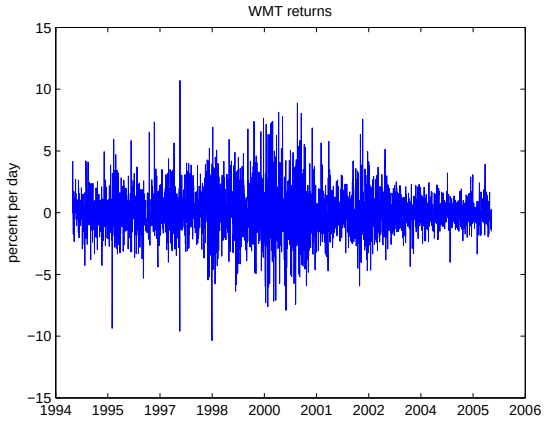
The table reports the ratio of the root mean squared errors from the nonlinear and log-linear models for one-day-ahead forecasts between 1997 and 2005 without bagging. Numbers below one indicate that nonlinearity improves the forecast performance. The numbers in parentheses report the p-value of Giacomini and White's (2006) test of conditional predictive ability.

Series	1997	1998	1999	2000	2001	2002	2003	2004	2005	1997–2005
AA	0.74	0.93	0.97	0.95	0.97	0.98	0.99	0.99	1.00	0.97
	(0.00)	(0.00)	(0.00)	(0.00)	(0.25)	(0.38)	(0.72)	(0.70)	(0.63)	(0.00)
AIG	0.84	1.01	0.96	0.98	1.02	1.01	0.99	1.01	1.01	1.04
	(0.07)	(0.00)	(0.04)	(0.42)	(0.20)	(0.01)	(0.63)	(0.43)	(0.37)	(0.00)
AXP	0.85	1.05	0.97	0.95	1.03	0.98	0.98	0.99	1.00	1.04
	(0.00)	(0.00)	(0.48)	(0.84)	(0.13)	(0.44)	(0.28)	(0.62)	(0.78)	(0.00)
BA	0.94	0.97	1.00_-	0.98	0.98	0.99	0.97	0.98	0.98	1.02
	(0.05)	(0.05)	(0.03)	(0.00)	(0.00)	(0.75)	(0.04)	(0.32)	(0.47)	(0.00)
CAT	0.95	0.97	0.95	0.97	1.03	0.99	0.99	1.01	1.00	1.05
	(0.06)	(0.00)	(0.09)	(0.48)	(0.10)	(0.15)	(0.23)	(0.27)	(0.54)	(0.00)
DD	0.73	0.89	0.95	0.99	0.99	0.99	0.97	1.00	1.00	0.98
	(0.01)	(0.37)	(0.03)	(0.01)	(0.81)	(0.57)	(0.02)	(0.93)	(0.58)	(0.00)
DIS	0.82	1.00_-	0.99	0.96	1.01	0.97	1.00	1.00	1.00	1.00_-
	(0.01)	(0.00)	(0.11)	(0.03)	(0.06)	(0.20)	(0.92)	(0.82)	(0.53)	(0.00)
GE	0.97	1.09	0.98	0.97	1.01	1.00	1.01	0.98	0.99	1.04
	(0.06)	(0.00)	(0.36)	(0.78)	(0.39)	(0.56)	(0.31)	(0.58)	(0.42)	(0.00)
GM	0.88	1.12	0.99	0.96	0.98	0.99	0.99	1.01	1.01	1.03
	(0.00)	(0.00)	(0.20)	(0.04)	(0.17)	(0.94)	(0.66)	(0.29)	(0.59)	(0.00)
HD	0.89	1.05	0.95	0.95	0.99	1.00	1.00	0.99	1.00	1.00_-
	(0.00)	(0.00)	(0.11)	(0.26)	(0.69)	(0.74)	(0.92)	(0.93)	(0.35)	(0.00)
HON	0.82	0.87	0.88	0.94	0.97	0.97	0.99	0.99	1.00	1.00_+
	(0.00)	(0.07)	(0.00)	(0.00)	(0.06)	(0.14)	(0.75)	(0.10)	(0.81)	(0.00)
IBM	0.95	0.99	1.01	0.99	1.00	0.97	0.99	0.98	1.01	1.02
	(0.03)	(0.13)	(0.07)	(0.18)	(0.89)	(0.26)	(0.87)	(0.12)	(0.34)	(0.01)
JNJ	0.86	0.96	0.96	0.95	0.97	1.00	1.00	1.00	1.00	1.00_+
	(0.16)	(0.10)	(0.04)	(0.10)	(0.01)	(0.20)	(0.58)	(0.14)	(0.19)	(0.07)
JPM	0.86	0.85	0.97	0.99	1.02	0.96	0.98	1.00	1.00	0.98
	(0.01)	(0.70)	(0.39)	(0.12)	(0.24)	(0.05)	(0.36)	(0.72)	(0.59)	(0.01)
KO	0.88	0.97	1.00	0.98	0.99	0.99	0.99	1.01	1.00	0.99
	(0.04)	(0.19)	(0.15)	(0.60)	(0.08)	(0.53)	(0.60)	(0.05)	(0.23)	(0.01)
MCD	1.01	1.00	0.90	0.94	1.00	0.98	0.98	1.00	0.99	1.00
	(0.03)	(0.49)	(0.07)	(0.06)	(0.96)	(0.11)	(0.04)	(0.34)	(0.71)	(0.08)
MMM	0.86	0.99	0.92	0.98	0.95	1.01	0.98	0.98	0.99	0.97
	(0.00)	(0.00)	(0.36)	(0.28)	(0.01)	(0.18)	(0.16)	(0.10)	(0.41)	(0.00)
MO	0.89	0.97	0.96	0.96	0.96	1.00	0.99	1.00	1.00	1.04
	(0.00)	(0.71)	(0.56)	(0.63)	(0.01)	(0.24)	(0.44)	(0.82)	(0.18)	(0.00)
MRK	0.80	1.02	0.94	0.99	1.00	1.00_+	1.02	1.01	0.98	1.00_+
	(0.08)	(0.03)	(0.11)	(0.07)	(0.92)	(0.07)	(0.06)	(0.07)	(0.12)	(0.00)
PFE	0.90	1.04	1.01	0.98	1.01	1.03	1.01	0.97	0.99	1.03
	(0.00)	(0.00)	(0.02)	(0.56)	(0.81)	(0.13)	(0.42)	(0.06)	(0.44)	(0.00)
PG	0.80	0.92	0.97	1.00	1.00	1.00	1.02	1.00	0.98	0.99
	(0.17)	(0.00)	(0.38)	(0.86)	(0.36)	(0.89)	(0.05)	(0.15)	(0.09)	(0.02)
UTX	0.75	0.83	0.95	0.92	0.95	0.98	0.99	0.98	0.99	0.91
	(0.00)	(0.00)	(0.44)	(0.01)	(0.00)	(0.20)	(0.61)	(0.34)	(0.68)	(0.00)
WMT	0.85	0.85	0.90	0.92	0.99	1.00	0.98	1.00_-	1.01	0.94
	(0.00)	(0.00)	(0.04)	(0.01)	(0.40)	(0.12)	(0.06)	(0.00)	(0.00)	(0.00)

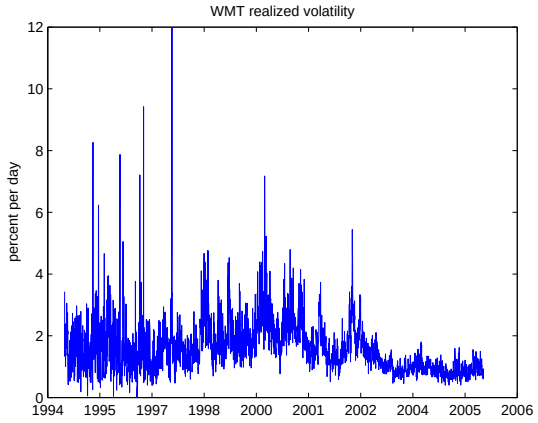
TABLE 5. NONLINEAR VS. LOG-LINEAR MODEL WITH BAGGING.

The table reports the ratio of the root mean squared errors from the nonlinear and log-linear models for one-day-ahead forecasts between 1997 and 2005 with bagging. Numbers below one indicate that nonlinearity improves the forecast performance. The numbers in parentheses report the p-value of Giacomini and White's (2006) test of conditional predictive ability.

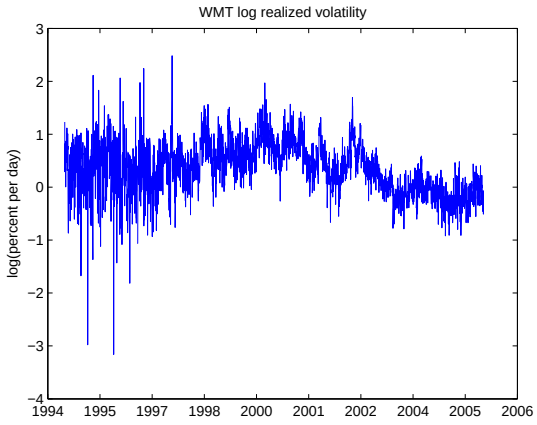
Series	1997	1998	1999	2000	2001	2002	2003	2004	2005	1997–2005
AA	0.96	1.01	1.00_-	1.04	1.01	1.01	0.99	1.00	1.00	1.01
	(0.02)	(0.03)	(0.00)	(0.00)	(0.17)	(0.39)	(0.06)	(0.99)	(0.18)	(0.00)
AIG	1.01	1.06	1.01	1.03	1.01	1.00_-	0.99	1.00	1.00	1.05
	(0.00)	(0.00)	(0.00)	(0.07)	(0.08)	(0.06)	(0.56)	(0.26)	(0.39)	(0.00)
AXP	1.02	1.15	1.01	1.00_-	1.01	1.00	0.99	0.99	0.99	1.07
	(0.00)	(0.00)	(0.00)	(0.00)	(0.52)	(0.26)	(0.65)	(0.02)	(0.11)	(0.00)
BA	1.02	1.00	1.01	1.00	1.04	0.99	1.00	1.00	1.00	1.02
	(0.00)	(0.83)	(0.02)	(0.00)	(0.04)	(0.44)	(0.79)	(0.24)	(0.69)	(0.00)
CAT	1.05	1.01	0.99	1.02	1.02	1.00	1.00	1.02	1.01	1.04
	(0.00)	(0.00)	(0.59)	(0.00)	(0.11)	(0.87)	(0.68)	(0.00)	(0.39)	(0.00)
DD	0.99	0.98	1.00_-	1.01	0.99	1.01	0.99	1.01	1.00	1.04
	(0.00)	(0.08)	(0.10)	(0.41)	(0.48)	(0.54)	(0.93)	(0.02)	(0.44)	(0.00)
DIS	0.98	1.05	1.03	0.99	1.01	1.00	1.00	1.01	0.99	1.02
	(0.08)	(0.00)	(0.02)	(0.24)	(0.39)	(0.17)	(0.36)	(0.01)	(0.49)	(0.00)
GE	1.05	1.04	1.02	1.01	1.00_+	1.03	1.01	0.99	0.99	1.05
	(0.00)	(0.00)	(0.18)	(0.13)	(0.07)	(0.13)	(0.00)	(0.01)	(0.00)	(0.00)
GM	0.99	1.07	1.04	1.02	1.01	1.01	1.01	1.00_-	1.00	1.03
	(0.23)	(0.00)	(0.00)	(0.00)	(0.22)	(0.21)	(0.34)	(0.06)	(0.17)	(0.00)
HD	0.98	1.08	1.01	1.01	1.01	1.03	1.00	0.99	0.99	1.03
	(0.13)	(0.00)	(0.08)	(0.00)	(0.20)	(0.04)	(0.56)	(0.38)	(0.50)	(0.00)
HON	1.00_+	1.02	1.01	1.00_+	1.01	1.00_+	1.01	1.00	1.02	1.03
	(0.05)	(0.00)	(0.00)	(0.00)	(0.16)	(0.01)	(0.48)	(0.65)	(0.02)	(0.00)
IBM	1.00	1.06	1.01	1.02	1.01	1.00	1.01	0.99	1.00	1.03
	(0.04)	(0.00)	(0.00)	(0.00)	(0.15)	(0.44)	(0.00)	(0.01)	(0.13)	(0.00)
JNJ	1.00	1.01	1.00	0.99	1.00	1.00	1.01	1.00_-	1.00	1.01
	(0.97)	(0.03)	(0.74)	(0.08)	(0.35)	(0.54)	(0.00)	(0.00)	(0.63)	(0.01)
JPM	1.00_+	1.01	1.00	0.99	1.00	0.98	0.99	1.01	0.99	1.03
	(0.00)	(0.00)	(0.30)	(0.01)	(0.93)	(0.85)	(0.38)	(0.09)	(0.10)	(0.00)
KO	1.00	1.02	1.00	0.97	1.02	1.01	0.99	1.01	1.00_+	1.02
	(0.46)	(0.02)	(0.77)	(0.59)	(0.02)	(0.67)	(0.01)	(0.00)	(0.01)	(0.00)
MCD	1.01	1.04	0.98	0.99	1.00	0.99	1.01	1.00_-	1.00	1.01
	(0.47)	(0.00)	(0.15)	(0.54)	(0.91)	(0.08)	(0.41)	(0.02)	(0.57)	(0.00)
MMM	1.00_-	1.04	0.99	0.99	1.00	1.00	0.99	0.99	1.00_-	1.02
	(0.00)	(0.00)	(0.88)	(0.89)	(0.23)	(0.93)	(0.92)	(0.49)	(0.02)	(0.00)
MO	0.99	1.00	1.00_+	0.99	1.00	1.00	1.00	1.01	1.00	1.02
	(0.00)	(0.52)	(0.06)	(0.46)	(0.29)	(0.64)	(0.54)	(0.02)	(0.46)	(0.00)
MRK	0.99	1.02	0.99	1.01	1.01	1.02	1.02	1.01	1.00	1.02
	(0.00)	(0.03)	(0.09)	(0.01)	(0.11)	(0.08)	(0.00)	(0.00)	(0.87)	(0.00)
PFE	1.00_-	1.00	1.06	1.02	1.01	1.08	1.01	1.01	1.00_-	1.03
	(0.02)	(0.86)	(0.06)	(0.37)	(0.31)	(0.01)	(0.01)	(0.02)	(0.08)	(0.00)
PG	0.99	1.00_-	0.99	0.99	1.00	1.00	0.99	1.00	1.00	1.01
	(0.00)	(0.01)	(0.45)	(0.74)	(0.95)	(0.25)	(0.46)	(0.84)	(0.22)	(0.00)
UTX	0.95	0.99	1.00	1.02	1.02	1.00	1.00	1.00	1.00	1.00
	(0.03)	(0.02)	(0.94)	(0.00)	(0.33)	(0.71)	(0.99)	(0.36)	(0.45)	(0.00)
WMT	1.01	1.10	1.01	1.02	1.04	1.03	1.05	1.06	1.01	1.07
	(0.08)	(0.00)	(0.16)	(0.00)	(0.00)	(0.04)	(0.00)	(0.00)	(0.00)	(0.00)



(a)



(b)



(c)

FIGURE 1. Time series of (a) returns, (b) realized volatility, and (c) logarithmic realized volatility of WMT.

(E. Hillebrand) DEPARTMENT OF ECONOMICS, LOUISIANA STATE UNIVERSITY, BATON ROUGE, LA 70803, PHONE: (225) 578-3795, FAX (225) 578-3807.

E-mail address: erhil@lsu.edu

(M. C. Medeiros) DEPARTMENT OF ECONOMICS, PONTIFICAL CATHOLIC UNIVERSITY OF RIO DE JANEIRO, RIO DE JANEIRO, RJ, BRAZIL.

E-mail address: mcm@econ.puc-rio.br

Departamento de Economia PUC-Rio
Pontifícia Universidade Católica do Rio de Janeiro
Rua Marques de Sâo Vicente 225 - Rio de Janeiro 22453-900, RJ
Tel.(21) 35271078 Fax (21) 35271084
www.econ.puc-rio.br
flavia@econ.puc-rio.br