Forecasting Realized Volatility with Linear and Nonlinear Models

Michael McAleer
Econometric Institute
Erasmus School of Economics
Erasmus University Rotterdam
and
Tinbergen Institute
The Netherlands
and
Center for International Research on the Japanese Economy (CIRJE)
Faculty of Economics
University of Tokyo

Marcelo C. Medeiros
Department of Economics
Pontifical Catholic University of Rio de Janeiro

EI2009-37

October 2009

*****The first author wishes to acknowledge the financial support of the Australian Research Council and National Science Council, Taiwan. The second author wishes to thank the CNPq/Brazil for partial financial support.

Abstract

In this paper we consider a nonlinear model based on neural networks as well as linear models to forecast the daily volatility of the S&P 500 and FTSE 100 indexes. As a proxy for daily volatility, we consider a consistent and unbiased estimator of the integrated volatility that is computed from high frequency intra-day returns. We also consider a simple algorithm based on bagging (bootstrap aggregation) in order to specify the models analyzed in the paper.

Keywords: Financial econometrics, volatility forecasting, neural networks, nonlinear models, realized volatility, bagging.

JEL Classifications: C22, C53, G12, G17.

1. Introduction

Given the rapid growth in financial markets and the continual development of new and more complex financial instruments, there is an ever-growing need for theoretical and empirical knowledge of the volatility inherent in financial time series. It is widely known that the daily returns of financial assets, especially of stocks, are difficult, if not impossible, to predict, although the volatility of the returns seems to be relatively easier to forecast. Therefore, it is hardly surprising that financial econometrics and, in particular, the modeling of financial volatility, has played such a central role in modern pricing and risk management theories.

There is, however, an inherent problem in using models where the volatility measure plays a central role. The conditional variance is latent, and hence is not directly observable. It can be estimated, among other approaches, by the (Generalized) Autoregressive Conditional Heteroskedasticity, or (G)ARCH, family of models proposed by Engle (1982) and Bollerslev (1986), stochastic volatility (SV) models (see, for example, Taylor (1986)), or exponentially weighted moving averages (EWMA), as advocated by the Riskmetrics methodology (see McAleer (2005) for a recent exposition of a wide range of univariate and multivariate, conditional and stochastic, models of volatility, and Asai, McAleer and Yu (2006) for a review of the growing literature on multivariate stochastic volatility models). However, as observed by Bollerslev (1987), Malmsten and Teräsvirta (2004), and Carnero, Peña, and Ruiz (2004), among others, most of the latent volatility models fail to describe satisfactorily several stylized facts that are observed in financial time series.

An empirical fact that standard latent volatility models fail to describe in an adequate manner is the low, but slowly decreasing, autocorrelations in the squared returns that are associated with high excess kurtosis of returns. Correctly describing the dynamics of the returns is important in order to obtain accurate forecasts of the future volatility which, in turn, is important in risk analysis and management. In this sense, the assumption of Gaussian standardized returns has been refuted in many studies, and heavy-tailed distributions have instead been used.

The search for an adequate framework for the estimation and prediction of the conditional variance of financial assets returns has led to the analysis of high frequency intraday data. Merton (1980) noted that the variance over a fixed interval can be estimated arbitrarily, although accurately, as the sum of squared realizations, provided the data are available at a sufficiently high sampling frequency. More recently, Andersen and Bollerslev (1998) showed that ex post daily foreign exchange volatility is best measured by aggregating 288 squared five-minute returns. The five-minute frequency is a trade-off between accuracy, which is theoretically optimized using the highest possible frequency, and microstructure noise that can arise through the bid-ask bounce, asynchronous trading, infrequent trading, and price discreteness, among other factors (see Madhavan (2000) and Biais, Glosten and Spatt (2005) for very useful surveys).

Ignoring the remaining measurement error, which can be problematic, the ex post volatility essentially becomes “observable”. Andersen and Bollerslev (1998) and Patton (2008) used this new volatility measure to evaluate the out-of-sample forecasting performance of GARCH models. As volatility becomes “observable”, it can be modeled directly, rather than being treated as a latent variable. Based on the theoretical results of Barndorff-Nielsen and Shephard (2002), Andersen, Bollerslev, Diebold and Labys (2003) and Meddahi (2002), several recent studies have documented the properties of realized volatilities constructed from high frequency data. However, microstructure effects introduce a severe bias on the daily volatility estimation. Zhang, Mykland and Aït-Sahalia (2005), Bandi and Russell (2006), Hansen and Lunde (2006), and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), among others, have discussed various solutions to the inconsistency problem.

In this paper we consider the forecasting of stock marketing volatility via nonlinear models based on a neural network version of the Heterogenous Autoregressive Model (HAR) of Corsi (2009). As in Hillebrand and Medeiros (2009) we evaluate the benefits of bagging (bootstrap aggregation) in forecasting daily volatility as well as the inclusion of past cumulated returns over different horizons as possible predictors. As the number of predictors can get quite large, the application of bagging is recommended as a device to improve forecasting performance.

The remainder of the paper is organized as follows. In Section 2, we briefly discuss the main concepts in construction realized volatility measures. In Section 3, the models considered in this paper are presented, while in Section 4 we describe the bagging methodology to specify the models and construct forecasts. The empirical results are presented in Section 5. Section 6 concludes the paper.

2. Realized Volatility

Suppose that, along day *t*, the logarithmic prices of a given asset follow a continuous time diffusion process, as follows:

$$dp(t+\tau) = \mu(t+\tau) + \sigma(t+\tau)dW(t+\tau), \qquad 0 \leq \tau \leq 1, t = 1,2,3,\ldots, \tag{1}$$

where $p(t+\tau)$ is the logarithmic price at time $t+\tau$ $\mu(t+\tau)$ is the drift component, $\sigma(t+\tau)$ is the instantaneous volatility (or standard deviation), and $W(t+\tau)$ is a standard Brownian motion.

Andersen, Bollerslev, Diebold and Labys (2003) and Barndorff-Nielsen and Shephard (2002) showed that daily returns, $r(t) = p(t) - p(t-1)$, are Gaussian conditionally on $\mathcal{F}_t = \mathcal{F}\{\mu(t+\tau-1), \sigma(t+\tau-1)\}_{\tau=0}^{1}$, the σ -algebra (information set) generated by the sample paths of $\mu(t+\tau-1)$ and $\sigma(t+\tau-1)$, $0 \leq \tau \leq 1$, such that

$$r_t|\mathcal{F}_t \sim \mathrm{N}\left[\int_0^1 \mu(t+\tau-1)d\tau, \int_0^1 \sigma(t+\tau-1)d\tau\right].$$

The term $IV_t = \int_0^1 \sigma(t+\tau-1)d\tau$ is known as the *integrated variance*, which is a measure of the day-*t* ex post volatility. The integrated variance is typically the object of interest as a measure of the true daily volatility.

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 $\{\tau_1, \cdots, \tau_n\}$, $0 = \tau_1 < \tau_2 \cdots < \tau_n = 1$. The length of the i th 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 $p_{t,i}, i = 1, \dots, n,$ be the i th log price observation during day t , such that $r_{t,i} = p_{t,i} - p_{t,i-1}$ is the i th intra-period return of day t . Realized variance is defined as

$$RV_t = \sum_{i=1}^{n} r_{t,i}^2. \tag{2}$$

Realized volatility is the square-root of (2).

Under regularity conditions, including the assumption of uncorrelated intraday returns, realized variance RV_t^2 is a consistent estimator of integrated variance, such that $RV_t^2 \xrightarrow{p} IV_t$. However, when returns are serially correlated, realized variance is a biased and inconsistent 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.

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.

Recently, Barndorff-Nielsen, Hansen, Lunde and Shephard (2008), hereafter BHLS (2008), proposed the flat-top kernel-based estimator:

$$RV_t^{(BHLS)} = RV_t + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)(\hat{\gamma}_h + \hat{\gamma}_{-h}), \tag{3}$$

where $k(x)$ for $x \in [0,1]$ is a non-stochastic weight function such that $k(0) = 1$ and $k(1) = 0$, RV_t is defined as in (2) and

$$\hat{\gamma}_h = \frac{n}{n-h} \sum_{j=1}^{n-h} r_{t,j} r_{t,j+h} \, .$$

BHLS (2008) discussed different kernels and provided all the technical details.

3. The Models

Let y_t be the square-root of the logarithm of a consistent and unbiased estimator for the integrated variance of day t , such as the estimator in (3), and call it the daily "realized volatility"[1]. Define daily accumulated logarithm returns over an h -period interval as

[1] In fact, there is an abuse of terminology here as "realized volatility" specifically refers to the square root of the sum of the squared intra-day returns, which is a biased and inconsistent estimator of the daily integrated volatility under the presence of micro-structure noise. However, to simplify notation and terminology, we will refer to any unbiased and consistent estimator as realized volatility.

$$r_{h,t} = \sum_{i=0}^{h-1} r_{t-i}, \tag{4}$$

where r_t is the daily return at day t. Furthermore, define the average log realized volatility over h -days as

$$y_{h,t} = \frac{1}{h}\sum_{i=0}^{h-1} y_{t-i}. \tag{5}$$

3.1. The Linear Heterogeneous Autoregressive Model

The Linear Heterogeneous Autoregressive (HAR) model proposed by Corsi (2009) is defined as

$$y_t = \beta_0 + \sum_{i \in I} \beta_i y_{i,t-1} + \varepsilon_t = \beta_0 + \beta' x_{t-1} + \varepsilon_t, \tag{6}$$

where $x_{t-1} = (y_{i_1,t-1}, \dots, y_{i_p,t-1})'$, $I = (i_1, i_2, \dots, i_p)$ is a set of p indices with $0 < i_1 < i_2 < \dots < i_p < \infty$ and $i = 1, \dots, p$. Throughout the paper, ε_t is a zero-mean and uncorrelated process with finite, but not necessarily constant variance (Corsi, Mittnik, Pigorsch, and Pigorsch 2008). Corsi (2009) advocated the use of $I = (1,5,22)$. His specification builds on the HARCH model proposed by Müller, Dacorogna, Dave, Olsen, Pictet, and von Weizsaecker (1997). This type of specification captures long-range dependence by aggregating the log realized volatility over the different time scales in I (daily, weekly, and monthly).

Hillebrand and Medeiros (2009) consider more lags than 1, 5, and 22, as well as, dummy variables for weekdays and macroeconomic announcements and past cumulated returns over different horizons as defined in (3). Hence,

$$y_t = \delta' d_t + \sum_{i_j \in I} \beta_i y_{i_j, t-1} + \sum_{\kappa_j \in K} \lambda_j r_{\kappa_j, t-1} + \varepsilon_t = \delta' d_t + \beta' x_{t-1} + \lambda' r_{t-1} + \varepsilon_t, \qquad (7)$$

where d_t is a vector of n dummy variables as described above, x_{t-1} is defined as in (6), $r_{t-1} = \left(r_{\kappa_1, t-1}, \cdots, r_{\kappa_q, t-1}\right)'$, $K = \left(\kappa_1, \kappa_2, \cdots, \kappa_q\right)$ is a set of q indices with $0 < \kappa_1 < \kappa_2 < \cdots < \kappa_q < \infty$ and $i = 1, \ldots, K$. The final set of variables in the model was determined by a bagging strategy as a flexible choice of the lag structure imposes high computational costs.

3.2. The Nonlinear HAR Model

McAleer and Medeiros (2008) proposed an extension of the linear HAR model by incorporating smooth transitions. The resulting model is called the Multiple-Regime Smooth Transition HAR (HARST) model and is defined as

$$y_t = \delta' d_t + \beta_0' x_{t-1} + \sum_{i=1}^{M} \beta_i' x_{t-1} f[\gamma_i (z_t - c_i)] + \varepsilon_t, \qquad (8)$$

where z_t is a transition variable, d_t and ε_t are defined as before, and

$$f[\gamma_i (z_t - c_i)] = \frac{1}{1 + e^{-\gamma_i (z_t - c_i)}} \qquad (9)$$

is the logistic function. The authors also presented a modeling cycle based on statistical arguments to select the set of explanatory variables as well as the number of regimes, M.

Hillebrand and Medeiros (2009) put forward a nonlinear version of the HAR model based on neural networks (NN). Their specification is defined as follows:

$$y_t = \beta_0' w_{t-1} + \sum_{i=1}^{M} \beta_i f(\gamma_i' w_{t-1}) + \varepsilon_t, \tag{10}$$

where $w_{t-1} = (d_t', x_{t-1}', r_{t-1}')'$, ε_t is defined as above, and $f(\gamma_i' w_{t-1})$ is the logistic function as in (9).

As first discussed in Kuan and White (1994), the model defined by equation (10) 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 w_{t-1} (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 (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.

As model (10) is, in principle, more flexible than model (8) we will consider only the NN-HAR model in our empirical experiment.

4. Bagging Linear and Nonlinear HAR Models

4.1. What is Bagging?

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 2008, Lee and Yang 2006). Applications include, among others, financial volatility (Huang and Lee 2007, Hillebrand and Medeiros 2009), equity premium (Huang and Lee 2008), and employment data (Rapach and Strauss 2008).

4.2. Bagging the Linear HAR Model

Using the same notation as in Section 3, set $w_{t-1} = (d_t', x_{t-1}', r_{t-1}')' \in \mathbb{R}^J$, $J = p + q + n$, and write (7) as

$$y_t = \theta' w_{t-1} + \varepsilon_t. \tag{11}$$

The bagging forecast for model (11) is constructed in steps as follows:

PROPOSAL 1: Bagging the linear HAR model

(1) *Arrange the set of tuples* $(y_t, w_{t-1}')', t = 1, \ldots, T$ *, in the form of a matrix* X *of dimension* $T \times J$ *;*

(2) *Construct bootstrap samples of the form* $\{(y^*_{(i)1}, w'^*_{(i)0}), \ldots, (y^*_{(i)T}, w'^*_{(i)T-1})\}, i = 1, \ldots, B$ *, by drawing blocks of* m *rows of* X *with replacement, where the block size* m *is chosen to capture possible dependence in the error term of the realized volatility series, such as conditional variance ("volatility of volatility");*

(3) *Compute the* i *th bootstrap one-step ahead forecast as*

$$\hat{y}^*_{(i)t|t-1} = \begin{cases} 0 & if\ |t_j| < c, \quad \forall j, \\ \hat{\theta}^{*\prime} \tilde{w}^*_{(i)t-1} & otherwise, \end{cases}$$

where t_j *is the* t *-statistic for the null hypothesis* $\mathrm{H}_0 : \theta_j = 0$, $\tilde{W}^*_{(i)t-1} = S^* w^*_{t-1}$,

S^* *is a diagonal selection matrix, which depends on the bootstrap sample, with the* j *th diagonal element given by*

$$S_{ij}^{*} = \begin{cases} 1 & if\ |t_j| \geq c, \\ 0 & otherwise, \end{cases} \qquad \forall j,$$

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

$$\hat{\theta} = \left[\sum_{t=1}^{T} \tilde{w}^{*}_{(i)t-1} \tilde{w}^{*\prime}_{(i)t-1}\right]^{-1} \sum_{t=1}^{T} \tilde{w}^{*}_{(i)t-1} y^{*}_{t}.$$

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

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

We choose a block size of $m = T^{\frac{1}{3}}$ for the bootstrap procedure described above. This allows for dependence in the error term of equation (11). The critical value c is set equal to 1.96, corresponding to a two-sided test at the 96% confidence level.

4.3. Bagging Nonlinear HAR Models

There are two main problems in specifying model (10): the selection of variables in the vector x and the number of hidden units M. There are many approaches in the literature to tackle these problems. For example, when model (10) is seen as a variant of parametric smooth transition models, Medeiros, Teräsvirta, and Rech (2006) proposed a methodology based on statistical arguments to variable selection and determination of M. However, this approach is not directly applicable here, as we advocate model (10) as a semi-parametric specification. On the other hand, as shown in Hillebrand and Medeiros

(2009), Bayesian regularization (MacKay 1992) is a viable alternative, which is equivalent to penalized quasi-maximum likelihood.

In this paper, we do not specify neither the elements of x nor the number of hidden units, M . In turn, in each bootstrap sample, we randomly select M from a uniform distribution on the interval [0,20], and the elements of x are selected as the ones with significant coefficients in the linear HAR case. The bagging procedure can be summarized as follows:

PROPOSAL 2: Bagging the NN-HAR model

(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 NN-HAR model randomly selecting M from a uniform distribution on the interval [0,20]. Compute the i th bootstrap one-step ahead forecast and call it $\hat{y}^{*}_{(i)t|t-1}$.

(3) Compute the average forecast over the bootstrap samples:

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

5. Empirical Results

We use high frequency tick-by-tick on S&P 500 futures from January 2, 1996 to March 29, 2007 (2796 observations) and FTSE 100 futures from January 2, 1996 to December 28, 2007 (3001 observations). In computing the daily realized volatilities, we employ the realized kerned estimator with the modified Tukey-Hanning kernel of BHLS (2008). As it is a standard practice in the literature, we focus on the logarithm of the daily realized volatilities. Figures 1 and 2 illustrate the data. The last 1000 observations are left out the estimation sample in order to evaluate the out-of-sample performance of different models.

In this paper we consider the following competing models: the standard heterogeneous autoregressive (HAR) model with average volatility over one, five, and 22 days as regressors (see equation (6)); the flexible HAR model where cumulated returns over one to 200 days and average past volatility over one to 60 days are initially included as possible regressors; the neural network HAR (NN-HAR) model estimated with Bayesian regularization (BR) and the same set of regressors as the flexible HAR model; and finally, the NN-HAR model estimated by nonlinear least squares (LS). Bagging is applied to all models apart from the standard HAR specification.

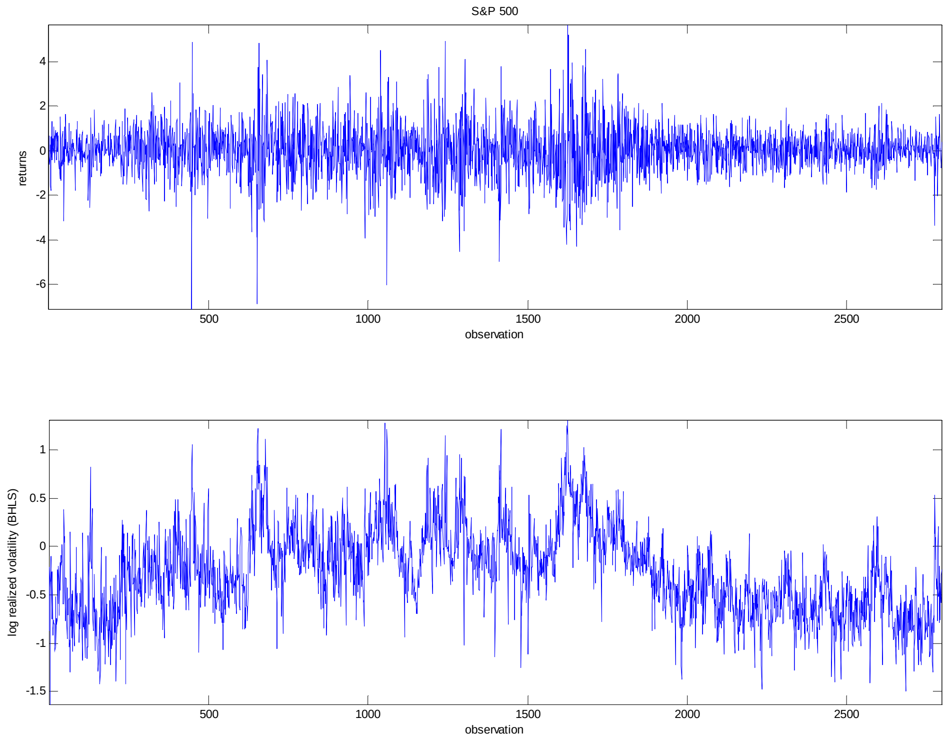

Figure 1. Upper panel: Daily returns for the S&P 500 index. Lower panel: Daily log realized volatility computed via the method described in BHLS (2008) and using the Tukey-Hanning kernel. We use high frequency tick-by-tick on S&P 500 futures from January 2, 1996 to March 29, 2007.

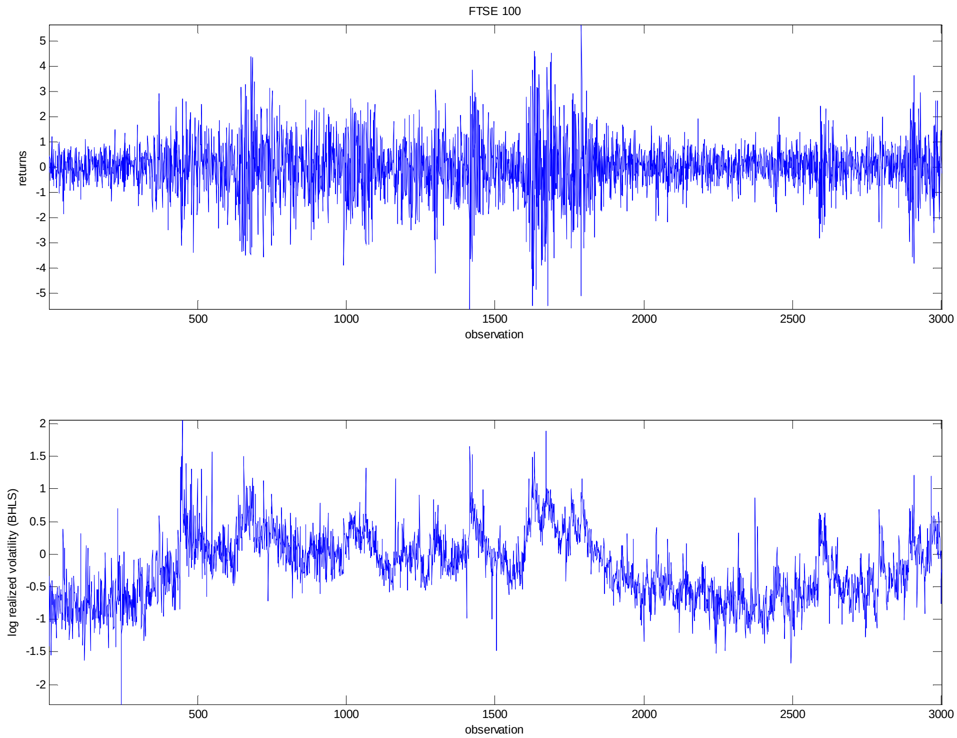

Figure 2. Upper panel: Daily returns for the FTSE index. Lower panel: Daily log realized volatility computed via the method described in BHLS (2008) and using the Tukey-Hanning kernel. We use high frequency tick-by-tick on FTSE 100 futures from January 2, 1996 to December 28, 2007.

The forecasting results are presented in Tables 1 and 2. Table 1 shows the root mean squared error (RMSE) and the mean absolute error (MAE) as well as the mean, the standard deviation, the maximum, and the minimum one-step-ahead forecast error for the four models considered in the empirical exercise. From the table it is clear that the flexible linear HAR model and the nonlinear HAR model estimated with Bayesian regularization (NN-HAR (BR)) are the two best models. However, the performance of the standard HAR specification is not much worse. On the other hand, the NN-HAR model without Bayesian regularization seems to be the worst model among the four competing ones. The results are similar for the S&P 500 and the FTSE 100.

Table 2 presents the *p*-value of the modified Diebold-Mariano test of equal predictive accuracy of different models with respect the benchmark standard HAR model. The test is applied to the squared errors as well as to the absolute errors. It is clear from the table that both the flexible linear HAR and the NN-HAR (BR) models have superior out-of-sample performance than the standard HAR model in the case of the S&P 500 index. For the FTSE 100, the NN-HAR (BR) model has a statistically superior performance than the standard HAR specification only when the absolute errors are considered.

6. Conclusions

In this paper we considered linear and nonlinear models to forecast daily realized volatility: the standard heterogeneous autoregressive (HAR) model with average volatility over one, five, and 22 days as regressors; the flexible HAR model where cumulated returns over one to 200 days and average past volatility over one to 60 days are initially included as possible regressors; the neural network HAR (NN-HAR) model estimated with Bayesian regularization (BR) and the same set of regressors as the flexible HAR model; and finally, the NN-HAR model estimated by nonlinear least squares (LS). Both the flexible HAR and the NN-HAR (BR) models outperformed the benchmark HAR model. The NN-HAR model estimated with nonlinear least squares was the worst model among all the alternatives considered.

Table 1. Forecasting Results: Main Statistics

The table shows the root mean squared error (RMSE) and the mean absolute error (MAE) as well as the mean, the standard deviation, the maximum, and the minimum one-step-ahead forecast error for the following models: the standard heterogeneous autoregressive (HAR) model; the flexible HAR model where cumulated returns over one to 200 days and average past volatility over one to 60 days are initially included as possible regressors; the neural network HAR (NN-HAR) model estimated with Bayesian regularization (BR) and the same set of regressors as the flexible HAR model; and the NN-HAR model estimated by nonlinear least squares (LS). Bagging is applied to all models, apart from the standard HAR specification.

Model	**RMSE**	**MAE**	**Mean**	**Std. D.**	**Max.**	**Min.**
			S&P 500			
Flexible HAR w/ bagging	0.228	0.180	-0.038	0.225	1.326	-0.853
NN-HAR (BR) w/ bagging	0.229	0.179	-0.043	0.225	1.305	-0.865
NN-HAR (LS) w/ bagging	0.247	0.195	-0.096	0.228	1.208	-0.870
HAR (1,5,22) w/o bagging	0.237	0.186	-0.041	0.233	1.268	-0.896
			FTSE 100			
Flexible HAR w/ bagging	0.264	0.198	-0.011	0.264	1.745	-0.900
NN-HAR (BR) w/ bagging	0.266	0.198	-0.015	0.266	1.720	-0.882
NN-HAR (LS) w/ bagging	0.292	0.224	-0.094	0.277	1.570	-1.000
HAR (1,5,22) w/o bagging	0.270	0.202	-0.016	0.268	1.694	-0.912

Table 2. Forecasting Results: Diebold-Mariano Test

The table shows the *p*-value of the modified Diebold-Mariano test of equal predictive accuracy of different models with respect the benchmark standard HAR model. The test is applied to the squared errors as well as to the absolute errors. The following models are considered : the flexible HAR model where cumulated returns over one to 200 days and average past volatility over one to 60 days are initially included as possible regressors; the neural network HAR (NN-HAR) model estimated with Bayesian regularization (BR) and the same set of regressors as the flexible HAR model; and the NN-HAR model estimated by nonlinear least squares (LS). Bagging is applied to all models, apart from the benchmark standard HAR specification.

Model	**Squared Errors**	**Absolute Errors**
	S&P 500	
Flexible HAR w/ bagging	4.52e-5	1.36e-4
NN-HAR (BR) w/ bagging	2.89e-4	3.23e-4
NN-HAR (LS) w/ bagging	0.001	0.004
	FTSE 100	
Flexible HAR w/ bagging	0.011	0.006
NN-HAR (BR) w/ bagging	0.144	0.016
NN-HAR (LS) w/ bagging	5.68e-11	1.30e-10

References

Andersen, T. and T. Bollerslev (1998) "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts." *International Economic Review*, 39(4): 885-906.

Andersen, T., Bollerslev, T., Diebold, F.X. and H. Ebens (2001) "The Distribution of Realized Stock Return Volatility." *Journal of Financial Economics*, 61(1): 43-76.

Andersen, T., Bollerslev, T., Diebold, F.X. and P. Labys (2003) "Modeling and Forecasting Realized Volatility." *Econometrica*, 71(2): 579-625.

Asai, M., McAleer, M. and J. Yu (2006) "Multivariate Stochastic Volatility: A Review." *Econometric Reviews*, 25(2): 145-175.

Bandi, F.M. and J.R. Russell (2006) "Separating Market Microstructure Noise from Volatility." *Journal of Financial Economics*, 79(3): 655-692.

Barndorff-Nielsen, O., Hansen, P., Lunde, A. and N. Shephard (2008) "Designing Realized Kernels to Measure the Ex-post Variation of Equity Prices in the Presence of Noise." *Econometrica*, 76(6): 1481-1536,

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(2): 253-280.

Biais, B., Glosten, L. and C. Spatt (2005) "Market Microstructure: A Survey of Microfoundations, Empirical Results, and Policy Implications." *Journal of Financial Markets*, 8(2): 217-264.

Bollerslev, T. (1986) "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics*, 21(3): 307-328.

Bollerslev, T. (1987) "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return." *Review of Economics and Statistics*, 69(3): 542-547.

Breiman, L. (1996) "Bagging Predictors." *Machine Learning*, 24(2): 123-140.

Bühlmann, P. and B. Yu (2002) "Analyzing Bagging." *Annals of Statistics*, 30(4): 927-961.

Campbell, J., Lo, A. and A. Mackinlay (1997) *The Econometrics of Financial Markets*. Princeton University Press: Princeton, NJ.

Carnero, M.A., Peña, D. and E. Ruiz (2004) "Persistence and Kurtosis in GARCH and Stochastic Volatility Models." *Journal of Financial Econometrics*, 2(2): 319-342.

Chen, X., Racine, J. and N.R. Swanson (2001) "Semiparametric ARX Neural-Network Models with an Application to Forecasting Inflation." *IEEE Transactions on Neural Networks*, 12(4): 674-683.

Chen, X. and X. Shen (1998) "Sieve Extremum Estimates for Weakly Dependent Data." *Econometrica*, 66(2): 289-314.

Chen, X. and H. White (1998) "Improved Rates and Asymptotic Normality for Nonparametric Neural Network Estimators." *IEEE Transactions on Information Theory*, 45(2): 682-691.

Corsi, F. (2009) "A Simple Approximate Long Memory Model of Realized Volatility." *Journal of Financial Econometrics*, 7(2): 174-196.

Corsi, F., Mittnik, S., Pigorsch, C. and U. Pigorsch (2008) "The Volatility of Realized Volatility." *Econometric Reviews*, 27(1-3): 46-78.

Engle, R.F. (1982) "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica*, 50(4): 987-1007.

Hansen, P.R. and A. Lunde (2006) "Realized Variance and Market Microstructure Noise" (with discussion). *Journal of Business and Economic Statistics*, 24(2): 127-218.

Hillebrand, E. and M.C. Medeiros (2009) "The Benefits of Bagging for Forecast Models of Realized Volatility." To appear in *Econometric Reviews*.

Hornik, K., Stinchombe, A. and H. White (1989) "Multi-Layer Feedforward Networks are Universal Approximators." *Neural Networks*, 2(5): 359-366.

Hornik, K., Stinchombe, A., White, H. and P. Auer (1994) "Degree of Approximation Results for Feedforward Networks Approximating Unknown Mappings and Their Derivatives." *Neural Computation*, 6(6): 1262-1274.

Huang, H. and T.-H. Lee (2007) "Forecasting Using High-Frequency Financial Time Series." Working Paper, University of California at Riverside.

Huang, H. and T.-H. Lee (2008) "To Combine Forecasts or to Combine Information." To appear in *Econometric Reviews*.

Inoue, A. and L. Killian (2004) “Bagging Time Series Models.” Discussion Paper No. 4333, Centre for Economic Policy Research (CEPR).

Inoue, A. and L. Killian (2008): “How Useful is Bagging in Forecasting Economic Time Series? A Case Study of U.S. Consumer Price Inflation.” *Journal of the American Statistical Association*, 103(482): 511-522.

Kuan, C.-M. and H. White (1994) “Artificial Neural Networks: An Econometric Perspective.” *Econometric Reviews*, 13(1): 1-91.

Lee, T.-H. and Y. Yang (2006) “Bagging Binary and Quantile Predictors for Time Series.” *Journal of Econometrics*, 135(1-2): 465-497.

MacKay, D.J.C. (1992) “A Practical Bayesian Framework for Backpropagation Networks.” *Neural Computation*, 4(3): 448-472.

Madhavan, A. (2000) “Market Microstructure: A Survey.” *Journal of Financial Markets*, 3(3): 205-258.

Malmsten, H., and T. Teräsvirta (2004) “Stylized Facts of Financial Time Series and Three Popular Models of Volatility.” Working Paper Series in Economics and Finance 563, Stockholm School of Economics.

McAleer, M. (2005) “Automated Inference and Learning in Modeling Financial Volatility.” *Econometric Theory*, 21(1): 232-261.

McAleer, M. and M. Medeiros (2008) “A Multiple Regime Smooth Transition Heterogenous Autoregressive Model for Long Memory and Asymmetries.” *Journal of Econometrics*, 147(1): 104-119.

Meddahi, N. (2002) “A Theoretical Comparison Between Integrated and Realized Volatility.” *Journal of Applied Econometrics*, 17(5): 479-508.

Medeiros, M.C., Teräsvirta, T. and G. Rech (2006) “Building Neural Network Models for Time Series: A Statistical Approach.” *Journal of Forecasting*, 25(1): 49-75.

Medeiros, M.C. and A. Veiga (2000) “A Hybrid Linear-Neural Model for Time Series Forecasting.” *IEEE Transactions on Neural Networks*, 11(6): 1402-1412.

Merton, R.C. (1980) “On Estimating the Expected Return on the Market: An Exploratory Investigation.” *Journal of Financial Economics*, 8(4): 323-361.

Müller, U.A., Dacorogna, M.M., Dave, R.D., Olsen, R.B., Puctet, O.V. and J. von Weizsacker (1997) “Volatilities of Different Time Resolutions – Analyzing the Dynamics of Market Components.” *Journal of Empirical Finance*, 4(2-3): 213-239.

Patton, A. (2008) “Volatility Forecast Evaluation and Comparison Using Imperfect Volatility Proxies.” To appear in *Journal of Econometrics*.

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.

Taylor, S.J. (1986) *Modelling Financial Time Series*. Wiley: Chichester.

Trapletti, A., Leisch, F. and K. Hornik (2000) “Stationary and Integrated Autoregressive Neural Network Processes.” *Neural Computation*, 12(10): 2427-2450.

van Dijk, D., Teräsvirta, T. and P.H. Franses (2002) “Smooth Transition Autoregressive Models - A Survey of Recent Developments.” *Econometric Reviews*, 21(1): 1-47.

Zhang L., Mykland, P. and Y. Ait-Sahalia (2005) “A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data.” *Journal of the American Statistical Association*, 100(472): 1394-1411.