Modeling and Forecasting Realized Volatility: the Role of Power Variation*

Chun Liu[†] John M. Maheu[‡]

November 2005

Abstract

How to measure and model volatility is an important issue in finance. Recent research uses high frequency intraday data to construct ex post measures of daily volatility. This measure, called *realized volatility*, permits the modeling of volatility by traditional time-series methods. Barndorff-Nielsen and Shephard(2004) have introduced additional volatility instruments called realized power variation and realized bipower variation. We investigate the benefits of these volatility instruments in modeling and forecasting volatility. The first contribution of this paper is to demonstrate that realized power variation can provide dramatic improvements in predicting volatility for foreign exchange and equity markets. Secondly, given the large number of possible models, we consider the benefits of Bayesian model averaging. The model average reduces the risk of choosing an individual model and provides overall strong performance for each volatility series and forecast horizon.

*We thank Tom McCurdy who contributed to the preparation of the data used in this paper, and the helpful comments from Mark Kamstra, Alex Maynard, Tom McCurdy, Angelo Melino and Neil Shephard. Maheu thanks the SSHRCC for financial support.

†Department of Economics, 150 St. George St., University of Toronto, Toronto, Canada, M5S 3G7, chun.liu@utoronto.ca.

‡Corresponding author: Department of Economics, 150 St. George St., University of Toronto, Toronto, Canada, M5S 3G7, phone 416-978-1495, fax 416-978-6713, jmaheu@chass.utoronto.ca.

1 Introduction

A large literature investigates the modeling of asset returns. An important feature of any model is the volatility dynamics. Volatility is directly related to the risk associated with holding financial securities, it affects consumption and investment decisions, portfolio choice, and is central to the theory and practice of asset pricing. Solutions to these problems often require a full characterization of the distribution of volatility.

How to measure and model volatility become important questions. Volatility is latent and not observed directly as other variables, such as prices or volume. Traditional parametric approaches, such as GARCH or stochastic volatility models depend upon specific parameterizations as well as distributional assumptions. It is unclear how robust these specifications are when these assumptions are invalid.

A new approach to modeling volatility dynamics has emerged which uses improved measures of ex post volatility constructed from high frequency data. Andersen, Bollerslev, Diebold and Labys (2001), Andersen, Bollerslev, Diebold and Ebens (2001) and Barndorff-Nielsen and Shephard (2002a,2002b) advocate an ex post estimate called realized volatility (RV). RV is constructed from the sum of high frequency squared returns and is a consistent estimate of integrated volatility plus a jump component for a broad class of continuous time models. In contrast to traditional measures of volatility, such as squared returns, realized volatility is more efficient. Recent work has demonstrated the usefulness of this approach in finance. For example, Bollerslev and Zhou (2002) use realized volatility to simplify the estimation of stochastic volatility diffusions, while Fleming, Kirby and Ostdiek (2003) demonstrate that investors who use realized volatility improve portfolio decisions.

This paper contributes to a growing literature that investigates time series models of realized volatility and their forecasting power. Recent contributions include Andersen, Bollerslev, Diebold and Labys (2003), Anderson, Bollerslev and Meddahi (2005), Andreou and Ghysels (2002), Barndorff-Nielsen and Shephard(2004a), Koopman, Jungbacker and Hol (2005), Maheu and McCurdy (2002), and Martens, Dijk, and Pooter (2004). These papers concentrate on pure time series specifications of RV, however, there may be benefits to including additional volatility instruments.

Barndorff-Nielsen and Shephard (2004) have defined several new measures of volatility, and associated estimators. *Realized power variation* (RPV), is a consistent estimate of the integral of the spot volatility process raised to a positive power (integrated power variation), while *realized bipower variation* is a consistent estimate of integrated volatility. RPV defines a range of estimators that are constructed from the sum of powers of the absolute value of high frequency returns, $|r_t|^p$, $p > 0$. Barndorff-Nielsen and Shephard (2004) show that under certain circumstances power variation is robust to jumps.

There are several reasons why RPV can improve the modeling of volatility. First, the absolute value of returns is less sensitive to large movements in prices. Thus models with RPV may provide better predictions during periods with jumps. Secondly, the absolute value of returns displays stronger persistence than squared returns (Ding, Granger and Engle (1993)), and therefore may provide a better signal for volatility. If higher order moments of returns do not exist, such as the fourth moment, the absolute value of returns will be more reliable since its variance is more likely to exist. Finally, Ghysels, Santa-Clara and Valkanov (2005) find the absolute value of high frequency returns improve forecasts of lower frequency realized volatility.

The first contribution of this paper is to provide a detailed investigation of the use of realized power variation in modeling and forecasting volatility. We construct realized volatility and realized power variation at a daily frequency using high frequency intraday observations on returns. We focus on modeling of daily, weekly and biweekly logarithmic average volatility and the forecasting of levels of average volatility. Our analysis relies on simple linear regression models which are easy to estimate. Using Bayesian methods we rank the predictive power in volatility models for foreign exchange and equity markets. Model specifications can include lag terms of power variation, realized volatility, bipower variation as well as a leverage effect. The heterogeneous autoregressive (HAR) model of Corsi(2004), a logarithmic version (log-HAR) and an ARFIMA model are popular specifications for volatility and are included for comparison.

Based on the marginal likelihood we find that the top models include power variation of orders 0.5, 1.0, and 1.5. In some cases the best model includes no lags of realized volatility, and only combinations of RPV terms. In general, our results show that good models of volatility can be obtained by adding in realized power variation into autoregressive specifications of realized volatility. Although there are generally improvements in all forecast horizons, the largest are from one-step ahead daily predictions. As the forecast horizon increases, lags of realized volatility become more important. The ARFIMA model improves upon all pure autoregressive specifications of RV. However, linear models which include RPV terms consistently improve upon the results of all the benchmark models.

In addition, our results from out-of-sample point forecasts support the use of RPV, however, the improvements are less dramatic. The point forecasts only assess one feature of the predictive distribution, the predictive mean, while the marginal likelihood provides a complete assessment of the predictive distribution. This indicates that RPV is most useful for improving the forecast distribution of volatility. The HAR performance, which uses levels of RV can be poor since the distribution of volatility is highly skewed. A leverage effect is found in all the best parameterizations for S&P 500 volatility. We find bipower variation has very limited forecasting power. With one exception, none of the top models include bipower variation regressors.

In all cases, no single specification dominates across markets and forecast horizons. In this situation it is natural to consider the benefits of Bayesian model averaging. Choosing one model ignores model uncertainty which understates the risk in forecasting and can lead to poor predictions. Recent examples of Bayesian model averaging in a macroeconomic context include Fernández, Ley, and Steel (2001), Jacobson and Karlsson (2004), Koop and Potter (2004), and Wright (2003).

The second contribution of this paper is to investigate the benefits of Bayesian model averaging for volatility prediction. There is considerable model uncertainty in all our applications. The ranking of individual models can change dramatically over data series and forecast horizons. The model average combines individual model forecasts based on their past predictive record. Therefore, models with good predictions receive large weights in the model average. The empirical results show the model average to be consistently ranked near the top of the best individual models. Considering all data series and forecast horizons, the Bayesian model average is the dominate model. Model averaging reduces the risk associated with selecting a single model. Our results are in line with Hibon and Evgeniou (2004) who argue that the advantage of model averaging is not necessarily model performance but a reduction in risk.

This paper is organized as follows. In Section 2, we briefly describe the theory behind the improved volatility measures: realized volatility, realized power variation and realized bipower variation. Section 3 presents the regression models in our study and estimation, forecasting and model averaging. Section 4 describes several benchmark models of volatility. Sections 5 details the data. Section 6 reports model comparisons, forecasting performance and estimates for foreign exchange and stock market volatility. The last section concludes with a discussion.

2 Realized Volatility, Power variation and Bipower variation

Suppose the price process belongs to the class of special semi-martingales with jumps, and for illustration, consider the following logarithmic price process:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + \kappa(t)dq(t),\ 0 \leq t \leq T, \tag{1}$$

where $\mu(t)$ is a continuous and locally bounded variation process, $\sigma(t)$ is the stochastic volatility process, $W(t)$ denotes a standard Brownian motion, $dq(t)$ is a counting process with $dq(t) = 1$ corresponding to a jump at time t and $dq(t) = 0$ otherwise, with jump intensity $\lambda(t)$, and $\kappa(t)$ refers to the size of the corresponding jumps. The increment in quadratic variation from time t to $t+1$ is

$$QV_{t+1} = \int_t^{t+1} \sigma^2(s)ds + \sum_{t<s\leq t+1, dq(s)=1} \kappa^2(s) \tag{2}$$

where the first component, called integrated volatility, is from the continuous component of (1), and the second term is the contribution from discrete jumps. Barndorff-Nielsen and Shephard (2004) consider integrated power variation of order p defined as

$$PV_{t+1}(p) = \int_t^{t+1} \sigma^p(s)ds \tag{3}$$

where $0 < p \leq 2$. Clearly $PV_{t+1}(2)$ is integrated volatility.

To consider estimation of these quantities, we normalize the daily time interval to unity and divide it into h periods. Each period has length $\Delta = 1/h$. Then define the Δ period return as $r_{t,j} = p(t + j\Delta) - p(t + (j-1)\Delta)$, $j = 1, ..., h$. Note that the daily return is $r_t = \sum_{j=1}^{h} r_{t,j}$. Barndorff-Nielsen and Shephard (2004) introduce the following estimator called *realized power variation* of order p defined as

$$RPV_{t+1}(p) = \mu_p^{-1}\Delta^{1-p/2} \sum_{j=1}^{h} |r_{t,j}|^p, \tag{4}$$

where

$$\mu_p = E\,|\mu|^p = 2^{p/2}\frac{\Gamma(\frac{1}{2}(p+1))}{\Gamma(\frac{1}{2})}$$

for $p > 0$ where $\mu \sim N(0,1)$ and they show that as $h \rightarrow \infty$,

$$RPV_{t+1}(p) \xrightarrow{p} \begin{cases} \int_t^{t+1} \sigma^p(s)ds & p \in (0,2) \\ QV_{t+1} & p = 2 \\ \infty & p > 2. \end{cases} \tag{5}$$

When $p \in (0,2)$, $RPV_{t+1}(p)$ is robust to jumps. Note that for the special case of $p = 2$ equation (4) becomes

$$RPV_{t+1}(2) = \sum_{j=1}^{h} {r_{t,j}}^2 = RV_{t+1} \tag{6}$$

and we have the realized volatility (RV_{t+1}) estimator discussed in Andersen, Bollerslev, Diebold and Labys (2001), Barndorff-Nielsen and Shephard (2002b), and Meddahi (2002). To avoid confusion we refer to $RPV(p)$ for $p < 2$ as realized power variation, and to (6) as realized volatility (RV).

A second estimator considered in Barndorff-Nielsen and Shephard (2004) is realized bipower variation which is,

$$RBP_{t+1} \equiv \mu_1^{-2} \sum_{j=2}^{h} |r_{t,j-1}||r_{t,j}|, \tag{7}$$

where $\mu_1 = \sqrt{2/\pi}$. As $h \rightarrow \infty$, $RBP_{t+1} \xrightarrow{p} \int_t^{t+1} \sigma_t^2(s)ds$, which also excludes the effect of jumps.

As advocated by Andersen, Bollerslev, Diebold and Labys (2003), integrated volatility which can be approximated very closely by realized volatility under weak regularity conditions, provides a natural ex post measure of the variance of the return process. For specifications including jumps, and a predetermined mean process, realized volatility will be an unbiased estimator of the conditional variance of returns. For further details on the relationship between RV and the second moments of returns see Barndorff-Nielsen and Shephard (2002a,2005) and Meddahi (2003).

3 Linear Models of Realized Volatility

It has been noted by Andersen, Bollerslev, Diebold and Labys(2001) and Andersen, Bollerslev, Diebold and Ebens (2001) that the logarithm of foreign exchange and equity volatility is approximately bell shaped. Therefore, we consider the following linear regression models of the form

$$y_{t,h} = W_{t-1}\theta + Z_{t-1}\gamma + u_{t,h} \quad u_{t,h} \backsim NID(0, \sigma^2). \tag{8}$$

where

$$y_{t,h} = \log(RV_{t,h}), \text{ and } RV_{t,h} = \frac{1}{h}\sum_{i=1}^{h} RV_{t+i-1}. \tag{9}$$

$RV_{t,h}$ is the h-step ahead average realized volatility which is an estimate of the corresponding average quadratic variation over the same time interval. A number of popular continuous time models are shown by Meddahi (2003) to imply ARMA forms for realized volatility. Therefore, we consider AR models as our base specification using lags of $\log(RV_t)$ as in,

$$W_{t-1} = [1, y_{t-1,1} \cdots , y_{t-p,1}]$$

and other regressors contained in Z_{t-1}, which may include lags of $\log(RPV_t(p))$, $0 < p < 2$, $\log(RBP_t)$, and a leverage effect parameterized as $\log(RV_{t-1})I(r_{t-1} < 0)$, where r_{t-1} is the daily return defined in Section 2, and $I(\cdot)$ is the indicator function. When $h > 1$ the overlapping nature of $y_{t,h}$ will result in a moving average autocorrelation structure in the innovations, however, most of the specifications we consider include an ample number of lags of the dependent variable which will account for this. Section 6 provides a list of the models included in the Bayesian model averaging. In this paper we evaluate the predictive distribution of future log-volatility $y_{t+1,h}$, as well as point forecasts of the levels of volatility, $RV_{t+1,h}$ based on time t information. How these quantities are computed is discussed below.

3.1 Bayesian Estimation

To conduct formal model comparisons and model averaging we use Bayesian estimation methods. Our model is the standard Normal linear regression

$$Y = X\beta + \epsilon \tag{10}$$

where $X = [W\ Z]$, $\beta = [\theta, \gamma]^T$ and $\epsilon \sim N(0, \sigma^2 I)$. By Bayes rule, the prior $p(\beta, \sigma^2)$, given data and a likelihood function $p(Y|\beta, \sigma^2)$, are updated to the posterior distribution,[1]

$$p(\beta, \sigma^2|Y) = \frac{p(Y|\beta, \sigma^2)p(\beta, \sigma^2)}{\int\int p(Y|\beta, \sigma^2)p(\beta, \sigma^2)d\beta d\sigma^2}. \tag{11}$$

We specify independent conditionally conjugate priors for β and σ^2. They are

$$p(\beta) \sim N(b_0, B_0), \quad \sigma^2 \sim IG\left(\frac{v_0}{2}, \frac{s_0}{2}\right), \tag{12}$$

where $IG(\cdot, \cdot)$ denotes the inverse Gamma distribution. Although the posterior is not a well known distribution we can obtain samples from the posterior based on a Gibbs sampling scheme. For instance, the conditional distributions used in sampling are

$$\beta|Y, \sigma^2 \sim N(M, V) \tag{13}$$

[1] To minimize notation we suppress the conditioning on X in the following derivations.

where

$$M = V^{-1}\left(\frac{X'Y}{\sigma^2} + B_0^{-1}b_0\right),\ V = \left(\frac{X'X}{\sigma^2} + B_0^{-1}\right)^{-1},$$

and

$$\sigma^2|Y,\beta \sim IG\left(\frac{v}{2},\frac{s}{2}\right) \tag{14}$$

where

$$v = T + v_0,\ s = (Y - X\beta)'(Y - X\beta) + s_0.$$

Good introductions to Gibbs sampling and Markov chain Monte Carlo methods can be found in Geweke(1997) and Chib (2001). Formally, Gibbs sampling involves the following steps. Choose a starting value, $\beta^{(0)}$ and $\sigma^{2(0)}$ and number of iterations N, then iterate on

- Take a random draw, $\beta^{(i)}$ from $p(\beta|Y,\sigma^{2(i-1)})$.
- Take a random draw, $\sigma^{2(i)}$ from $p(\sigma^2|Y,\beta^{(i)})$.

Repeating these steps N times produces the draws $\{\beta^{(i)},\sigma^{2(i)}\}_{i=1}^N$. To eliminate the effect of starting values, we drop the first N_0 draws and collect the next N. For a sufficiently large sample this Markov chain converges to draws from the stationary distribution which is the posterior distribution. From this posterior sample, any function of interest can be consistently estimated. For example,

$$\widehat{g(\beta,\sigma^2)} = \frac{1}{N}\sum_{i=1}^N g(\beta^{(i)},\sigma^{2(i)}) \tag{15}$$

is a simulation consistent estimate of $E\left[g\left(\beta,\sigma^2\right)\right]$, the posterior mean of $g(\beta,\sigma^2)$.

Forecasting a future observation y^* given the information I_t is based on the predictive density defined as

$$p(y^*|I_t) = \int\int p(y^*|\beta,\sigma^2,I_t)p(\beta,\sigma^2|I_t)d\beta d\sigma^2 \tag{16}$$

which integrates out the parameter uncertainty. In this paper we compute out-of-sample forecasts of $RV_{t+1,h}$. Since $RV_{t+1,h}$ is conditionally log-normal, we use the following calculation for the predictive mean of a particular model[2]:

$$E_tRV_{t+1,h} = \frac{1}{N}\sum_{i=1}^N \exp\left[E_t\left[y_{t+1,h}|\beta^{(i)},\sigma^{2(i)}\right] + \frac{1}{2}\mathrm{Var}_t\left[y_{t+1,h}|\beta^{(i)},\sigma^{2(i)}\right]\right] \tag{17}$$

where

$$E_t\left[y_{t+1,h}|\beta^{(i)},\sigma^{2(i)}\right] = W_t\theta^{(i)} + Z_t\gamma^{(i)}$$

and

$$\mathrm{Var}_t\left[y_{t+1,h}|\beta^{(i)},\sigma^{2(i)}\right] = \sigma^{2(i)}.$$

Note that in forecasting $RV_{t+1,h}$ for $h = 5$, and 10 we are careful to compute true out-of-sample forecasts. For instance, if we used data till time t for estimation, the last regressand would be $y_{t-h+1,h} = \log\left((RV_{t-h+1} + \cdots + RV_t)/h\right)$, then the forecast is computed based on this information set as $E_tRV_{t+1,h}$.

[2]The use of the analytical moments in (17) has been referred to as a Rao-Blackwellization, and can be expected to be more accurate than a fully simulated approach to calculating the posterior mean.

3.2 Model Averaging

In a Bayesian context it is straightforward to entertain many models and combine their information and forecasts in a consistent fashion. There are several justifications for model averaging. Min and Zellner (1993) show that the model average minimizes the expected predicted squared error when the models are exhaustive. More generally, Fernández-Villaverde and Rubio-Ramíez (2004) show that even when all models are misspecified and/or non-nested, Bayes factors consistently choose the best model according to the Kullback-Leibler measure. This implies that asymptotically the model average will put a probability of 1 on the *best* model. For an introduction to Bayesian model averaging see Koop (2003). Suppose we have K different models, M_k, $k = 1, \ldots, K$ and data $Y_t = \{y_1, \ldots, y_t\}$, the probability of model M_k given Y_t is,

$$p(M_k|Y_t) = \frac{p(Y_t|M_k)p(M_k)}{\sum_{i=1}^{K} p(Y_t|M_i)p(M_i)}. \tag{18}$$

where

$$p(Y_t|M_k) = \int\int p(Y_t|\beta, \sigma^2, M_k)p(\beta, \sigma^2|M_k)d\beta d\sigma^2. \tag{19}$$

In this equation, $p(M_k)$ is the prior model probability, $p(Y_t|M_k)$ is the marginal likelihood, $p(Y_t|\beta, \sigma^2, M_k)$ the likelihood and $p(\beta, \sigma^2|M_k)$ the prior for model M_k.

Given the information set at time t denoted by I_t, we can predict future unobserved data (or function of) y^* using the predictive density $p(y^*|I_t)$ of the model average, defined as,

$$p(y^*|I_t) = \sum_{k=1}^{K} p(y^*|I_t, M_k)p(M_k|I_t), \tag{20}$$

where each model's predictive density is defined in (16). The predictive mean of the function $g(y^*)$ is,

$$E[g(y^*)|I_t] = \sum_{k=1}^{K} E[g(y^*)|I_t, M_k]p(M_k|I_t), \tag{21}$$

which is a weighted average, using the model probabilities, of model specific predictive means. In our calculations $g(y^*) = RV_{t+1,h}$, and $E[g(y^*)|I_t, M_k]$ is calculated from (17).

3.3 Model Comparison

The Bayesian approach allows for the comparison and ranking of models by Bayes factors or posterior odds. In contrast to the classical likelihood based test statistics, model comparison can be conducted for both nested and non-nested models. The Bayes factor for M_0 versus M_1 based on the data Y_t is defined as

$$BF_{01} = p(Y_t|M_0)/p(Y_t|M_1) \tag{22}$$

which is the ratio of marginal likelihoods and summarizes the evidence for model M_0 versus M_1. An advantage of using Bayes factors for model comparison is that they automatically include Occam's razor effect in that they penalize highly parameterized models that do not deliver improved predictive content.[3] Kass and Raftery (1995) recommend considering twice the logarithm of the Bayes factor for model comparison, as it has the same scaling as the likelihood ratio statistic. They suggest the following interpretation of support for M_0 based on $2\log BF_{01}$: 0 to 2 not worth more than a bare mention, 2 to 6 positive, 6 to 10 strong, and greater than 10 as very strong.

There is a long tradition in the Bayesian literature of comparing models based on predictive distributions (Box (1980), Gelfand and Dey (1994), and Gordon (1997)). In a similar fashion to the Bayes factor which is based on all the data, we can compare the performance of models on a specific out-of-sample period. In this case the *predictive likelihood* (Geweke (1995,2005)) is defined for the data $y_s, ..., y_t$, $s < t$ as

$$p(y_s, ..., y_t|I_{s-1}, M_k) = \int\int p(y_s, ..., y_t|\beta, \sigma^2, I_{s-1}, M_k)p(\beta, \sigma^2|I_{s-1}, M_k)d\beta d\sigma^2. \tag{23}$$

This is the same as the marginal likelihood if $s = 1$, and $t = T$. The predictive likelihood is the predictive density evaluated at the realized outcome $y_s, ..., y_t$. Note that integration is performed with respect to the posterior distribution based on the data I_{s-1}.

Both the marginal likelihood and the predictive likelihood contain the out-of-sample prediction record of a model, making them the central quantity of interest for model evaluation (Geweke and Whiteman (2005)). For example, (23) is simply the product of the individual predictive likelihoods:

$$p\left(y_s, ..., y_t|I_{s-1}, M_k\right) = \prod_{j=s}^{t} p\left(y_j|I_{j-1}, M_k\right). \tag{24}$$

Models with good past predictions have larger predictive likelihoods and receive larger weights in the model average.

In this paper we report estimates of the marginal likelihood for the full sample of data and the predictive likelihood corresponding to an out-of-sample period in which point forecasts are also investigated. In exactly the same way as a Bayes factor is defined, a *predictive* Bayes factor can be computed as $PBF_{01} = p(y_s, ..., y_t|I_{s-1}, M_0)/p(y_s, ..., y_t|I_{s-1}, M_1)$.

3.4 Calculations

Many of the above results require the calculation of the marginal likelihood or the predictive likelihood for each model. One approach to obtain these quantities is due to Geweke (1995). This method incorporates the information step-by-step and is useful when recursive out-of-sample forecasts are calculated. It is less computationally demanding than other methods such as Gelfand and Dey (1994) that require a full evaluation of the likelihood function for every draw from the posterior simulator and it is also very accurate.[4]

[3] For the advantages of the use of Bayes factors see Koop and Potter (1999),

[4] For instance, we can replicate our marginal likelihood estimates using the Gelfand and Dey approach.

Each of the individual terms of the right hand side of (24) can be estimated consistently from the Gibbs sampler output as

$$p\left(y_j|I_{j-1},M_k\right)\approx\frac{1}{N}\sum_{i=1}^{N}p\left(y_j|\beta^{(i)},\sigma^{2(i)},I_{j-1},M_k\right)\tag{25}$$

where $p(y_j|\beta^{(i)},\sigma^{2(i)},I_{j-1},M_k)$ denotes the normal density with mean $X_{j-1}\beta^{(i)}$ and variance $\sigma^{2(i)}$, evaluated at y_j, and the Gibbs sampler draws are obtained using the data I_{j-1}. Computing the predictive likelihood for all observations and models allows us to calculate model probabilities in (18) and the predictive likelihood (marginal likelihood) for the model average using a similar expression to (20) where the predictive density is replaced by the predictive likelihood estimates. Note that the predictive likelihood of the model average must be bounded above by the largest predictive likelihood from the individual models.

4 Benchmark Models

Besides several AR specifications we include two versions of the heterogeneous autoregressive model (HAR) of realized volatility by Corsi (2004). Corsi (2004) shows that this model can account for many of the features of volatility including long-memory. In formal model comparisons we consider a logarithmic version (HAR-log) implemented by Andersen, Bollerslev, and Diebold (2003),

$$y_{t,h}=\beta_0+\beta_1y_{t-1,1}+\beta_2y_{t-5,5}+\beta_3y_{t-22,22}+\gamma y_{t-1,1}I(r_{t-1}<0)+\epsilon_t,\ \ \epsilon_t\sim NID(0,\sigma^2).\tag{26}$$

This model postulates three factors that affect volatility: daily log-volatility $y_{t-1,1}$, weekly log-volatility $y_{t-5,5}$ and monthly log-volatility $y_{t-22,22}$. Compared to Andersen, Bollerslev, and Diebold (2003), we have not included a jump term, but we have included a leverage term in the case of equity. For FX volatility we omit the leverage effect, $\gamma=0$. We also include a levels version (HAR), in which y in (26) is replaced by RV, when we compute out-of-sample loss functions for forecasting average realized volatility. Once again a leverage term is included for equity. The methods discussed above for parameter estimation, forecasting and calculation of the predictive likelihood are applicable to these models.

The final comparison model is an ARFIMA(p,d,0) specification in log-volatility,

$$\Phi(L)(1-L)^d(y_{t,1}-\mu)=\epsilon_t,\ \ \epsilon_t\sim NID(0,\sigma^2),\tag{27}$$

where short-run autoregressive components are modeled through the lag polynomial $\Phi(L)=(1-\phi_1L-\cdots-\phi_pL^p-\gamma I(r_{t-1}<0)L)$. γ captures the impact of a leverage effect for equity volatility and we restrict estimation to $0\leq d<1/2$. For a $d>0$ the autocorrelation function decays at a hyperbolic rate and offers an intermediate case between the exponential decay of short memory and the infinite persistence of a unit root. We follow Beran (1995) and Chung and Baillie(1993) and consider joint estimation of this model by using a truncated expansion of the fractional differencing operator $(1-L)^d$.[5] Chung and Baillie(1993) show this approximation to work well for low order models in a maximum likelihood context. Since we are dealing with large datasets of several thousand observations the finite expansion of $(1-L)^d$ should provide a sufficiently accurate likelihood for our purposes.

Bayesian estimation is more involved for the ARFIMA model. In this paper we use a random walk Metropolis-Hasting routine to jointly sample the full parameter vector from the posterior distribution. Joint sampling is known to reduce the dependence in the chain output and provide more accurate posterior estimates. Let $\theta^{(i)}$ denote the last parameter vector drawn and $\pi(\theta)$ the posterior density. Then a new proposal is obtained as follows

- propose $\theta \sim q(\theta^{(i)}, \theta)$
- calculate $\alpha = \min\{\pi(\theta)/\pi(\theta^{(i)}), 1\}$
- Set $\theta^{(i+1)} = \theta$ with probability α and otherwise $\theta^{(i+1)} = \theta^{(i)}$.

Here $q(\theta^{(i)}, \theta)$ is a fat tailed multivariate mixture of 2 Normals both centered around $\theta^{(i)}$ where one component has a variance-covariance matrix 100 times the other. The latter distribution is sampled from with probability 0.1. To capture the parameter correlations in the posterior and ensure efficient proposals, we calibrate the variance-covariance of the mixture with a series of preliminary runs that provide an approximate estimate to this matrix.[6] As a result, our proposal density efficiently explores the posterior distribution and the dependence in the output from the chain quickly diminishes.[7] Even so, posterior simulation remains computationally expensive compared to linear models. As such we do not update the posterior as we do with the other models when forecasting. Instead we fix the posterior based on the in-sample data. This also means we do not obtain a full decomposition of the marginal likelihood as discussed above. Instead we compute the marginal likelihood based on the simulation output using the Gelfand and Dey (1994) method as implemented by Geweke (1999). We can always compute the predictive likelihood for observations $y_s, ..., y_t$ $s < t$ through the identity $p(y_s, ..., y_t | y_1, ..., y_{s-1}) = p(y_1, ..., y_{s-1}, y_s, ..., y_t)/p(y_1, ..., y_{s-1})$ which requires us to separately estimate the marginal likelihood of $y_1, ..., y_t$ and $y_1, ..., y_{s-1}$ (Geweke (2005)). This is what we do to compute the predictive likelihood and allow for a training sample as discussed in the results section.

Finally, note that in contrast to the other models the ARFIMA is a one period model. Therefore, to calculate the predictive mean we take a random draw from the posterior, and simulate the model forward. This entails simulating the short memory component forward, applying the long memory operator with $-d$ and then exponentiating to obtain

[5]Note that $(1-L)^d = \sum_{j=0}^{\infty} \pi_j L^j$ where $\pi_0 = 1$, and $\pi_j = -\frac{d-j+1}{j}\pi_{j-1}$, $\forall j \geq 1$. We replace $(1-L)^d(y_{t,1} - \mu)$ in (27) with $\sum_{j=0}^{t-1} \pi_j (y_{t-j,1} - \mu)$.

[6]The very first estimate is based on a single move random walk proposal in which each parameter is sampled conditional on all other parameters.

[7]For instance, the autocorrelation function of the sampled parameter values are close to 0 before 50 lags.

levels of $\tilde{RV}_{t+i}$, $i = 1, ..., h$, from which a simulated value of $\tilde{RV}_{t+1,h}$ is obtained.[8] The average of a large number of draws is used as the predictive mean, $E_t RV_{t+1,h}$.

5 Data

High frequency foreign exchange data on the JPY-USD and DEM-USD spot rates are from Olsen & Associates. The data is updated from Maheu and McCurdy (2002). We adopt the official conversion rate between DEM and Euro after January 1, 1999 to obtain the DEM-USD rate. Spot rates on a five minute grid for a 24 hour day are constructed from the nearest logarithmic middle prices. The end of a day is defined as 21:00:00 GMT and the start as 21:05:00 GMT. Weekends (21:05:00 GMT Friday - 21:00:00 GMT Sunday) and slow trading dates (December 24-26, 31 and January 1-2) and the moving holidays Good Friday, Easter Monday, Memorial Day, July Fourth, Labor Day, Thanksgiving and the day after were removed. Also a small number of slow trading days were removed. From this remaining data, 5 minute returns where constructed, from which a MA(4) and MA(10) filter was applied to JPY-USD and DEM-USD returns respectively to remove any market microstructure effects that may bias the daily volatility estimators.[9] From the filtered 5 minute return data, daily returns, realized volatility, bipower variation and power variation of order 0.5 and 1, were constructed. The sample period for JPY-USD is from December 16, 1986 to December 31, 2002 (4001 observations), while for the DEM-USD it is November 4, 1986 to December 31, 2002 (4026 observations). For both data series we reserve the first 35 observation as startup values for the models. This results in the final FX data of 3966 observations (JPY-USD) and 3991 observations (DEM-USD).

For equity we consider the S&P 500 index by using the Spyder (Standard & Poor's Depository Receipts), which is an Exchange Traded Fund that represents ownership in the S&P 500 Index. The ticker symbol is SPY. Since this asset is actively traded, it avoids the stale price effect of the S&P 500 index. The Spyder price transaction data are obtained from the Trade and Quotes (TAQ) database. After removing errors from the transaction data[10], a 5 minute grid from 9:30 to 16:00 was constructed by finding the closest transaction price before or equal to each grid point time. The first observation of the day occurring just after 9:30 was used for the 9:30 grid time. From this grid, 5 minute intraday log returns are constructed. Next intraday periodicities were removed using a MA(1) model on returns. Finally, this filtered return data was used to construct daily returns, realized volatility, bipower variation and power variation of order

[8] In the case of equity there is the added complication of the leverage effect. For $h > 1$ we simulated returns based on the assumption they follow $N(0, \tilde{RV}_{t+i})$ where $\tilde{RV}_{t+i}$ is the simulated value from the model.

[9] See Bandi and Russell (2005), Hansen and Lunde (2006), Oomen (2005) and Zhang et al. (2005) for more details on the effects of market microstructure noise on volatility estimation.

[10] Data was collected with a TAQ correction indicator of 0 (regular trade) and when possible a 1 (trade later corrected). We also excluded any transaction with a sale condition of Z, which is a transaction reported on the tape out of time sequence, and with intervening trades between the trade time and the reported time on the tape. We also checked any price transaction change that was larger than 3%. A number of these were obvious errors and were removed.

0.25,0.5,0.75,1.0,1.25,1.5, and 1.75. The final data ranges from January 29, 1993 to March 30, 2004, and conditioning on the first 35 observations leaves 2778 observations.

Statistics for daily logarithmic realized volatility for all three data series are presented in Table 1. The distributions are approximately bell shaped. The time series movements of volatility are shown in Figure 1.

6 Results

6.1 Choice of prior

For our time series models, the coefficients of the regressors do not have a clear economic meaning. It is difficult to set a reasonable "informative" prior. On the other hand, when priors are very diffuse, model probabilities can be very small. This can be a problem for many of our highly parameterization models. These models will tend to provide imprecise predictions early in the sample, even though they can perform much better given accurate parameters values. As a result, poor performance at the start of the sample will diminish their contributions to the model averaging forecasts in the latter part of the sample. To avoid these issues we follow Geweke (1995) among others and use a training sample of data to set an informative prior for each model. The information in this presample of data, which is common to all models, along with our proto-prior is combined to form an updated prior. We first specify highly uninformative proto-priors, then assign some data as the training sample.

We do Bayesian model comparison, and model averaging conditional on the updated prior. For the linear models proto-priors are specified as $\beta \sim N(0, 100I)$, and $\sigma^2 \sim IG(v_0/2, s_0/2)$, where $v_0 = 5,\ s_0 = 5$. For β this is highly uninformative while the proto-prior on σ^2 represents a wide range of plausible values for volatility data. We use the same priors for the ARFIMA model parameters and assume d is uniformly distributed on the interval $(0, .5)$.

In our model averaging exercises, we set the first 1000 observations as the training sample for each model. After this, the prior model probabilities are set to $P(M_i) = 1/K,\ i = 1, ..., K$, and are updated based on (18). For the linear models the first 100 Gibbs draws were discarded and the next 5000 were collected for posterior inference. The output from the Gibbs sampler mixed very well with a fast decaying autocorrelation function. We obtain a complete decomposition of the marginal likelihood and update all estimates through time for forecasting. For the ARFIMA model we discard the first 5000 draws and collect the next 5000 for posterior inference from the final run. Due to the computational cost associated with estimation we do not update the posterior in out-of-sample forecasts. We compute the marginal likelihood as discussed in Section 4.

6.2 Foreign Exchange Volatility

Before we begin a more detailed discussion of the estimation and forecasting results for FX volatility we list our results. The model specifications included in the model average are listed in Table 2, marginal likelihood and predictive likelihood estimates are recorded in Tables 3 - 4, selected model estimates in Table 5, and out-of-sample forecast results in Tables 6 and 7. The cumulative probabilities for selected models are displayed in Figures 2 and 3. For both currencies, the first 3000 observations are split into the training sample (1–1000) and an estimation sample (1001-3000). This leaves an out-of-sample period of February 17, 1999 – December 31, 2002 (966 observations) for the JPY-USD, and January 14, 1999 – December 31, 2002 (991 observations) for the DEM-USD.

The models that enter into the model average are selected based on augmenting autoregressive models of $\log(RV_t)$, and are listed in Table 2. They are not an exhaustive list of all the potential specifications, but do represent combinations that we felt worth study, and being linear are simple to estimate. We have not included as many configurations with bipower variation since our initial analysis found this regressor to have low forecasting value.

Table 3 lists the model specifications for the JPY-USD, the full sample marginal likelihood and the out-of-sample predictive likelihood estimates for each of the forecast horizons $h = 1, 5$, and 10.[11] This corresponds to daily, weekly and biweekly average volatility. The table includes results for the model average (MA), HAR-log, and ARFIMA models discussed in Section 4.[12] Models 1 – 5 are pure autoregressive specifications of realized volatility, while models 6 and 7 allow for only lags of realized bipower variation. All remaining models include realized power variation regressors of order 0.5 or 1. Models 6 – 13 have no lags of realized volatility.

The results for $h = 1$ show the pure AR specifications to be strongly dominated by models which include power variation terms. For instance, based on the marginal likelihood values for the JPY-USD, the best pure AR model is the AR(20) (model 5) with a value of -1830.9 while every model that includes at least one RPV term provides a dramatic improvement.[13] The best model is 32 with a value of -1761.8, and has 10 lags of RV and 5 lags of RPV(0.5). There is strong support for the realized power variation regressors, since $2\log(BF)$ for model 32 against 5 is 138.2. Similar results are obtained for the longer horizons of $h = 5, 10$ with some strengthening of the pure AR models. Based on the marginal likelihood values for $h = 5$, and 10 the model with 10 lags of RPV(1.0) performs best, however, based on the predictive likelihood the pure AR(20) model in RV is preferred.

To see the improvements for different h consider the AR(10) (model 3) and compare it to an AR(10) with 1 extra RPV term (models 35 and 36) for $h = 1, 5, 10$. In each case the addition of the RPV regressor provides an improvement over the pure AR(10) specification. The gains from including RPV appear to be greatest for small h.

The model average (MA), the HAR-log specification, and an ARFIMA(5,d,0) are listed in the last three rows of the table.[14] The model average is close to the best model for all forecast horizons based on the marginal likelihood or predictive likelihood values.

[11] Bayes factors or predictive Bayes factors can be used to compare any of the models in these tables.

[12] Note that the model average does not include the HAR-log or ARFIMA model.

[13] The exception being model 14 which does not contain sufficient lags to account for the dynamics of volatility.

[14] The ARFIMA(5,d,0) specification is used in Andersen, Bollerslev, Diebold and Labys (2003) for a similar FX data series. The posterior mean of d was 0.45 based on the in-sample data.

For example, based on rank the MA places 7, 14, 4, 3, 3, and 4 in the criteria listed in Table 3. The HAR-log model is in general not very competitive with the majority of alternatives, although its performance improves as the forecast horizon lengthens. The ARFIMA is better than the HAR-log and any of the pure autoregressive models (1-5), but there are still dramatic improvements found in the linear specifications which include power variation.

Figure 2, shows selected model probabilities through time. The first panel shows considerable changes in model performance and hence probability through time for the top models. For the JPY-USD $h = 5$ case it is interesting to note that model 9 has the highest probability for most of the sample, however, model 5 which has the highest predictive likelihood over the out-of-sample period has essentially 0 weight except at the start of the sample prior to 1992. Hence model selection based on the out-of-sample period alone may be misleading. What is clear from the plots and Table 3 is that no single model always dominates.

Now we turn to the results for the DEM-USD listed is Table 4. The results are similar to the JPY-USD and show large improvements for models with RPV terms at $h = 1$. Based on the marginal likelihood the models 36, 38 and 39 all improve upon the AR specifications (models 1-5) for $h = 1, 5, 10$. Each of these models adds RPV terms to an AR(10).

We see that as the forecast horizon lengthens the AR(20) (model 5) specification performs best according to the predictive likelihood. However, there are several models with power variation that have performance quite close. For instance, models $36 - 38$ are quite competitive with the AR(20) model. Like the JPY-USD results, the bipower regressors do not appear to offer any improvements for models that include them.[15] Including 1 lag of power variation in an AR(10) appears to be a useful specification as model 36 performs well over $h = 1, 5$, and 10, and dominates the HAR-log model.[16]

The model average is very near the top model for all criteria and forecast horizons. Based on the marginal and predictive likelihood values in the table, the model average ranks as 3, 9, 3, 7, 4, and 6 respectively out of 42 models. The model average out performs both the HAR-log and ARFIMA(5,d,0) specification for each h. The model probabilities for DEM-USD $h = 10$ in Figure 3 give a good illustration of the changing fortunes of individual models through time. The model average combines the best models available at any point in time, based on their past predictive record.

To see the effect that power variation has on an autoregressive model, Table 5 reports full sample posterior estimates for an AR(5), AR(5) with RPV(0.5), and AR(5) with RPV(1) terms respectively, for $h = 1$.[17] The coefficients on realized power variation are both large and accurately estimated. The autoregressive coefficients diminish when the RPV terms are added. This shows that RPV contains considerable information on the persistence of RV. Bayes factors favor the AR(5) model with a power variation regressor.

[15] A direct comparison of model 3 (AR(10) in $\log(RV_t)$) and model 7 (AR(10) in $\log(BPV_t)$) in Tables 3 and 4 yields mixed results, with no clear winner.

[16] This is not true for the JPY-USD where the HAR-log is preferred for $h = 5, 10$ according to the predictive likelihood.

[17] These are models 2, 20 and 21 in Table 3.

Next we consider the out-of-sample point forecasts of average volatility based on the predictive mean. The out-of-sample period corresponds exactly to the period used to calculate the predictive likelihood. Tables 6 and 7 report a summary of the results for both currencies. These tables list the root mean squared forecast error (RMSE) and the R^2 obtained from a regression of average realized volatility on a constant and a model forecast. Forecast performance is listed for the best model an AR(5), AR(20), ARFIMA, HAR, HAR-log, the model average, and the worst model. The rank is also reported.

Models which include power variation for $h = 1$ perform the best. These are models 15 and 38 for the JPY-USD and DEM-USD respectively. For $h = 5$, and 10 the best models are pure AR models of RV and bipower variation for the JPY-USD. This is consistent with the good performance of AR models based on the predictive likelihood. In contrast to this, it is interesting to find that models with power variation are the best forecasters (model 39) for DEM-USD for $h = 5, 10$.

Based on the RMSE the model average performs well against the HAR, HAR-log, and ARFIMA models and is always close to the best model. The model average does particularly well for the DEM-USD. It should be noted that the RMSE of many of the models are not separated by much, which suggests this criteria may not be useful to discriminate among models.[18] The HAR-log model does better than the levels version, particularly for the DEM-USD volatility. This is expected since RV_t is highly skewed while $\log(RV_t)$ is approximately normal.

In summary for FX volatility we find: models with RPV terms provide dramatic improvements based on the marginal and predictive likelihoods for 1 period ahead forecasts and smaller improvements for longer forecast horizons. With one exception, bipower variation regressors offer no improvement over realized volatility regressors or power variation regressors. The Bayesian model average consistently performs well. With the exception of the model average, there is no dominate individual model for both currencies. In many cases, the benchmark models are beaten with models that include RPV terms.

6.3 S&P500 Volatility

The model specifications for S&P500 volatility are listed in Table 8, the full sample marginal likelihood and out-of-sample predictive likelihood estimates in Table 9, additional forecast results are in Table 10. The first 2000 observations are split into the training sample (1–1000) and an estimation sample (1001-2000). The out-of-sample period in which the predictive likelihood and predictive mean are calculated is February 21, 2001 to March 30, 2004 (778 observations).

Given the encouraging results in the FX application, we have expanded the number of power variation terms to $RPV(p)$, $p = .25, .5, .75, 1, 1.25, 1.5$, and 1.75, and dropped the bipower variation regressors. This results in a total of 60 models that enter into the model average. In addition, we include a HAR, HAR-log and ARFIMA(5,d,0) model each with a leverage term for comparison.

[18] An advantage of model comparison based on the predictive or marginal likelihood is that it takes into consideration the whole predictive distribution, not just one part of it such as the predictive mean.

Table 9 records the full sample marginal likelihood and the out-of-sample predictive likelihood for all model specifications. For $h = 1$ the top ranked model 51 includes 10 lags of RPV(0.5) and RPV(1) and no lags of realized volatility. This model provides a large improvement over the HAR-log model. For instance, the marginal likelihood is -1244.9 for this model and -1288.1 for the HAR-log.[19] The ARFIMA model is much stronger for the equity series as compared to the FX case. The full sample marginal likelihood results favor models with power variation terms. However based on the predictive likelihood the pure autoregressive models of RV along with a leverage term dominate for $h = 5$, and $h = 10$.

Similar to the FX results, the model average is consistently a top performer. The rank of the MA based on marginal likelihood values is 3, 5, and 2 for $h = 1, 5, 10$, respectively, while it is 5, 4, and 5 based on the predictive likelihood. Note the improvements that a leverage effect has on all of the pure autoregressive specifications (models 1-10) for each forecast horizon $h = 1, 5, 10$.

The performance of the out-of-sample forecasts (predictive mean) for average realized volatility is summarized in Table 10. Compared to the ranking in Table 9 for $h = 1$, a very similar specification with RPV(p) terms of $p = 0.5, 1$, and 1.5 has the smallest RMSE. The model average is quite close to the best model, while both HAR models fall behind. In contrast to the predictive likelihood ranking for $h = 5$, and 10, models which include RPV terms of 1.5 and 1.75 display the best forecasts according to RMSE.[20] Once again the MA is very competitive with the best model and all benchmarks.

In summary, models with power variation provide improvements in describing volatility for the S&P 500. Power variation regressors appear to be most useful for short term forecasts of one day. They are also included in the top forecasting models for longer horizons. All of the best models include a leverage term. Over all cases, the MA performs well, better than any individual model.

7 Discussion

This paper uses Bayesian methods to evaluate the importance of additional volatility instruments in modeling and forecasting realized volatility. We compare these new models which include realized power variation, to a number of standard models of log-volatility. We discuss and summarize our findings.

First, for the benckmark HAR, HAR-log and ARFIMA models of volatility, based on point forecasts the HAR-log and ARFIMA models provide similar results with the HAR specification being slightly worse. However, based on the marginal likelihood the ARFIMA model is strongly favored.

We find robust improvements in both FX and equity markets when power variation terms are included for models of one period ahead volatility. Our results also show power

[19] Similarly, the predictive likelihood supports this specification with a value of -504.7 while it is -519.3 for the HAR-log.

[20] The RMSE of the AR(15)+LE favored by the predictive likelihood is 0.8120 and 0.7802 for $h = 5$, and 10, respectively.

variation to be useful for longer horizon forecasts. A summary of the five top models according to the marginal likelihood is listed in Table 11. In most cases, significant improvements to autoregressive models can be obtained by adding RPV terms of order 0.5, 1, and 1.5. There are a few cases in which the best model only includes RPV regressors and no RV lags. In general, we can recommend an AR(10) with the inclusion of 1 - 5 lags of realized power variation. As we have shown, these specification compare favorably with the HAR, HAR-log and ARFIMA models. In constructing RV, there is little to no cost in also computing RPV of various orders. Our empirical results offer significant improvements in modeling volatility with only the use of linear models. Given the smaller gains in point forecasts, RPV is most useful for improving the forecast distribution of volatility.

Our results lead to the question of why power variation is a useful instrument. As we mentioned in Section 2, power variation is robust to jumps. Jumps are generally thought to be large outliers that may have a strong effect on model estimates and forecasts. However, bipower variation, which is also robust to jumps, shows no benefits in forecasting in our study. Forsberg and Ghysels (2004) investigate why the absolute value of returns is such a good forecaster. They demonstrate that the absolute value of returns forecasts future quadratic variation better than squared returns. They argue that improvements are due to the absolute value having higher predictability, less sampling error and a robustness to jumps. Our results show that these desirable features appear to extend to power variation terms, notably RPV(0.5), RPV(1.0), and RPV(1.5).

With all the possible models, we show the benefits to model averaging. Although not always number one, the Bayesian model average, ranked by any of the criteria studied in this paper, is always near the top in performance. Moreover, the model average is the top performer when all forecast horizons and data series are considered.[21] This is not surprising, given that Bayesian model averaging fully accounts for model risk.

References

[1] Andersen T. G., T. Bollerslev and F. X. Diebold (2003): "Some like it sooth, and some Like it rough: Untangling continuous and jump components in measure, modeling, and forecasting asset return volatility," Manuscript, University of Pennsylvania.

[2] Andersen T. G., T. Bollerslev, F. X. Diebold and H. Ebens (2001): "The distribution of realized stock return volatility," *Journal of Financial Economics*, 61, 43-76

[3] Andersen T. G., T. Bollerslev, F.X.Diebold and P. Labys (2003): "Modeling and forecasting realized volatility," *Econometrica*, 71(2), March, 579-625

[4] Andersen T .G., T. Bollerslev, F. X. Diebold and P. Labys (2001): "The distribution of exchange rate volatility," *Journal of the American Statistical Association*, 96, 42-55.

[21] The models for the JPY-USD and DEM-USD are identical while the S&P 500 models do differ but are similar in spirit.

[5] Andersen T. G., T. Bollerslev and N. Meddahi (2005): "Correcting the errors: volatility forecast evaluation using high-frequency data and realized volatilities," *Econometrica*, 73(1), 279-296.

[6] Andreou E. and E. Ghysels (2002): "Rolling-sample volatility estimators: some new theoretical, simulation, and empirical results," *Journal of Business & Economic Statistics*, 20(3), 363-76.

[7] Bandi F. M. and J. R. Russell "Separating Microstructure Noise from Volatility," forthcoming *Journal of Financial Economics*

[8] Barndorff-Nielsen O. E. and N. Shephard (2005): "Variation, jumps, market frictions and high frequency data in financial econometrics", working paper, Nuffield College, University of Oxford.

[9] Barndorff-Nielsen O. E. and N. Shephard (2004): "Power and bipower variation with stochastic volatility and jumps", *Journal of Financial Econometrics*, 2(1), 1-37.

[10] Barndorff-Nielsen O. E. and N. Shephard (2002a): "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society*, Series B, 64, 253-280.

[11] Barndorff-Nielsen O. E. and N. Shephard (2002b): "Estimating quadratic variation using realised variance," *Journal of Applied Econometrics*, 17, 457-477.

[12] Beran J. (1995) "Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models", *Journal of the Royal Statistical Society*, 57, 659-672.

[13] Bollerslev T. and H. Zhou(2002): "Estimating stochastic volatility diffusion using conditional moments of integrated volatility," *Journal of Econometrics*, 109(1), 33-65.

[14] Box G. E. P. (1980): "Sampling and Bayes' inference in scientific modelling and robustness," *Journal of the Royal Statistical Society*, A, 143, 383-430.

[15] Chib S. (2001): "Markov Chain Monte Carlo Methods: Computation and Inference," Handbook of Econometrics: volume 5 (eds J.J. Heckman and E. Leamer), North Holland, Amsterdam, 3569-3649.

[16] Corsi F. (2004): "A simple long memory model of realized volatility," Working paper, University of Southern Switzerland.

[17] Chung C. F. and R. T. Baillie (1993) "Small sample bias in conditional sum-of-squares estimators of fractionally integrated ARMA models", *Empirical Economics*, 18, 791-806.

[18] Ding Z. and 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.

[19] Fernandez C. E. Ley and M. J. F. Steel,(2001): "Model Uncertainty in Cross-Country Growth Regressions", *Journal of Applied Econometrics*, 16(5), 563-576.

[20] Fernández-Villaverde J. and J.F. Rubio-Ramíez (2004): "Comparing dynamic equilibrium models to data: a Bayesian approach", *Journal of Econometrics*, 123, 153-187.

[21] Fleming J., C. Kirby and B. Ostdiek (2003): "The economic value of volatility timing using 'realized' volatility," *Journal of Financial Economics*, 67, 473-509.

[22] Forsberg L. and E. Ghysels (2004): "Why Do Absolute Returns Predict Volatility So Well? " Working paper, Department of Economics, University of North Carolina - Chapel Hill.

[23] Gelfand A. E. and D. Dey (1994): "Bayesian model choice: asymptotic and exact calculations," *Journal Royal Statistical Society*, B, 56, 501-514.

[24] Geweke J. (2005): Contemporary Bayesian Econometrics and Statistics, John Wiley & Sons Ltd.

[25] Geweke J. (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication", *Econometric Reviews* 18, 1-126.

[26] Geweke J.(1995): "Bayesian comparison of Econometric Models," Working paper, Research Department, Federal Reserve Bank of Minneapolis.

[27] Geweke J. and C. Whiteman (2005) "Bayesian Forecasting", Forthcoming in Handbook of Economic Forecasting, Edited by Elliott, Granger and Timmermann, Elsevier.

[28] Ghysels E., P. Santa-Clara and R. Valkanov (2005): "Predicting volatility: How to get most out of returns data sampled at different frequencies," forthcoming in *Journal of Econometrics*.

[29] Gordon S. (1997): "Stochastic trends, deterministic trends, and business cycle turning points", *Journal of Applied Econometrics*, 12, 411-434.

[30] Hansen P. R. and A.Lunde (2006): " Realized Variance and Market Microstructure Noise", forthcoming *Journal of Business and Economic Statistics*.

[31] Hibon M. and T. Evgeniou (2004): "To combine or not combine: selecting among forecasts and their combinations", *International Journal of Forecasting*, 21(1), 15-24.

[32] Jacobson T. and S. Karlsson (2004): "Finding good predictors for Inflation: A Bayesian Model Averaging Approach," forthcoming, *Journal of Forecasting*.

[33] Kass, R. E. and A. E. Raftery (1995): "Bayes factors and model uncertainty," *Journal of the American Statistical Association*, 90, 773-795.

[34] Koop G. (2003): Bayesian Econometrics, John Wiley & Sons Ltd.

[35] Koop G. and S. Potter (1999): "Bayes factors and nonlinearity: Evidence from economic time series," *Journal of Econometrics*, 88, 251-282

[36] Koop G. and S. Potter(2004): "Forecasting in dynamic factor models using Bayesian model averaging," *The Econometrics Journal*, 7(2), 550-565.

[37] Koopman S. J., B. Jungbacker and E. Hol (2005): "Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements," *Journal of Empirical Finance*, 12(3), 445-475.

[38] Maheu J. and T. H. McCurdy (2002): "Nonlinear Features of Realized FX Volatility," *Review of Economics and Statistics* 84(4), 668-681.

[39] Martens M., D. v. Dijk and Michiel de Pooter (2004): "Modeling and forecasting S&P 500 volatility: Long memory, structural breaks and nonlinearity", Discussion Paper, Tinbergen Institute.

[40] Meddahi N. (2002): "A Theoretical Comparison Between Integrated and Realized Volatility" *Journal of Applied Econometrics*, 17, 479-508.

[41] Meddahi N. (2003): "ARMA Representation of Integrated and Realized Variances," *The Econometrics Journal*, 6, 334-355.

[42] Min C. and A. Zellner (1993) "Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates", *Journal of Econometrics*, 56(1-2), 89-118.

[43] Oomen R. C. A. (2005) "Properties of bias-corrected realized variance under alternative sampling schemes" *Journal of Financial Econometrics*, 3 (4), 555-577.

[44] Wright J. H.(2003): "Forecasting U.S. inflation by Bayesian Model Averaging," International Finance Discussion Paper, The Federal Reserve Board.

[45] Zhang L. and P. Mykland and Y. Ait-Sahalia (2005) "A tale of two time scales: determining integrated volatility with noisy high-frequency data", forthcoming *Journal of American Statistical Association.*

Table 1: Summary statistics for $\log(RV_t)$

	JPY-USD	DEM-USD	S&P500
Mean	-0.7865	-0.8657	-0.2495
Median	-0.8276	-0.8854	-0.1671
Min	-3.5400	-3.2995	-3.2938
Max	3.5377	2.3861	3.4951
Std Dev.	0.6949	0.6225	0.9660
Skewness	0.3757	0.3287	0.0949
Excess Kurtosis	1.0404	0.8139	0.1220

JPY-USD data February 12, 1987 to December 31, 2002 (3966 observations).
DEM-USD data January 5, 1987 to December 31, 2002 (3991 observations).
S&P500 data March 22, 1993 to March 24, 2004 (2778 observations).

Table 2: Model Specifications for FX market

	Lags			
Model	RV	RPV(.5)	RPV(1)	BP
1	1	0	0	0
2	5	0	0	0
3	10	0	0	0
4	15	0	0	0
5	20	0	0	0
6	0	0	0	5
7	0	0	0	10
8	0	0	5	0
9	0	0	10	0
10	0	5	0	0
11	0	10	0	0
12	0	5	5	0
13	0	5	5	5
14	1	1	1	0
15	1	5	5	0
16	1	5	0	0
17	1	5	0	5
18	1	10	0	0
19	1	0	10	0
20	5	1	0	0
21	5	0	1	0
22	5	0	0	1
23	5	1	1	0
24	5	1	1	1
25	5	2	0	0
26	5	0	2	0
27	5	0	0	2
28	5	2	2	0
29	5	5	0	0
30	5	0	5	0
31	5	0	0	5
32	5	10	0	0
33	5	0	10	0
34	5	0	0	10
35	10	1	0	0
36	10	0	1	0
37	10	0	0	1
38	10	5	0	0
39	10	0	5	0
40	10	0	0	5

Table 3: Model Comparison, JPY-USD Volatility

	h=1		h=5		h=10	
Model	log(ML)	log(PL)	log(ML)	log(PL)	log(ML)	log(PL)
1	-2013.5	-596.5	-1688.9	-434.3	-1668.3	-413.2
2	-1841.6	-539.3	-1421.6	-354.8	-1407.1	-341.5
3	-1831.9	-537.7	-1391.6	-349.3	-1361.1	-327.7
4	-1831.4	-536.5	-1380.0	-342.9	-1337.0	-317.3
5	-1830.9	-536.0	-1372.6	**-340.0**	-1330.9	**-314.1**
6	-1842.5	-537.9	-1425.9	-352.0	-1402.6	-330.4
7	-1833.0	-536.3	-1395.5	-346.4	-1357.7	-317.6
8	-1773.0	-519.3	-1376.7	-347.9	-1361.3	-330.0
9	-1768.8	-519.1	**-1356.7**	-344.5	**-1326.4**	-320.5
10	-1777.5	-523.4	-1402.2	-364.9	-1375.8	-342.5
11	-1773.9	-523.5	-1384.8	-362.0	-1344.1	-334.3
12	-1764.1	**-513.9**	-1379.7	-345.3	-1366.9	-329.6
13	-1766.3	-513.9	-1382.4	-350.9	-1371.7	-335.3
14	-1849.9	-542.2	-1558.8	-398.0	-1549.8	-375.2
15	-1764.9	-514.2	-1380.2	-345.4	-1366.6	-328.3
16	-1776.2	-523.1	-1391.2	-356.8	-1370.9	-338.2
17	-1770.9	-518.0	-1389.4	-352.4	-1372.4	-336.1
18	-1771.8	-523.2	-1372.5	-353.6	-1337.8	-329.6
19	-1764.8	-517.4	-1357.5	-344.9	-1327.3	-320.7
20	-1766.6	-516.2	-1395.3	-351.9	-1385.0	-337.7
21	-1768.0	-514.8	-1388.9	-349.3	-1379.1	-334.3
22	-1838.4	-536.6	-1421.6	-354.3	-1406.4	-340.2
23	-1766.6	-515.3	-1390.8	-347.8	-1380.9	-332.1
24	-1766.7	-515.7	-1388.9	-347.3	-1381.5	-334.0
25	-1767.8	-516.1	-1395.2	-352.6	-1384.2	-338.0
26	-1769.0	-515.0	-1387.4	-350.1	-1376.7	-334.1
27	-1839.0	-536.6	-1422.6	-354.8	-1407.1	-340.3
28	-1767.1	-517.1	-1389.0	-347.5	-1378.6	-330.0
29	-1768.6	-516.5	-1388.1	-351.9	-1373.0	-336.2
30	-1768.6	-515.6	-1378.3	-348.7	-1362.7	-330.2
31	-1839.5	-535.8	-1423.6	-354.4	-1406.0	-337.6
32	**-1761.8**	-515.5	-1365.9	-347.1	-1336.5	-326.0
33	-1763.9	-514.8	-1359.3	-344.8	-1330.2	-321.1
34	-1829.7	-533.9	-1392.5	-347.8	-1360.9	-324.2
35	-1762.3	-516.0	-1369.6	-347.2	-1343.7	-325.2
36	-1763.9	-514.8	-1364.1	-345.1	-1338.9	-322.5
37	-1828.9	-535.1	-1391.8	-349.0	-1360.8	-326.8
38	-1764.1	-515.5	-1369.0	-348.2	-1341.8	-325.9
39	-1765.2	-514.9	-1361.9	-346.0	-1334.7	-321.8
40	-1830.3	-534.4	-1394.6	-350.2	-1362.2	-326.2
MA	-1764.6	-515.6	-1360.0	-343.5	-1329.7	-317.6
HAR-Log	-1835.2	-538.7	-1384.2	-341.3	-1343.0	-314.9
ARFIMA	-1817.1	-534.7				

This table reports each model's full sample log-marginal likelihood (log(ML)) and the out-of-sample log-predictive likelihood (log(PL)). MA denotes the model average, HAR-log and ARFIMA models are discussed in Section 4. The bold entries are the largest values for the models 1-40.

Table 4: Model Comparison, DEM-USD Volatility

	h=1		h=5		h=10	
Model	log(ML)	log(PL)	log(ML)	log(PL)	log(ML)	log(PL)
1	-1886.8	-530.6	-1420.9	-360.0	-1310.2	-324.9
2	-1702.3	-466.0	-1101.4	-258.9	-972.5	-216.5
3	-1685.2	-462.7	-1050.8	-244.3	-918.6	-194.9
4	-1683.8	-458.7	-1047.3	-239.5	-915.1	-190.2
5	-1680.0	-456.6	-1051.0	**-239.3**	-921.5	**-190.2**
6	-1702.5	-464.2	-1113.2	-258.6	-989.6	-217.5
7	-1686.6	-460.8	-1065.3	-244.1	-937.3	-196.4
8	-1661.6	-450.9	-1104.0	-260.5	-991.2	-221.7
9	-1652.4	-450.2	-1068.2	-248.4	-951.1	-204.8
10	-1707.5	-463.7	-1190.4	-290.4	-1077.2	-249.7
11	-1697.0	-463.6	-1157.7	-278.9	-1039.6	-234.3
12	-1641.3	-442.2	-1070.2	-248.2	-948.5	-209.4
13	-1643.1	-442.4	-1068.4	-249.0	-947.8	-210.3
14	-1748.9	-483.2	-1286.6	-316.6	-1193.8	-290.2
15	-1642.1	-442.2	-1069.6	-248.0	-948.6	-209.6
16	-1683.8	-459.0	-1142.0	-275.5	-1031.2	-235.9
17	-1656.9	-445.4	-1105.1	-256.6	-987.0	-216.5
18	-1674.4	-458.4	-1112.0	-264.2	-995.9	-220.4
19	-1653.0	-450.3	-1064.5	-247.3	-945.5	-203.1
20	-1652.6	-446.1	-1094.2	-255.8	-970.1	-215.3
21	-1650.8	-446.4	-1087.9	-253.5	-966.3	-214.1
22	-1697.1	-463.2	-1101.6	-257.8	-974.4	-216.0
23	-1652.1	-446.9	-1079.1	-251.0	-958.8	-212.7
24	-1652.6	-447.1	-1078.3	-251.9	-958.4	-213.2
25	-1653.1	-446.3	-1094.3	-255.7	-969.9	-215.2
26	-1650.6	-446.6	-1088.4	-253.6	-966.7	-214.2
27	-1697.6	-463.1	-1103.1	-258.0	-976.0	-216.3
28	-1648.6	-445.9	-1072.3	-248.2	-953.7	-211.0
29	-1651.0	-445.0	-1095.0	-256.0	-970.9	-215.5
30	-1651.0	-446.4	-1088.6	-254.3	-967.4	-214.6
31	-1700.5	-463.5	-1109.2	-258.4	-981.5	-216.9
32	-1638.4	-443.6	-1059.8	-242.6	-928.6	-197.3
33	-1637.5	-445.0	-1047.9	-239.9	-920.0	-194.5
34	-1684.8	-460.4	-1062.1	-244.2	-929.8	-196.4
35	-1637.8	-443.9	-1043.5	-241.6	-916.1	-194.0
36	-1636.0	-444.3	**-1037.8**	-239.6	-912.7	-193.0
37	-1680.0	-459.8	-1050.9	-243.2	-920.5	-194.4
38	**-1632.3**	**-442.1**	-1042.8	-240.7	-912.3	-192.7
39	-1633.4	-443.9	-1038.6	-239.4	**-911.0**	-192.3
40	-1682.7	-460.2	-1058.1	-243.8	-925.7	-195.4
MA	-1635.7	-443.9	-1041.1	-241.4	-914.3	-192.9
HAR-Log	-1696.4	-468.4	-1069.8	-249.7	-943.0	-203.6
ARFIMA	-1680.8	-458.4				

This table reports each model's full sample log-marginal likelihood (log(ML)) and the out-of-sample log-predictive likelihood (log(PL)). MA denotes the model average, HAR-log and ARFIMA models are discussed in Section 4. The bold entries are the largest values for the models 1-40.

Table 5: Model Estimates for JPY-USD

	AR(5)		AR(5)+RPV(0.5)		AR(5)+RPV(1)	
Parameters	mean	stdev	mean	stdev	mean	stdev
Intercept	-0.1208	0.1241	0.0826	0.0194	0.0967	0.0202
θ_1	0.4729	0.0158	0.1028	0.0311	-0.1834	0.0502
θ_2	0.1324	0.0173	0.0962	0.0173	0.0835	0.0176
θ_3	0.0559	0.0176	0.0507	0.0171	0.0451	0.0171
θ_4	0.0726	0.0174	0.0560	0.0171	0.0562	0.0171
θ_5	0.1132	0.0158	0.0866	0.0157	0.0883	0.0157
γ			1.8040	0.1310	1.5280	0.1113
σ^2	0.2226	0.0050	0.2123	0.0048	0.2124	0.0048

This table reports posterior mean and standard deviations for model parameters from the following specifications,

$$y_t = \phi_0 + \sum_{i=1}^{5} \phi_i y_{t-i} + \gamma \log(RPV_{t-1}(p)) + e_t, \;\; e_t \sim NID(0, \sigma^2)$$

with $y_t = \log(RV_t)$, AR(5) sets $\gamma = 0$, AR(5)+RPV(0.5) $p = 0.5$, and AR(5)+RPV(1) $p = 1.0$. Sample period starts from February 12, 1987 to December 31, 2002 (3966 observations).

Table 6: Out-of-Sample Forecast Results, JPY-USD, $RV_{t,h}$

$h = 1$

	RMSE		R^2	
Model	value	Rank	value	Rank
Best model	0.3181(M15)	1	0.3452(M29)	1
AR(5)	0.3299	40	0.3048	37
AR(20)	0.3290	37	0.3063	35
ARFIMA	0.3321	42	0.3029	41
HAR-log	0.3308	41	0.2992	42
HAR	0.3377	43	0.2677	43
MA	0.3201	17	0.3381	17
Worst model	0.3551(M1)	44	0.2275(M1)	44

$h = 5$

	RMSE		R^2	
Model	value	Rank	value	Rank
Best model	0.2135(M4)	1	0.4642(M4)	1
AR(5)	0.2153	18	0.4590	10
AR(20)	0.2135	2	0.4635	2
ARFIMA	0.2182	36	0.4554	17
HAR-log	0.2156	22	0.4484	24
HAR	0.2220	41	0.4155	42
MA	0.2174	31	0.4394	34
Worst model	0.2438(M1)	44	0.3411(M1)	44

$h = 10$

	RMSE		R^2	
Model	value	Rank	value	Rank
Best model	0.2030(M7)	1	0.4155(M5)	1
AR(5)	0.2082	27	0.3962	14
AR(20)	0.2035	2	0.4155	1
ARFIMA	0.2099	35	0.4102	3
HAR-log	0.2047	5	0.3390	42
HAR	0.2097	34	0.3587	41
MA	0.2080	24	0.3830	30
Worst Model	0.2317(M1)	44	0.2892(M1)	44

Model forecasts are the predictive mean. This table reports root mean square error (RMSE) for the forecast errors, and the R^2 from a forecast regression of realized volatility on a constant and a model forecast from different models. The out-of-sample period goes from February 17, 1999 to December 31, 2002 (966 observations). MA is the Bayesian model average. Best(Worst) model denotes the model with the best(worst) out-of-sample performance according to RMSE or R^2 criteria. The model label appears in parenthesis. Rank = relative ranking among all models.

Table 7: Out-of-Sample Forecast Results, DEM-USD, $RV_{t,h}$

$h = 1$

	RMSE		R^2	
Model	value	Rank	value	Rank
Best model	0.4171(M38)	1	0.2144(M38)	1
AR(5)	0.4247	40	0.1830	40
AR(20)	0.4233	30	0.1882	30
ARFIMA	0.4278	42	0.1732	41
HAR-log	0.4278	42	0.1726	42
HAR	0.4411	44	0.1400	44
MA	0.4173	2	0.2136	4
Worst model	0.4411(HAR)	44	0.1400(HAR)	44

$h = 5$

	RMSE		R^2	
Model	value	Rank	value	Rank
Best model	0.2391(M39)	1	0.3426(M39)	1
AR(5)	0.2439	35	0.3154	35
AR(20)	0.2409	12	0.3328	12
ARFIMA	0.2437	33	0.3241	24
HAR-log	0.2484	40	0.2938	40
HAR	0.2611	43	0.2361	43
MA	0.2392	3	0.3421	3
Worst model	0.2640(M1)	44	0.2103(M1)	44

$h = 10$

	RMSE		R^2	
Model	value	Rank	value	Rank
Best model	0.1982(M39)	1	0.3790(M39)	1
AR(5)	0.2027	34	0.3503	35
AR(20)	0.1991	8	0.3734	8
ARFIMA	0.2012	19	0.3717	11
HAR-log	0.2071	40	0.3259	40
HAR	0.2188	42	0.2631	43
MA	0.1985	2	0.3769	2
Worst model	0.2232(M1)	44	0.2248(M1)	44

See notes to Table 6. Out-of-sample period is from January 14, 1999 to December 31, 2002 (991 observations).

Table 8: Model Specifications for S&P500

Model	Lag length of power variation, $\log(RPV(p))$ p=.25	.5	0.75	1	1.25	1.5	1.75	2	Lev.Effect
1	0	0	0	0	0	0	0	1	1
2	0	0	0	0	0	0	0	1	0
3	0	0	0	0	0	0	0	5	1
4	0	0	0	0	0	0	0	5	0
5	0	0	0	0	0	0	0	10	1
6	0	0	0	0	0	0	0	10	0
7	0	0	0	0	0	0	0	15	1
8	0	0	0	0	0	0	0	15	0
9	0	0	0	0	0	0	0	20	1
10	0	0	0	0	0	0	0	20	0
11	5	0	0	0	0	0	0	1	0
12	0	5	0	0	0	0	0	1	0
13	0	0	5	0	0	0	0	1	0
14	0	0	0	5	0	0	0	1	0
15	0	0	0	0	5	0	0	1	0
16	0	0	0	0	0	5	0	1	0
17	0	0	0	0	0	0	5	1	0
18	1	0	0	0	0	0	0	5	1
19	0	1	0	0	0	0	0	5	1
20	0	0	1	0	0	0	0	5	1
21	0	0	0	1	0	0	0	5	1
22	0	0	0	0	1	0	0	5	1
23	0	0	0	0	0	1	0	5	1
24	0	0	0	0	0	0	1	5	1
25	0	5	0	0	0	0	0	1	1
26	0	0	0	5	0	0	0	1	1
27	0	0	0	0	0	5	0	1	1
28	0	5	0	0	0	0	0	5	0
29	0	0	0	5	0	0	0	5	0
30	0	0	0	0	0	5	0	5	0
31	0	5	0	0	0	0	0	5	1
32	0	0	0	5	0	0	0	5	1
33	0	0	0	0	0	5	0	5	1
34	0	5	0	0	0	0	0	10	0
35	0	0	0	5	0	0	0	10	0
36	0	0	0	0	0	5	0	10	0
37	0	5	0	0	0	0	0	10	1
38	0	0	0	5	0	0	0	10	1
39	0	0	0	0	0	5	0	10	1
40	0	5	0	0	0	0	0	0	0

Table 8: Model Specifications for S&P500 (Continue)

	Lag length of power variation, $\log(RPV(p))$								
Model	p=.25	.5	0.75	1	1.25	1.5	1.75	2	Lev.Effect
41	0	0	0	5	0	0	0	0	0
42	0	0	0	0	0	5	0	0	0
43	0	5	0	5	0	5	0	0	0
44	0	5	0	0	0	0	0	0	1
45	0	0	0	5	0	0	0	0	1
46	0	0	0	0	0	5	0	0	1
47	0	5	0	5	0	5	0	0	1
48	0	10	0	0	0	0	0	0	1
49	0	0	0	10	0	0	0	0	1
50	0	0	0	0	0	10	0	0	1
51	0	10	0	10	0	0	0	0	1
52	0	0	0	10	0	10	0	0	1
53	0	10	0	0	0	10	0	0	1
54	0	10	0	10	0	10	0	0	1
55	0	10	0	0	0	0	0	10	1
56	0	0	0	10	0	0	0	10	1
57	0	0	0	0	0	10	0	10	1
58	0	10	0	0	0	0	0	5	1
59	0	0	0	10	0	0	0	5	1
60	0	0	0	0	0	10	0	5	1

The list of models: The first column is the model index. Column 2 to column 9 is the lags of realized power variation terms where p=0.25,0.5,0.75,1,1.25,1.5,1.75,2 respectively. The last column is the index to leverage effect(i.e, including leverage when it has value 1, no leverage effect when it has value 0).

Table 9: Model Comparison, S&P500 Volatility

	h=1		h=5		h=10	
Model	log(ML)	log(PL)	log(ML)	log(PL)	log(ML)	log(PL)
1	-1493.0	-599.3	-1295.5	-476.9	-1280.2	-466.6
2	-1495.5	-601.3	-1299.0	-478.4	-1283.4	-468.0
3	-1285.1	-515.4	-945.8	-346.3	-922.4	-339.5
4	-1288.2	-517.6	-951.1	-348.1	-928.4	-341.1
5	-1260.6	-508.4	-896.8	-333.9	-865.0	**-322.5**
6	-1263.9	-510.9	-902.8	-336.2	-871.9	-324.6
7	-1257.1	-505.7	**-889.1**	**-330.2**	-858.5	-323.3
8	-1260.2	-508.2	-894.9	-332.5	-865.5	-325.7
9	-1256.8	-505.7	-892.4	-335.2	-864.5	-333.8
10	-1260.0	-508.2	-898.3	-337.6	-871.5	-336.0
11	-1344.0	-575.7	-1052.8	-459.0	-1032.4	-476.9
12	-1323.9	-561.3	-1017.5	-434.0	-994.4	-447.9
13	-1303.8	-547.4	-982.3	-409.4	-956.7	-419.4
14	-1287.3	-535.4	-955.0	-388.9	-927.4	-395.5
15	-1275.1	-526.2	-937.2	-373.9	-908.4	-377.7
16	-1267.4	-519.7	-927.6	-364.0	-898.7	-365.7
17	-1263.1	-515.6	-924.4	-358.2	-895.9	-358.3
18	-1274.8	-521.3	-940.1	-369.9	-916.6	-379.3
19	-1273.2	-519.3	-937.7	-366.4	-913.1	-373.5
20	-1270.8	-517.8	-934.6	-364.1	-909.2	-369.6
21	-1268.1	-516.5	-931.3	-362.0	-905.1	-366.1
22	-1265.5	-515.1	-928.3	-359.8	-901.5	-362.6
23	-1263.3	-513.9	-925.8	-357.8	-898.5	-359.3
24	-1261.5	-512.7	-923.7	-355.8	-896.2	-356.1
25	-1322.8	-559.8	-1014.1	-432.8	-990.4	-447.0
26	-1285.8	-533.9	-951.3	-387.6	-923.2	-394.6
27	-1265.8	-518.2	-923.8	-362.7	-894.8	-364.8
28	-1275.6	-524.0	-952.8	-381.2	-938.2	-398.1
29	-1272.0	-522.1	-946.4	-377.6	-929.1	-391.7
30	-1268.4	-520.2	-939.1	-373.4	-918.6	-384.3
31	-1274.2	-522.3	-948.9	-379.6	-933.9	-396.8
32	-1270.6	-520.5	-942.6	-376.2	-924.7	-390.6
33	-1267.0	-518.7	-935.3	-372.2	-914.4	-383.4
34	-1257.8	-513.8	-911.9	-358.9	-891.1	-368.1
35	-1255.5	-512.6	-907.9	-356.6	-884.6	-363.5
36	-1252.8	-511.3	-902.7	-353.6	-876.8	-358.0
37	-1256.1	-511.9	-907.3	-356.8	-885.6	-366.3
38	-1253.7	-510.8	-903.2	-354.7	-879.1	-361.8
39	-1251.1	-509.5	-898.1	-351.9	-871.5	-356.5
40	-1497.9	-681.8	-1227.2	-595.7	-1172.6	-593.1

Table 9: Model Comparison, S&P500 Volatility (Continued)

	h=1		h=5		h=10	
Model	log(ML)	log(PL)	log(ML)	log(PL)	log(ML)	log(PL)
41	-1326.7	-566.5	-1005.1	-434.0	-957.9	-427.0
42	-1265.0	-516.6	-921.9	-355.2	-888.4	-347.9
43	-1260.8	-511.8	-926.9	-358.4	-917.4	-376.6
44	-1461.8	-657.8	-1183.9	-566.1	-1131.3	-564.6
45	-1319.4	-561.0	-994.1	-427.0	-946.9	-420.1
46	-1263.8	-515.0	-918.2	-353.5	-884.1	-346.3
47	-1259.2	-510.0	-922.6	-356.6	-912.3	-374.8
48	-1463.5	-659.0	-1185.9	-570.5	-1132.1	-568.8
49	-1317.4	-560.7	-986.2	-428.0	-935.2	-419.8
50	-1252.1	-511.2	-890.7	-347.5	**-849.6**	-337.0
51	**-1244.9**	**-504.7**	-895.4	-345.6	-875.1	-355.1
52	-1251.9	-509.2	-905.8	-354.1	-882.2	-362.4
53	-1250.6	-508.4	-904.0	-352.6	-882.4	-362.0
54	-1247.8	-505.3	-892.8	-343.9	-873.9	-352.1
55	-1258.6	-513.3	-915.1	-361.1	-892.4	-370.7
56	-1256.4	-512.0	-911.5	-358.9	-887.1	-367.2
57	-1253.9	-510.6	-906.8	-356.1	-880.7	-363.1
58	-1260.6	-515.6	-920.2	-366.4	-903.2	-381.1
59	-1255.9	-512.6	-909.1	-359.3	-886.2	-368.7
60	-1251.9	-510.2	-900.0	-353.4	-873.4	-358.7
MA	-1248.9	-506.9	-893.0	-331.6	-853.6	-323.9
HAR-log	-1288.1	-519.3	-947.6	-349.5	-933.1	-348.7
ARFIMA	-1249.2	-509.4				

This table reports each model's full sample log-marginal likelihood (log(ML)) and the out-of-sample log-predictive likelihood (log(PL)). MA denotes the model average, HAR-log and ARFIMA models are discussed in Section 4. The bold entries are the largest values for the models 1-60.

Table 10: Out-of-Sample Forecast Results, S&P500, $RV_{t,h}$

$h = 1$

	RMSE		R^2	
Model	value	Rank	value	Rank
Best model	1.0975(M54)	1	0.5172(M54)	1
AR(5)	1.1275	45	0.4783	59
AR(5)+LE	1.1082	10	0.4955	37
AR(20)	1.1205	34	0.4848	54
AR(20)+LE	1.1041	5	0.5004	30
ARFIMA	1.1171	30	0.4931	42
HAR-log	1.1200	33	0.4848	55
HAR	1.1153	26	0.4899	49
MA	1.1024	4	0.5129	3
Worst model	1.2787(M40)	64	0.4158(M2)	64

$h = 5$

	RMSE		R^2	
Model	value	Rank	value	Rank
Best model	0.7890(M46)	1	0.6360(M44)	1
AR(5)	0.8197	41	0.5815	56
AR(5)+LE	0.7935	4	0.6061	35
AR(20)	0.8361	53	0.5636	61
AR(20)+LE	0.8170	39	0.5825	55
ARFIMA	0.8376	54	0.5675	59
HAR-log	0.8273	47	0.5702	57
HAR	0.8335	50	0.5675	59
MA	0.8030	14	0.6047	39
Worst model	0.9592(M40)	64	0.4848(M2)	64

$h = 10$

	RMSE		R^2	
Model	value	Rank	value	Rank
Best model	0.7498(M24)	1	0.6126(M33)	1
AR(5)	0.7895	50	0.5570	55
AR(5)+LE	0.7645	21	0.5801	49
AR(20)	0.8070	56	0.5348	60
AR(20)+LE	0.7890	49	0.5525	56
ARFIMA	0.8066	56	0.5332	61
HAR-log	0.7960	52	0.5411	59
HAR	0.8235	59	0.5058	62
MA	0.7611	12	0.5832	45
Worst Model	0.8813(M40)	64	0.4716(M2)	64

Model forecasts are the predictive mean. This table reports root mean square error (RMSE) for the forecast errors, and the R^2 from a forecast regression of realized volatility on a constant and a model forecast. The out-of-sample period is from February 21, 2001 to March 30, 2004 (778 observations). MA is the Bayesian model average. Best(Worst) model denotes the model with the best(worst) out-of-sample performance according to RMSE or R^2 criteria. The model label appears in parenthesis. Rank = relative ranking among all models. LE=leverage effect.

Table 11: Model Recommendations: Top Models by Category

		JPY-USD			
	h=1		h=5		h=10
Model	Model Specification	Model	Model Specification	Model	Model Specification
32	$AR(5) + RPV_5(.5)$	9	$RPV_{10}(1)$	9	$RPV_{10}(1)$
35	$AR(10) + RPV_1(.5)$	19	$AR(1) + RPV_{10}(1)$	19	$AR(1) + RPV_{10}(1)$
33	$AR(5) + RPV_{10}(1)$	33	$AR(5) + RPV_{10}(1)$	33	$AR(5) + RPV_{10}(1)$
36	$AR(10) + RPV_1(1)$	39	$AR(10) + RPV_5(1)$	5	$AR(20)$
12	$RPV_5(.5) + RPV_5(1)$	36	$AR(10) + RPV_1(1)$	39	$AR(10) + RPV_5(1)$
		DEM-USD			
	h=1		h=5		h=10
Model	Model Specification	Model	Model Specification	Model	Model Specification
38	$AR(10) + RPV_5(.5)$	36	$AR(10) + RPV_1(1)$	39	$AR(10) + RPV_5(1)$
39	$AR(10) + RPV_5(1)$	39	$AR(10) + RPV_5(1)$	38	$AR(10) + RPV_5(.5)$
36	$AR(10) + RPV_1(1)$	38	$AR(10) + RPV_5(.5)$	36	$AR(10) + RPV_1(1)$
33	$AR(5) + RPV_{10}(1)$	35	$AR(10) + RPV_1(.5)$	4	$AR(15)$
35	$AR(10) + RPV_1(.5)$	4	$AR(15)$	35	$AR(10) + RPV_1(.5)$
		S&P 500			
	h=1		h=5		h=10
Model	Model Specification	Model	Model Specification	Model	Model Specification
51	$RPV_{10}(.5) + RPV_{10}(1)$	7	$AR(15)$	50	$RPV_{10}(1.5)$
54	$RPV_{10}(.5) + RPV_{10}(1) + RPV_{10}(1.5)$	50	$RPV_{10}(1.5)$	7	$AR(15)$
53	$RPV_{10}(.5) + RPV_{10}(1.5)$	9	$AR(20)$	9	$AR(20)$
39	$AR(10) + RPV_5(1.5)$	54	$RPV_{10}(.5) + RPV_{10}(1) + RPV_{10}(1.5)$	5	$AR(10)$
60	$AR(5) + RPV_{10}(1.5)$	8*	$AR(15)$	8*	$AR(15)$

This table lists the top 5 models for different horizons and across markets. The odd columns are the model indices in our lists of models and even columns present the best models according to marginal likelihood. The subscript of $RPV(p)$ is the lag length of the power variations. For example, $AR(10) + RPV_5(1)$ means the regressors of this model include 10 lags of $\log(RV_t)$ and 5 lags of the logarithm of power variation of order 1. All S&P 500 models include a leverage effect except model 8.

Figure 1: Time Series of Daily Log-Realized Volatility

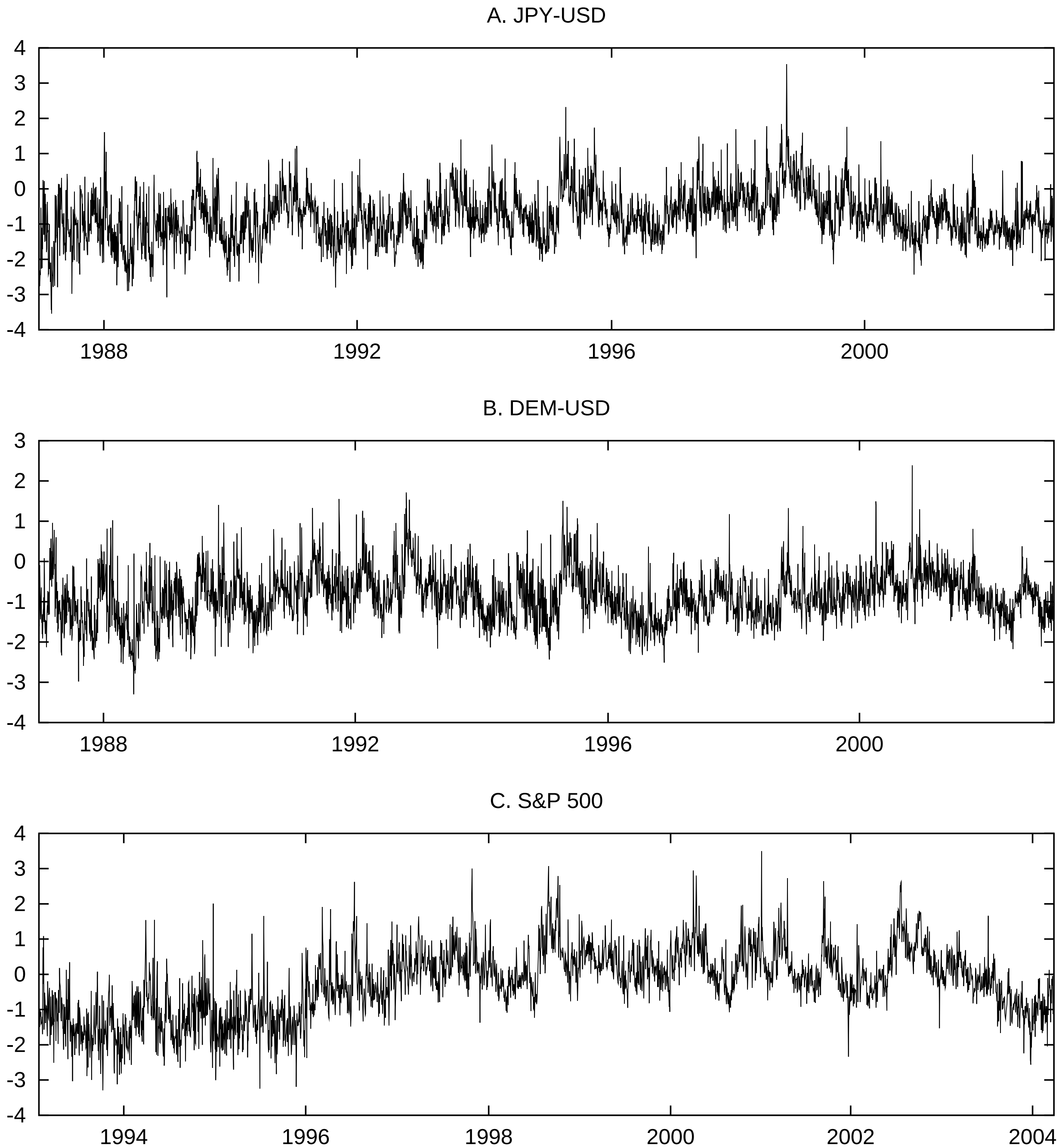

Figure 2: Cumulative Model Probabilities for JPY-USD

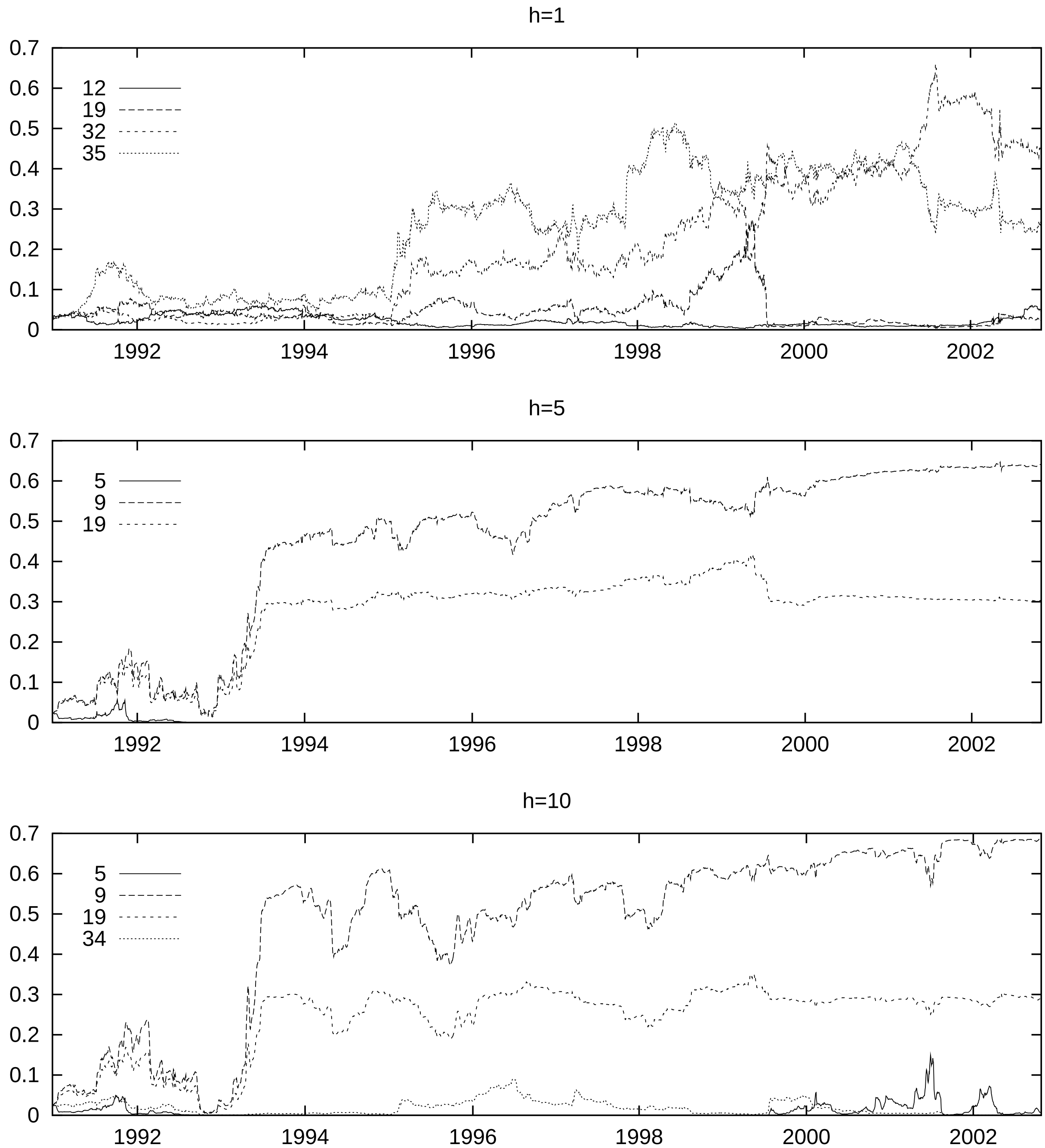

Figure 3: Cumulative Model Probabilities for DEM-USD