Do high-frequency measures of volatility improve forecasts of return distributions?

JOHN M. MAHEU and THOMAS H. MCCURDY*

June 2006

Abstract

Many finance questions require a full characterization of the distribution of returns. We propose a bivariate model of returns and realized volatility (RV); and explore which features of that time-series model contribute to superior density forecasts over horizons of 1 to 60 days out-of-sample. We investigate the importance of: the intraday information embodied in the daily RV estimates; the impact of alternative adjustments to RV estimates for market microstructure dynamics; the parsimony of the functional form for $\log(RV)$ dynamics; and the assumed distribution of return innovations. For our data and models, utilizing high-frequency information becomes less important the further one forecasts out-of-sample. In contrast to in-sample estimates, parsimony is important and Normal return innovations always produce better out-of-sample density forecasts than do fat-tailed innovations. A bivariate model of returns with Normal innovations and a mean-reverting, exponentially decaying function of $\log(RV)$ provides the best density forecasts over a range of out-of-sample horizons.

*Maheu (jmaheu@chass.utoronto.ca), Department of Economics, University of Toronto; McCurdy (tmccurdy@rotman.utoronto.ca), Joseph L. Rotman School of Management, University of Toronto, and Associated Fellow, CIRANO. We thank Guangyu Fu for excellent research assistance. We are also grateful to the SSHRC for financial support.

1 Introduction

Many finance questions require a full characterization of the distribution of returns. Examples include option pricing which uses the forecast density of the underlying spot asset, or value-at-risk which focuses on a quantile of the forecasted distribution. Once we move away from the simplifying assumptions of Normally distributed returns or quadratic utility, portfolio choice also requires a full specification of the return distribution.

The purpose of this paper is to study the accuracy of forecasts of return densities produced by alternative models. Specifically, we focus on the value that high frequency measures of ex post volatility provide in characterizing the forecast density of returns. We propose new bivariate models of returns and realized volatility and explore which features of those time-series models contribute to superior density forecasts over multiperiod horizons out-of-sample.

Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold, and Labys (2001), Andersen, Bollerslev, Diebold, and Ebens (2001), Andreou and Ghysels (2002), Barndorff-Nielsen and Shephard (2002a), Barndorff-Nielsen and Shephard (2002b), and Meddahi (2002), among others,[1] have established the theoretical and empirical properties of the estimation of quadratic variation for a broad class of stochastic processes in finance. Although theoretical advances continue to be important, part of the research in this new field has focused on the time-series properties and forecast improvements that realized volatility provide. Examples include Andersen, Bollerslev, Diebold, and Labys (2003), Andersen, Bollerslev, and Diebold (2005), Andersen, Bollerslev, and Meddahi (2004), Ghysels and Sinko (2006), Ghysels, Santa-Clara, and Valkanov (2006), Koopman, Jungbacker, and Hol (2005), Maheu and McCurdy (2002), Martens, van Dijk, and de Pooter (2003), and Taylor and Xu (1997).

Few papers have studied the benefits of incorporating RV into the return distribution. Andersen, Bollerslev, Diebold, and Labys (2003), and Giot and Laurent (2004) consider the value of RV for forecasting and Value-at-Risk. These approaches decouple the return and volatility dynamics and assume that RV is a sufficent statistic for the conditional variance of returns. Ghysels, Santa-Clara, and Valkanov (2005) find that high frequency measures of volatility identify a risk and return tradeoff at lower frequencies. Their filtering approach to volatility measurement does not provide a law of motion for volatility and therefore multiperiod forecasts cannot be computed in that setting.

We propose new bivariate models in which cross-equation restrictions on the conditional moments of RV are linked to the conditional variance of returns. The models are jointly estimated by maximum likelihood. Since our system provides a law of motion for both return and RV at the daily frequency, multiperiod forecasts of returns, RV or the density of returns are available. The models recognize that RV is an ex post measure of volatility and in general not equivalent to the conditional variance of returns. The approach extends previous work in Maheu and McCurdy (2006). Related to our work is Bollerslev, Kretschmer, Pigorsch, and Tauchen (2005) who model returns, bipower

[1]Recent reviews include Andersen, Bollerslev, and Diebold (2004), Barndorff-Nielsen and Shephard (2007).

variation and realized jumps in a multivariate setting.[2] They focus on improving the in-sample fit. We analyze the value of RV in improving the forecast density of returns.

Two types of bivariate models of returns and RV are proposed. The first model allows different components of $\log(RV)$ to have different decay rates. A second model uses a heterogenous autoregressive (HAR) specification (Corsi (2003), Andersen, Bollerslev, and Diebold (2005)) of $\log(RV)$. Both of our bivariate models allow for a so-called leverage or asymmetric effects of past negative versus positive return innovations. Both models are stationary and consistent with mean reversion in RV. We dicuss how to implement variance targeting for these specifications.

The main method of model comparison uses the predictive likelihood of returns. This is the forecast density of a model evaluated at the realized return; it provides a measure of the likelihood of the data being consistent with the model. Intuitively, better forecasting models will have higher predictive likelihood values. Therefore our focus is on the relative accuracy of the models. The forecast density of the models is not available in closed form; however, we discuss accurate simulation methods that can be used to evaluate the forecast density and the predictive likelihood.

An important feature of our approach is that we can directly compare traditional volatility specifications such as GARCH with our bivariate models of return and RV since we focus on a common criteria – forecast densities. A term structure of average predictive likelihoods, which is the average predictive likelihood for different forecast horizons, allow us to investigate the relative contributions of RV over short to long forecast horizons.

Using the average predictive likelihood, we find that market microstructure adjustments to RV along the lines of Hansen and Lunde (2006) are important, particularly for short-term density forecasts. In general, we find that our models with RV do provide improvements over traditional GARCH-type models estimated from daily return data. The main benefits are in short-run forecasts up to about 10 days, thereafter we find an EGARCH specification provides a very competitive alternative. The best models are very parsimonuous and feature variance targeting. A robust result is that models in which return innovations are Normally distributed dominate alternatives with t-innovations. This is true for all the models we investigate and true for the entire term structure of average predictive likelihoods.

The out-of-sample results strongly favor parsimony in contrast to the in-sample estimates. The full-sample estimates provide evidence of fat tails and stronger persistence in RV. These are exactly the type of specifications that performed poorly in density forecasts.

This paper is orgnaized as follows. the next section introduces the data used to construct daily returns and daily RV measures. Section 3 discusses the measurement of volatility and the adjustments to realized volatility to remove the effects of market microstructure. The new bivariate models of returns and RV are introduced in Section 4, while the benchmark models based on daily return data are detailed in Section 5. The calculation of density forecasts and the predictive likleihood are discussed in Section 6, and results are presented in Section 7. Section 8 concludes.

[2] For definition and development of bipower variation and realized jumps see, for example, Barndorff-Nielsen and Shephard (2004).

2 Data

We investigate the S&P 500 equity index using the Standard & Poor's Depository Receipt (Spyder) which is a tradeable security (Exchange Traded Fund) that represent 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 transaction price data are obtained from the New York Stock Exchange's Trade and Quotes (TAQ) database. Our sample covers the period January 29, 1993 to March 30, 2004.

After removing errors from the transaction data,[3] a 5-minute grid from 9:30 to 16:00 EST was constructed by finding the closest transaction price before or equal to each grid point time. From this grid, 5-minute intraday continuously compounded (log) returns are constructed. These returns are denoted as $r_{t,i}, i = 1, ..., M$, where M is the number of intraday returns in day t.

3 Realized Volatility Estimation

A traditional proxy for *ex post* latent volatility has been daily squared returns. As shown by Andersen and Bollerslev (1998), this measure of volatility is very noisy and of limited use in assessing features of volatility such as its time-series properties.

Better estimates of *ex post* latent volatility are available. The increment of quadratic variation is a natural measure of *ex post* variance over a time interval. A popular estimator of this type is realized variance or realized volatility (RV) computed as the sum of squared returns over this time interval. As shown by Andersen, Bollerslev, Diebold, and Labys (2001), as the sampling frequency is increased, the sum of squared returns converges to the quadratic variation over a fixed time interval for a broad class of models. Thus RV is a consistent estimate of *ex post* variance for that period. The asymptotic distribution of RV has been studied by Barndorff-Nielsen and Shephard (2002a) who provide conditions under which RV is also an unbiased estimate.

Given the intraday returns, $r_{t,i}, i = 1, ..., M$, the RV estimator of daily *ex post* variance would be

$$RV_{t,u} = \sum_{i=1}^{M} r_{t,i}^2. \tag{3.1}$$

However, in the presence of market-microstructure dynamics, RV can be biased and inconsistent for quadratic variation (Bandi and Russell (2005) and Zhang, Mykland, and Aït-Sahalia (2005)). Therefore, we consider several adjustments to our estimates based on Hansen and Lunde (2006) and gauge their statistical performance in our model comparisons.[4]

[3]Data were 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.

[4]For alternative approaches to dealing with market microstructure dynamics see Ait-Sahalia, Mykland, and Zhang (2005), Bandi and Russell (2006), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006), Oomen (2005), Aït-Sahalia, Mykland, and Zhang (2006) and Zhou (1996).

A first-order correction for bias gives the estimate,

$$RV_{t,AC1} = \sum_{i=1}^{M} r_{t,i}^2 + 2\sum_{i=1}^{M-1} r_{t,i} r_{t,i+1}. \tag{3.2}$$

This estimator suffers from the possibility of a negative value for RV. Hansen and Lunde (2006) suggest the use of Bartlett weights to rule this out. Following this approach, the corrected RV estimator is

$$\begin{aligned} RV_{t,ACqb} &= \omega_0 \hat{\gamma}_0 + 2\sum_{j=1}^{q} \omega_j \hat{\gamma}_j \qquad (3.3) \\ \hat{\gamma}_j &= \sum_{i=1}^{M-j} r_{t,i} r_{t,i+j}, \end{aligned}$$

in which the weights follow a Bartlett scheme $\omega_j = 1 - \frac{j}{q+1}$, $j = 0, 1, ..., q$. We consider $q = 1$ and $q = 2$.

In summary, we construct RV with no adjustment, that is RV_u as in equation (3.1), and, alternatively, RV_{ACqb} using a kernel-based adjustment for market-microstructure dynamics, as in equation system (3.3). We will compare the implications for forecasts of return densities of using adjusted versus unadjusted RV estimates.

In order to match the volatility measures, daily returns are computed as the logarithmic difference of the closing price and the opening price. Our data cover the sample January 29, 1993 to March 30, 2004 with a total of $T = 2813$ daily returns. Our time-series models condition on the first 24 observations.

Table 1 displays summary statistics for returns and different RV estimates. If we take the sample variance of daily returns as a benchmark estimate of volatility in which no market microstructure effects are present, and compare this to the sample mean of RV, we see a clear bias for unadjusted RV. The basic first-order correction, $RV_{t,AC1}$ gives a clear improvement. However, this estimator produces a negative estimate since the minimum is -0.3796. Below this are the Bartlett adjustments which provide a smaller bias adjustment but do not suffer from negative RV estimates. Figure 1 displays daily returns, and Figures 2 and 3 the associated square-root RV estimates for our sample period. Unless otherwise stated, we use $RV_t \equiv RV_{t,ACqb}$ with $q = 1$ in the remainder of the paper.

4 Joint Return-RV Models

As discussed in the Introduction, an integrated model of returns and realized volatility is needed to deal with common questions in finance which require a forecast density of returns for multiple horizons.[5] In this section, we introduce two joint specifications of daily returns and realized volatility. These two bivariate models are distinguished by

[5] Similar to most of the existing literature, although see Meddahi (2003), we ignore aggregation issues linking an underlying diffusion to dynamics of daily returns and RV. We focus on the empirical value of using daily RV in daily return models.

alternative assumptions about RV dynamics. In each case, cross-equation restrictions link the conditional variance of returns and our realized volatility specification. In what follows, we define the information set $\Phi_t = \{r_t, RV_t, r_{t-1}, RV_{t-1}, ..., r_1, RV_1\}$.

4.1 Linking Conditional Variance of Returns to RV

Corollary 1 of Andersen, Bollerslev, Diebold, and Labys (2003) shows that, under empirically realistic conditions, the conditional expectation of quadratic variation (QV_t) is equal to the conditional variance of returns, that is, $E_{t-1}(QV_t) = \text{Var}_{t-1}(r_t) \equiv \sigma_t^2$. If RV is an unbiased estimator of quadratic variation,[6] it follows that the conditional variance of returns can be linked to RV as follows

$$\sigma_t^2 = E_{t-1} RV_t. \tag{4.1}$$

4.2 Component Specifications

We begin with a bivariate specification for daily returns and RV in which conditional returns are driven by Normal innovations and the dynamics of $\log(RV_t)$ are captured by components with different decay rates. In particular, this bivariate system can be summarized as follows:

$$r_t = \mu + \epsilon_t, \ \epsilon_t = \sigma_t u_t, \ \ u_t \sim NID(0,1) \tag{4.2}$$

$$\log(RV_t) = \omega + \sum_{i=1}^{2} \phi_i s_{i,t} + \gamma u_{t-1} + \delta|u_{t-1}| + \eta v_t, \ \ v_t \sim NID(0,1) \tag{4.3}$$

$$s_{i,t} = (1-\alpha_i)\log(RV_{t-1}) + \alpha_i s_{i,t-1}, \ \ 0 < \alpha_i < 1, \ i = 1,2 \tag{4.4}$$

$$\sigma_t^2 = E_{t-1} RV_t = \exp\left(E_{t-1}\log(RV_t) + \frac{1}{2}\text{Var}_{t-1}(\log(RV_t))\right). \tag{4.5}$$

This bivariate specification of daily returns and RV imposes cross-equation restrictions that relate the conditional variance of daily returns to the conditional expectation of daily RV, as in equation (4.1). Equation (4.5) implements that restriction when we assume that RV has a log-Normal distribution. For this specification of our bivariate model, we parameterize the dynamics of daily $\log(RV_t)$ as in equation (4.3). Joint estimation of the bivariate system in equations (4.2) to (4.5) is by maximum likelihood.

Note that this dynamic model for $\log(RV)$ includes the sum of two exponential-smoothing components that are allowed to decay at different rates associated with their smoothing parameters α_i. That is, for component i, a small value of α_i puts more weight on $\log(RV_{t-1})$, and less weight on the past $s_{i,t-1}$. Conversely, an α_i close to 1 puts less weight on recent observations and more weight on past $s_{i,t-1}$ which smooths the data. Startup values of $s_{i,0}$ are set to the average of LogRV based on a reserved presample of data.

Although infinite exponential smoothing provides parsimonious estimates, it possesses several drawbacks. For instance, it does not allow for mean reversion in volatility;

[6] We assume that any stochastic component in the intraperiod conditional mean is negligible compared to the total conditional variance.

and, as Nelson (1990) has shown in the case of squared returns or squared innovations to returns, the model is degenerate in its asymptotic limit. To circumvent these problems, but still retain the parsimony of exponential smoothing, our dynamic model for $\log(RV_t)$, given by equation (4.3), weights each component i by the parameter $0 < \phi_i < 1$ and adds an intercept, ω. Note that when the model is stationary, variance forecasts will mean revert to $(\omega + \delta E(|u_{t-1}|)/(1 - \phi_1 - \phi_2)$. This result can be used to do variance targeting and eliminate the parameter ω from the model.[7] This model implies an infinite expansion in $\log(RV_{t-j})$ with coefficients of $\phi_1(1-\alpha_1)\alpha_1^{j-1} + \phi_2(1-\alpha_2)\alpha_2^{j-1}$, $j = 1, 2, ...$[8]

Finally, since our application is to S&P500 index returns, it is important to allow for asymmetric effects in volatility. To facilitate comparisons with a benchmark EGARCH model, our parameterization in equation (4.3) includes two asymmetry terms, γu_{t-1} and $\delta|u_{t-1}|$, both of which are functions of the standardized return innovation, u_{t-1}. The impact coefficient for a negative innovation to returns will be $-\gamma + \delta$, whereas the impact of a postive innovation will be $\gamma + \delta$. Typically, $\hat{\gamma} < 0$, which means that a negative innovation to returns implies a higher conditional variance for next period. Note that, unlike EGARCH, our parameterization in equation (4.3) does not propagate the asymmetry further into future volatility.

In-sample fit of GARCH models have generally favoured return innovations with tails that are fatter than those implied by a Normal distribution. In this paper, we evaluate whether or not that result obtains for our bivariate model of returns and RV. That is, we replace equation (4.2) with

$$r_t = \mu + \epsilon_t, \ \epsilon_t = \sigma_t u_t, \ \ u_t \sim t_\nu(0, 1), \tag{4.6}$$

in which t_ν denotes a t-distribution with mean 0, variance 1, and ν degrees of freedom. The rest of the bivariate dynamic system for this case is the same as in equations (4.3) to (4.5) above. We compare this bivariate system with t-distributed return innovations to that with Normally distributed innovations, not only for in-sample fit, but also for out-of-sample forecasts of the return distribution.

4.3 HAR Specifications

Corsi (2003) and Andersen, Bollerslev, and Diebold (2005) use Heterogeneous AutoRegressive (HAR) functions of RV in order to parsimoniously capture long-memory dependence. Motivated by that work, we define

$$\log(RV_{t-h,h}) = \frac{1}{h}\sum_{i=0}^{h-1} \log(RV_{t-h+i}), \quad \log(RV_{t-1,1}) = \log(RV_{t-1}). \tag{4.7}$$

For example, $\log(RV_{t-22,22})$ averages $\log(RV)$ over the most recent 22 days, $\log(RV_{t-5,5})$ over the most recent 5 days, etc.

This leads to our bivariate specification for daily returns and RV with the dynamics of $\log(RV_t)$ modeled as an asymmetric HAR function of past $\log(RV)$. This bivariate

[7] That is, set $\omega = \text{mean}(\log(RV))(1 - \phi_1 - \phi_2)$ for $\delta = 0$. In our application we found $\delta E(|u_{t-1}|)$ to be very small and omit it in the empirical work.

[8] Expanding (4.4) gives $s_{i,t} = (1 - \alpha_i) \sum_{n=0}^{\infty} \alpha_i^n \log(RV_{t-1-n})$.

system is summarized as follows:

$$\begin{aligned} r_t &= \mu + \epsilon_t, \ \epsilon_t = \sigma_t u_t, \ u_t \sim NID(0,1) & (4.8) \\ \log(RV_t) &= \omega + \phi_1 \log(RV_{t-1}) + \phi_2 \log(RV_{t-5,5}) + \phi_3 \log(RV_{t-22,22}) \\ &+ \gamma u_{t-1} + \delta|u_{t-1}| + \eta v_t, \ v_t \sim NID(0,1), & (4.9) \\ \sigma_t^2 = E_{t-1} RV_t &= \exp\left(E_{t-1} \log(RV_t) + \frac{1}{2} \mathrm{Var}_{t-1}(\log(RV_t) \right). & (4.10) \end{aligned}$$

Therefore, the difference between this bivariate dynamic system and that in section 4.2 above, is that, the component model of $\log(RV)$ in equation (4.3) is replaced by the HAR function in equation (4.9). If the model is stationary, the unconditional mean of $\log(RV)$ is $(\omega + \delta E(|u_{t-1}|)/(1 - \phi_1 - \phi_2 - \phi_3)$. This can be used to enforce variance targeting as discussed in the previous section.

Again, in order to evaluate the potential importance of t-distributed return innovations for this alternative bivariate specification, we replace equation (4.8) with

$$r_t = \mu + \epsilon_t, \ \epsilon_t = \sigma_t u_t, \ \ u_t \sim t_\nu(0,1), \tag{4.11}$$

The remainder of the bivariate dynamic system for this case is the same as in equations (4.9) and (4.10).

5 Benchmark Specification

One way to ascertain whether or not high-frequency (intraperiod) information contributes to improved forecasts of return distributions, is to compare density forecasts from the bivariate specifications of returns and $\log(RV)$ summarized above with those from a traditional EGARCH model (Nelson (1991)) specified as

$$\begin{aligned} r_t &= \mu + \epsilon_t, \ \ \epsilon_t = \sigma_t u_t \ \ \ u_t \sim NID(0,1), & (5.1) \\ \log(\sigma_t^2) &= \omega + \beta \log(\sigma_{t-1}^2) + \gamma u_{t-1} + \alpha|u_{t-1}|. & (5.2) \end{aligned}$$

Note that this EGARCH volatility model conditions on daily return innovations rather than RV which incorporates intraday information. We also investigate a GARCH and GJR-GARCH model both with Normal innovations and with t-innovations .

6 Density Forecasts

Our focus is on the return distribution and we evaluate each of the models ability to provide accurate density forecasts of daily returns. A popular approach to assess the accuracy of a model's density forecasts is the predictive likelihood or logarithmic score (Amisano and Giacomini (2006), Bao, Lee, and Saltoglu (2006), and Weigend and Shi (2000)). This approach evaluates the model's density forecast at the realized return. This is generally done for a one-step-ahead forecast density as multiperiod density forecasts are often not available in closed form. In this paper we advocate multiperiod forecasts since they provide more information to discern among models. The details of the muliperiod predictive likelihood and how to calculate it are discribed below.

Define the average predictive likelihood over the out-of-sample observations $t = \tau + k_{max} - 1, ..., T - k$, as

$$D_{M,k} = \frac{1}{T - k - \tau - k_{max} + 2} \sum_{t=\tau+k_{max}-1}^{T-k} \log f_{M,k}(r_{t+k}|\Phi_t, \theta), \quad k \geq 1, \tag{6.1}$$

where $f_{M,k}(x|\Phi_t, \theta)$ is the k-period ahead predictive density for model M, given Φ_t and parameter θ, evaluated at the realized return $x = r_{t+k}$. Intuitively, models that better account for the data produce larger $D_{M,k}$.

As we will see below for our application, $T = 2813, \tau = 1201$, $k_{max} = 60$ so that $\tau + k_{max} - 1 = 1260$. $D_{M,k}$ is computed for each k using the out-of-sample returns $r_{1260}, ..., r_{2813}$. That is, if $k = 1$, $D_{M,1}$ is computed using out-of-sample returns $r_{1260}, ..., r_{2813}$. For $k = 2$, $D_{M,2}$ is computed using the same out-of-sample returns, etc. This gives us a *term structure of average predictive likelihoods*, $D_{M,1}, ..., D_{M,60}$, to compare the performance of alternative models, M, over an *identical set of out-of-sample data points.*

6.1 Computations

For all $k > 1$ the term $f_{M,k}(r_{t+k}|\Phi_t, \theta)$ will be unknown for the models we consider. However, given that we have fully specified the law of motion for daily returns and RV, we can accurately estimate this quantity by standard Monte Carlo methods. A conventional approach to estimate the forecast density would be to simulate the model out k periods a large number of times and apply a kernel density estimator to these realizations. However, using the kernel density estimator to estimate the forecast density ignores the fact that, in our applications, conditional on the variance we know the distribution. The use of conditional analytic results has been referred to as Rao-Blackwellization and is a standard approach to reduce the variance of a Monte Carlo estimate (Robert and Casella (1999)). This is particularly useful in density estimation which is our context.

To illustrate consider our basic benchmark EGARCH model in (5.1). Note that in this univariate case the information set, Φ_t, just includes past returns. Our estimate is

$$\begin{aligned} f_{M,k}(r_{t+k}|\Phi_t, \theta) &= \int f(r_{t+k}|\mu, \sigma^2_{t+k}) p(\sigma^2_{t+k}|\Phi_t) d\sigma^2_{t+k} && (6.2) \\ &\approx \frac{1}{N} \sum_{i=1}^{N} f(r_{t+k}|\mu, \sigma^{2(i)}_{t+k}), \quad \sigma^{2(i)}_{t+k} \sim p(\sigma^2_{t+k}|\Phi_t) && (6.3) \end{aligned}$$

where $f(r_{t+k}|\mu, \sigma^{2(i)}_{t+k})$ is a Normal density with mean μ and variance σ^2_{t+k}, evaluated at return r_{t+k}; and $\sigma^{2(i)}_{t+k}$ is simulated out N times according to the EGARCH specification, $p(\sigma^2_{t+k}|\Phi_t)$, which is conditional on time t quantities σ^2_t, u_t, and $\hat{\theta}$, the maximum likelihood estimate of the parameter vector based on Φ_t.

For the joint models of returns and RV, we do a similar exercise to compute the predictive likelihood for returns. In this case, we simulate out both the return and RV dynamics, which implicitly integrates out the unknown σ^2_{t+k}. For each simulation of $RV^{(i)}_{t+1}, ..., RV^{(i)}_{t+k-1}$, $i = 1, ..., N$ we can compute $\sigma^{2(i)}_{t+k} = E_{t+k-1} RV^{(i)}_{t+k}$ using (4.1). The standard Monte Carlo numerical standard error of the standard deviation of $f_t(r_{t+k}|\mu, \sigma^{2(i)}_{t+k})$ divided by $\sqrt{N}$ can be used to assess accuracy.[9] In our application we found $N = 10000$ to provide sufficient accuracy. For example, the numerical standard error is typically well below 1% of $\hat{D}_{M,k}$.

Figure 4 displays an example of the predictive density for one of our models at horizons $k = 1, k = 20$, and $k = 60$. The difference in the densities is due to the mixing that occurs since the future conditional variance is a random variable. Note that even though this model has Normal innovations to returns, the forecast densities for $k > 1$ will be a mixture of Normals and as a result will produce thick tails. The density is computed analogously to (6.3) for $f_{M,k}(x|\Phi_t, \theta)$ and traced out for values of x from $[-6, 6]$ for a component model. The predictive likelihood is obtained by evaluating these densities at the realized return.

7 Results

The next two subsections report, respectively, out-of-sample density forecasts evaluated using predictive likelihoods, and full sample estimates of selected models.

7.1 Density Forecasts

The first density forecasts begin at 1997/10/28 and condition on data from 1993/1/29 – 1997/10/27 (1200 observations). The predictive likelihoods are computed using returns from 1997/10/28 – 2004/3/30. For each given information set, Φ_t, we compute k-period ahead predictive likelihood values associated with realized r_{t+k} for $k = 1, ..., 60$. After this, the information set is updated to Φ_{t+1} and we calculate predictive likelihoods for r_{t+1+k}. This is repeated for all of the out-of-sample data $t = \tau, ..., T-k$, where $\tau = 1201$. Models are re-estimated every 50 observations.

To calculate the average predictive likelihoods, defined by equation (6.1), we drop the first 59 predictive likelihood values so that the remaining out-of-sample returns enter into average predictive likelihood values for every k. In other words, the average predictive likelihoods are computed using the out-of-sample returns $r_{1260}, ..., r_{2813}$ for each and every k. This ensures that we have used an identical set of return observations over all forecast horizons k, so that the average predictive likelihoods are not only comparable across models but also over different forecast horizons for a particular model.

[9]To calculate a numerical standard error for $\hat{D}_{M,k}$: let v^2 denote the sample variance of the draws of $f_t(r_{t+k}|\mu, \sigma^{2(i)}_{t+k})$, then the numerical standard error for $\hat{f}_{M,k}(r_{t+k}|\Phi_t, \theta)$ is $\nu/\sqrt{N}$. Using the delta rule to calculate $\hat{Var}(\log \hat{f}_{M,k}(r_{t+k}|\Phi_t, \theta))$; the numerical standard error of $\hat{D}_{M,k}$ is $\sqrt{\sum_{t=\tau+k_{max}-1}^{T-k} \hat{Var}(\log \hat{f}_{M,k}(r_{t+k}|\Phi_t, \theta))/(T-k-\tau-k_{max}+2)}$.

7.1.1 Effect of the forecast horizon

All of the average predictive likelihood term structures display a negative slope. This is because the conditioning information is most useful for small k. As we forecast further and further out-of-sample, the value of the current information diminishes. All of our models are stationary so that multiperiod forecast densities converge to the unconditional distribution. Using the same data points to evaluate the predictive likelihood for different k, we can see how accuracy of forecasts deteriorate for longer horizons.

7.1.2 Effect of adjustments to RV

Figures 5 and 6 display the term structure of average predictive likelihoods for our main bivariate specifications of return and RV using alternative estimates of RV. Those figures illustrate the term structure of average predictive likelihoods using the unadjusted estimate of RV, from equation (3.1) which uses the raw 5-minute returns, versus RV which has been adjusted for first-order autocorrelation using Bartlett weights, as in equation (3.3). Recall that models which explain the data better have larger predictive likelihood values. Both figures show a difference in the accuracy of density forecasts, particularly at short horizons. We also explored higher-order adjustments but found only minor differences with the first-order adjustment. Based on these results we use the first-order-adjusted RV, $RV_{t,AC1b}$ from equation (3.3), for the remainder of the paper.

7.1.3 Conditional variance specification

Bollerslev, Kretschmer, Pigorsch, and Tauchen (2005) model returns, bipower variation and realized jumps in a multivariate context. They assume that RV_t is synonymous with the conditional variance of returns. In our models, the conditional variance is derived from time $t-1$ information based on an assumption of log Normality. Athough RV_t is an ex post estimate of volatility and contains measurement errors, it is of interest to know if this information difference is important. Therefore, we consider the following model motivated by Bollerslev, Kretschmer, Pigorsch, and Tauchen (2005) in the context of our specfications,

$$\begin{aligned} r_t &= \mu + \epsilon_t, \;\; \epsilon_t = \sqrt{RV_t} u_t, \;\; u_t \sim D(0,1) & (7.1)\\ \log(RV_t) &= \omega + \phi_1 s_{1,t} + \gamma u_{t-1} + \delta|u_{t-1}| + \eta v_t, \; v_t \sim NID(0,1), & (7.2)\\ s_{1,t} &= (1-\alpha_1)\log(RV_{t-1}) + \alpha_1 s_{1,t-1}, & (7.3) \end{aligned}$$

where ω is set by variance targeting. This is the one-component-RV model with variance targeting but using RV_t as the conditional variance. Figure 7 shows the predictive likelihood for our model, as well as the contemporaneous version that uses RV as the variance. These results suggest no loss in information and favor our approach.

7.1.4 Alternative forecasting functions for RV

Figures 8 and 9 display results for our joint return-RV models. Included are variance-targeting versions and one and two-component specifications. Both figures show variance targeting to provide superior forecast accuracy, and the gains are largest for long-term forecasts. In Figure 8, the more parsimonious one-component versions improve upon the two-component version.

7.1.5 Effect of intraday information

To investigate the statistical value of using RV versus using squared daily returns, we consider density forecasts for a one-component model in RV and an analogous version based on only daily return information. The one-component model follows from Section 4.2, while the model based on daily returns is

$$
\begin{aligned}
r_t &= \mu + \epsilon_t, \;\; \epsilon_t = \sigma_t u_t, \;\; u_t \sim NID(0,1) && (7.4)\\
\sigma_t^2 &= \exp(\omega + \phi_1 s_{1,t} + \gamma u_{t-1} + \delta |u_{t-1}|), && (7.5)\\
s_{1,t} &= (1-\alpha_1)\log(\epsilon_{t-1}^2) + \alpha_1 s_{1,t-1}. && (7.6)
\end{aligned}
$$

The parameters $\mu, \omega, \phi_1, \gamma$, and δ are common to both specifications. The main difference is that we have replaced RV_{t-1} with ϵ_{t-1}^2. The predictive likelihood is displayed for these specifications in Figure 10. The model which uses RV, and treats it as stochastic, provides clear gains in accuracy for about the first 35 periods out; thereafter the model based on daily return information is marginally better. This result suggests that the superior signal that RV provides does improve density forecasts; the largest gains are from short-term forecasts, approximately two months out-of-sample.

7.1.6 Comparing models

To futher investigate whether the use of RV improves upon traditional volatility models in characterizing and forecasting the return density, we tried several popular volatility models estimated with daily returns only. These included GARCH, GJR-GARCH, and EGARCH with both Normal and t-innovations to returns. The EGARCH model, summarized in Section 5, provided the best performance among all traditional models for these equity index returns. Figure 11 compares that EGARCH model with Normal innovations to our alternative joint return-RV models. RV results in better forecast density accuracy from 1 to 10 days out. After this all models are similar until the component model dominates again for $k \geq 45$. Both RV models provide robust gains for short-term forecasts, however, the improvements in longer-term forecasts may be sensitive to the model specification of RV.

A natural question is whether or not the gains in accuracy are statistically significant. For example, consider the statistic $\widehat{DIFF}_k = \hat{D}_{1-component,k} - \hat{D}_{EGARCH,k}$, which is the difference in the average predictive likelihood from the one-component model versus the EGARCH model for a particular k. We follow the approach of Kim, Shephard, and Chib (1998) who use simulation methods to calculate the finite sample distribution of a test statistic for non-nested models. Under the null hypothesis of the component model, we use the full-sample estimate $\hat{\theta}$ of the model parameters to generate 300 samples of data identical in length to our S&P500 data. For each sample we compute the average predictive likelihood, based on observations 1260 − 2813, for both models and hence calculate $DIFF_k$ for selected k. The number of repetitions (samples) is 300 and, due to the computational costs, we reduce the simulations used to calculate the predictive likelihoods to 2000. This gives us an empirical distribution of $DIFF_k$ under the null hypothesis that the true model is the component specification.[10]

Column 3 of Table 2 reports the probability under the null of being less than our observed test statistic for selected k. Intuitively, probabilities near 0 would be evidence against the null hypothesis. Generally, we find that for $k = 5$ ($k = 3$ to $k = 7$) our results are consistent with the component model being superior, as are the cases of $k \geq 35$. However, there are some k which are not consistent with the component model. In these situations, relative to our simulated distribution, we would have expected a larger $\widehat{DIFF}_k$ from our data.

7.1.7 Alternative innovations to returns

For all of the models investigated in this paper, versions with Normal innovations improved upon versions with t-innovations. This was true for the joint return-RV models as well as all the GARCH models not reported. For example, the term structures of average predictive likelihoods for our favored component model, using Normal versus t-distributed return innovations, are illustrated in Figure 12. The model with Normal innovations is more accurate for the whole forecast horizon.

Following the simulation approach used in the previous subsection we consider the significance of these results in Table 3. In this case, define $\widehat{DIFF}_k = \hat{D}_{t-innov,k} - \hat{D}_{N-innov,k}$, as the difference in the average predictive likelihood for the one-component-RV model with variance targeting and t-innovations versus the identical model with Normal innovations. Column 2 of this table displays the counterpart to Figure 12 for selected k. $\widehat{DIFF}_k$ is always negative meaning the model with Normal innovations is preferred. Under the null hypothesis of the t-innovation model, simulation methods[11] are used to produce an empirical distribution of $DIFF_k$ and hence $P(DIFF_k < \widehat{DIFF}_k)$, based on 300 repetitions. The last column shows the results and indicates that $\widehat{DIFF}_k$ computed from the S&P500 data are completely inconsistent with the t-innovations version of the model, leading to a strong rejection of the model with t-innovations for out-of-sample density forecasts.

Figure 13 shows the origin of the differences in the two specifications. This figure displays the individual forecast densities for $k = 1$, $k = 20$ and $k = 60$ given information up to 1999/6/1. The vertical line denotes the actual realized return 1, 20 and 60 periods ahead. Where this line intersects the forecast density gives the value of the predictive likelihood for that observation. Note that the average predictive likelihood reported for a particular k in Figure 12 is computed as in equation (6.1) from 1554 of these k-period-ahead, out-of-sample, return realizations.

There are pronounced differences in the models at each horizon. The t-innovation versions appear to be too thick tailed or too peaked around the mean. Recall that, as shown in Figure 4, the model with Normal innovations to returns will have forecast densities for $k > 1$ that are a mixture of Normals and as a result will already have thick tails.

[10] The reverse setup with EGARCH as the null model cannot be used since simulating EGARCH does not provide RV data which our joint model uses.

[11] Based on full sample MLE results for the assumed DGP.

7.1.8 Out-of-sample summary

In summary, we have shown that in terms of forecast density accuracy out-of-sample, models with RV can provide improvements over traditional models. We find parsimonious specifications dominate and variance targeting produces the best results. A robust result is that fat-tailed t-innovations in the return equation produce inferior results to Normally distributed innovations. Although no model is the best for every forecast horizon a one component model with variance targeting and Normal innovations to returns porforms well. There is no strong evidence against this model, neverthless, an EGARCH model based on daily return data remains a strong competitor. In the next section we contrast this out-of-sample performance with full sample model estimates.

7.2 Full-Sample Estimates

Tables 4 and 5 presents full-sample estimates for both joint return-RV models. We include the best model based on the density forecasts as well as a richer specification. For example, we include two components and t-innovations for the component model. In each case the richer specification is strongly favored. We also find strong evidence of fat-tailed innovations. For instance in the case of the HAR model the degree of freedom parameter is accurately estimated and approximately 6.56. In the case of the restricted HAR model there is some evidence it is not fully capturing the volatility dynamics as measured by the Ljung-Box statistic on the squared return innovations. By any measure we would reject the parsimonuous versions of these models. This is in direct contrast to our out-of-sample forecast results in Section 7.1 which strongly favor the restricted versions of our models.

8 Conclusion

This paper proposes a joint specification of daily returns and RV which links the conditional moments of RV to the conditional variance of returns and exploits the benefits of using intraperiod information to obtain accurate measures of volatility. Our focus is on out-of-sample forecasts of the return distribution generated by our bivariate models of return and RV. We explore which features of the time-series models contribute to superior density forecasts over horizons of 1 to 60 days out-of-sample.

Our main method of model comparison uses the predictive likelihood of returns, the forecast density evaluated at the realized return, which provides a measure of the likelihood of the data being consistent with the model. An identical set of return observations is used to compute a term structure of test statistics over a range of forecast horizons, so that the average predictive likelihoods are not only comparable across models but also over different forecast horizons for a particular model.

Two alternative joint specifications of daily returns and realized volatility were investigated. These two bivariate models are distinguished by alternative assumptions about RV dynamics. The first model allows different components of $\log(RV)$ to have different decay rates. A second model uses a heterogenous autoregressive (HAR) specification of $\log(RV)$. Both of these bivariate models allow for asymmetric effects of past negative versus positive return innovations. Both models are stationary and consistent with mean reversion in RV. We dicuss how to implement variance targeting for these specifications.

Using the predictive likelihood, we find that market microstructure adjustments to RV are important, particularly for short-term density forecasts. In general, we find that our models with RV do provide improvements over traditional GARCH-type models estimated from daily return data. The main benefits are in short-run forecasts up to about 10 days, thereafter we find an EGARCH specification provides a very competitive alternative. The best models are very parsimonuous and feature variance targeting. A particularly robust result is that models in which return innovations are Normally distributed dominate alternatives with t-innovations. This is true for all the models we investigate and true for the entire term structure of predictive likelihoods.

The out-of-sample results strongly favor parsimony in contrast to the in-sample estimates. The full-sample estimates provide evidence of fat tails and stronger persistence in RV. These are exactly the type of specifications that performed poorly in density forecasts. A bivariate model of returns with Normal innovations and a mean-reverting, exponentially decaying function of $\log(RV)$ provides the best density forecasts over a range of out-of-sample horizons.

Table 1: Summary Statistics: Returns and Realized Volatility

	Mean	Variance	Min	Max
r_t	-0.0115	0.9675	-7.5035	8.2359
RV_u	1.2445	2.7500	0.0366	33.2162
RV_{AC1}	0.9749	2.4003	-0.3796	28.3522
RV_{AC1b}	1.1097	2.4755	0.0356	30.7842
RV_{AC2b}	1.1431	2.4292	0.0366	29.2199
RV_{AC3b}	1.1239	2.3804	0.0366	29.7705

r_t are daily returns constructed from open and close prices. RV_u is constructed from raw 5-minute returns with no adjustment.

$$\begin{aligned} RV_{AC1} &= \sum_{i=1}^{M} r_{t,i}^2 + 2\sum_{i=1}^{M-1} r_{t,i}r_{t,i+1} \\ RV_{ACqb} &= \hat{\gamma}_0 + 2\sum_{j=1}^{q}\omega_j\hat{\gamma}_j, \quad \hat{\gamma}_j = \sum_{i=1}^{M-j} r_{t,i}r_{t,i+j}, \quad \omega_j = 1 - \frac{j}{q+1},\ j = 0,1,...,q, \\ q &= 1,2,3 \end{aligned}$$

Figure 1: Time Series of Daily Returns

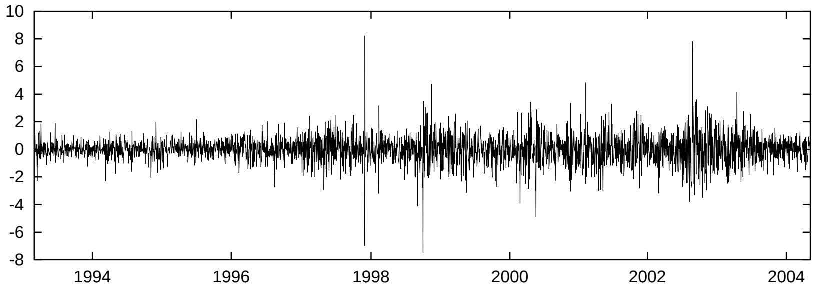


Figure 2: Time Series of $\sqrt{RV}$

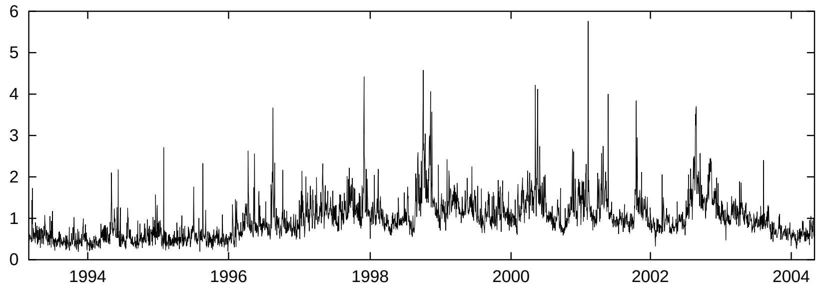


Figure 3: Time Series of $\sqrt{RV_{AC1b}}$

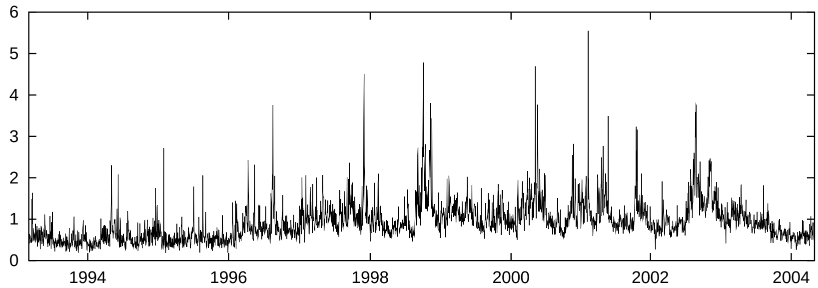
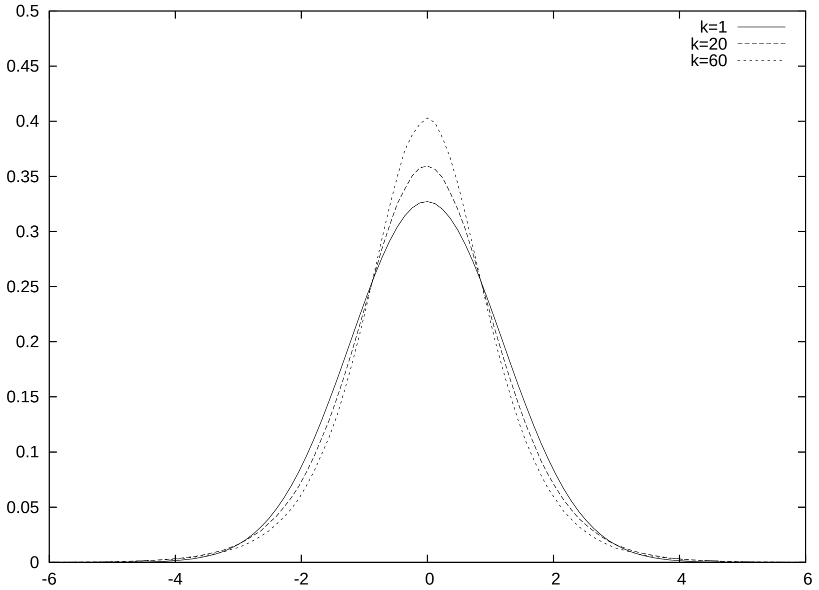


Figure 4: Forecast Densities for Different Forecast Horizons



Forecast densities k-periods ahead for the one-component model with variance targeting and Normal innovations to returns. The forecast densities are based on information up to and including observation 1999/6/1.

Figure 5: Results using different RV estimates: Component Model

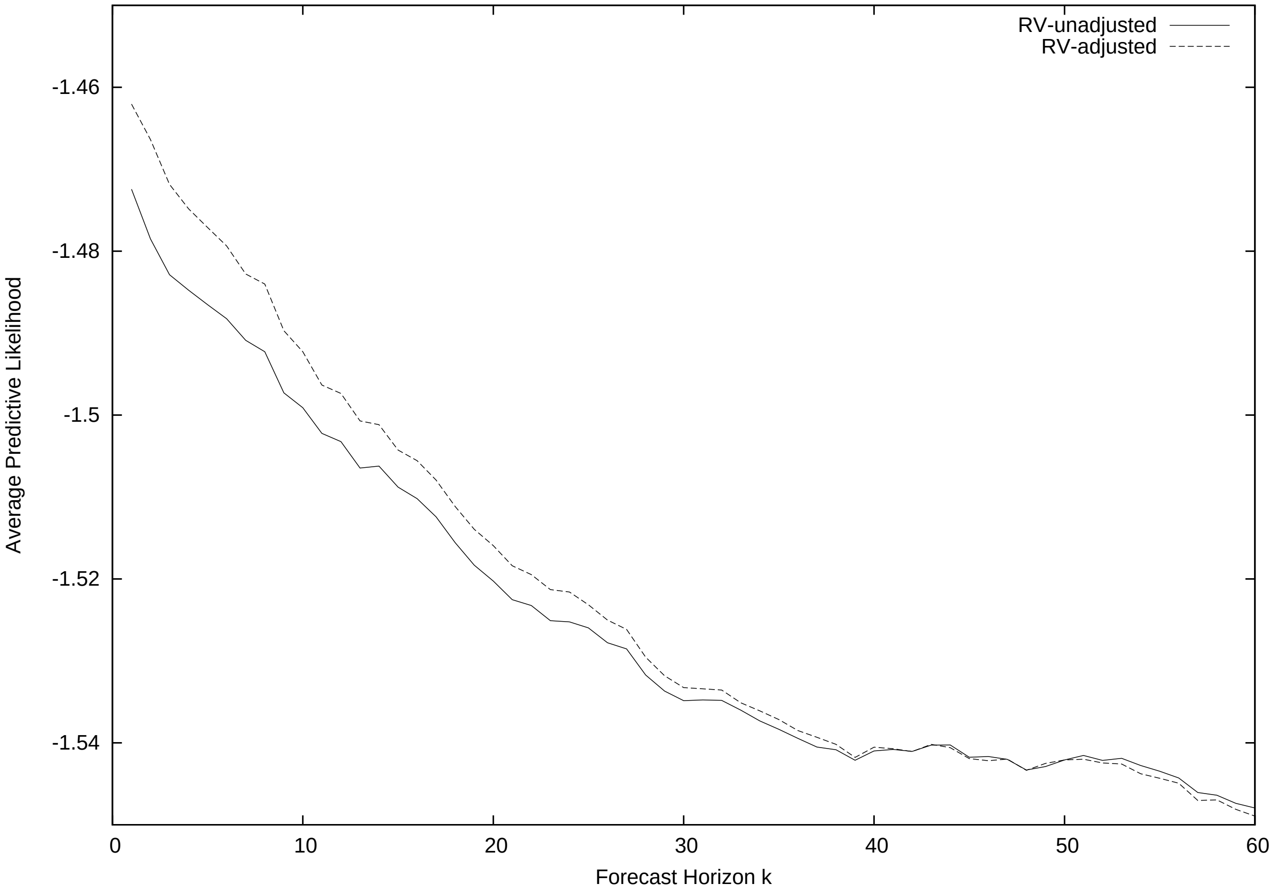


Figure 6: Results using different RV estimates: HAR Model

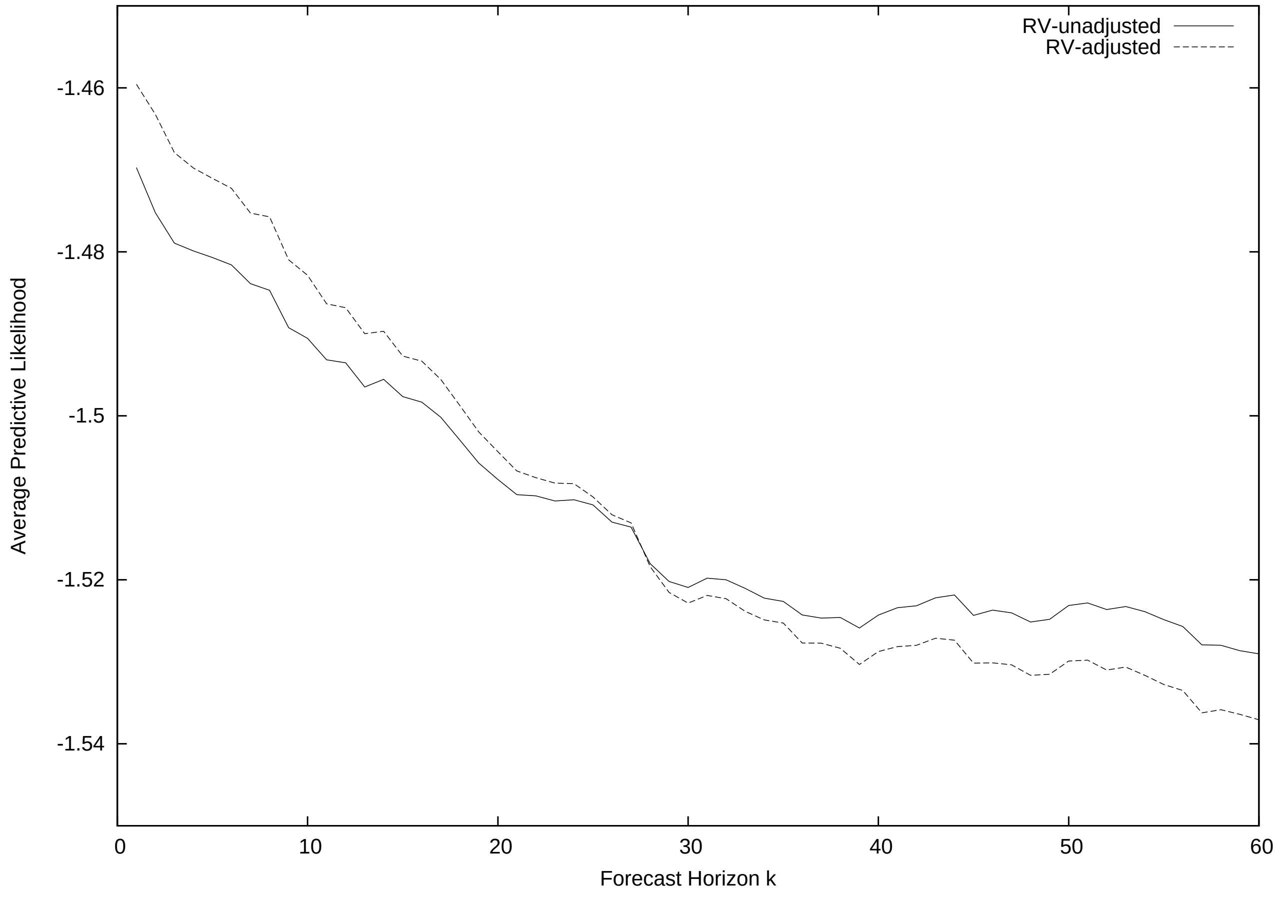


Figure 7: Standard Model versus Contemporaneous RV

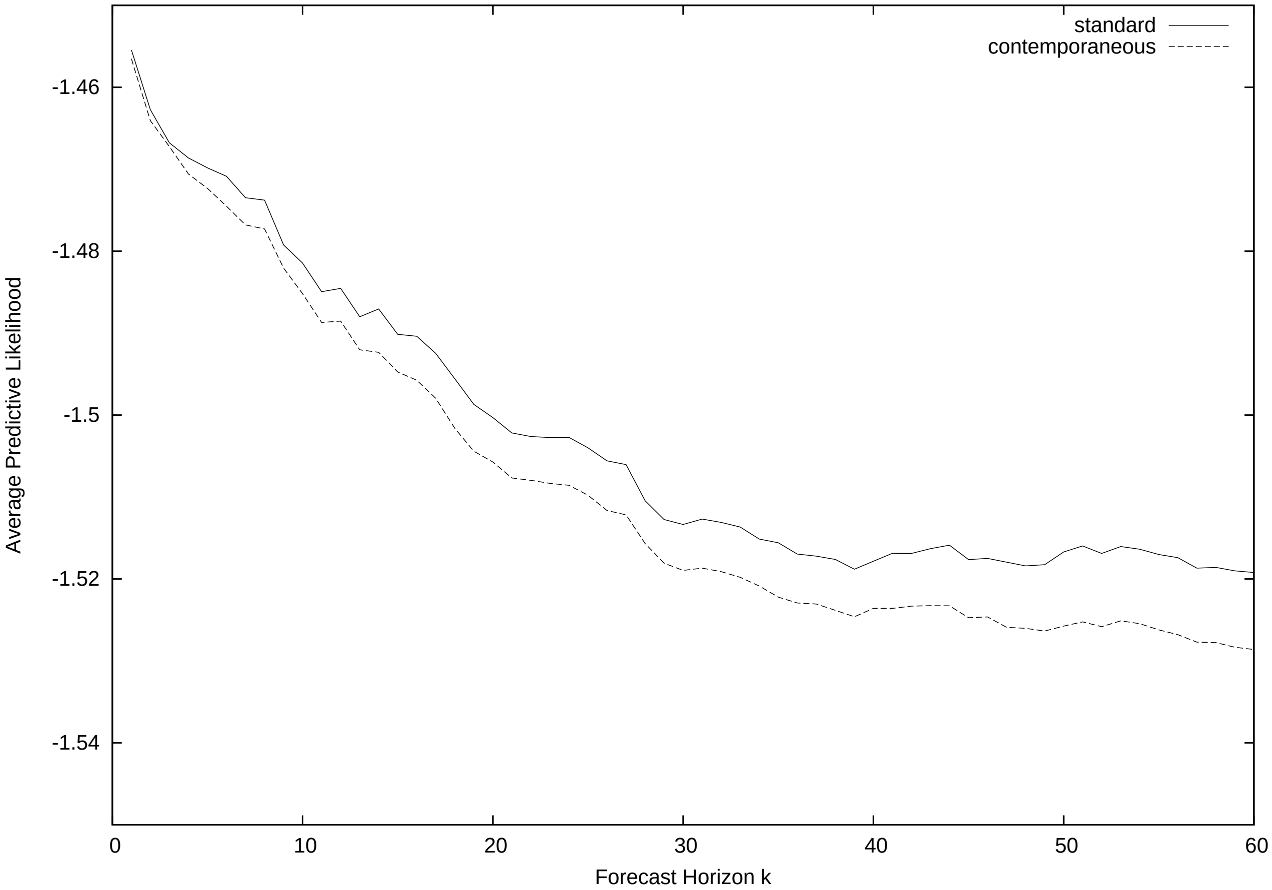


Figure 8: Component Models

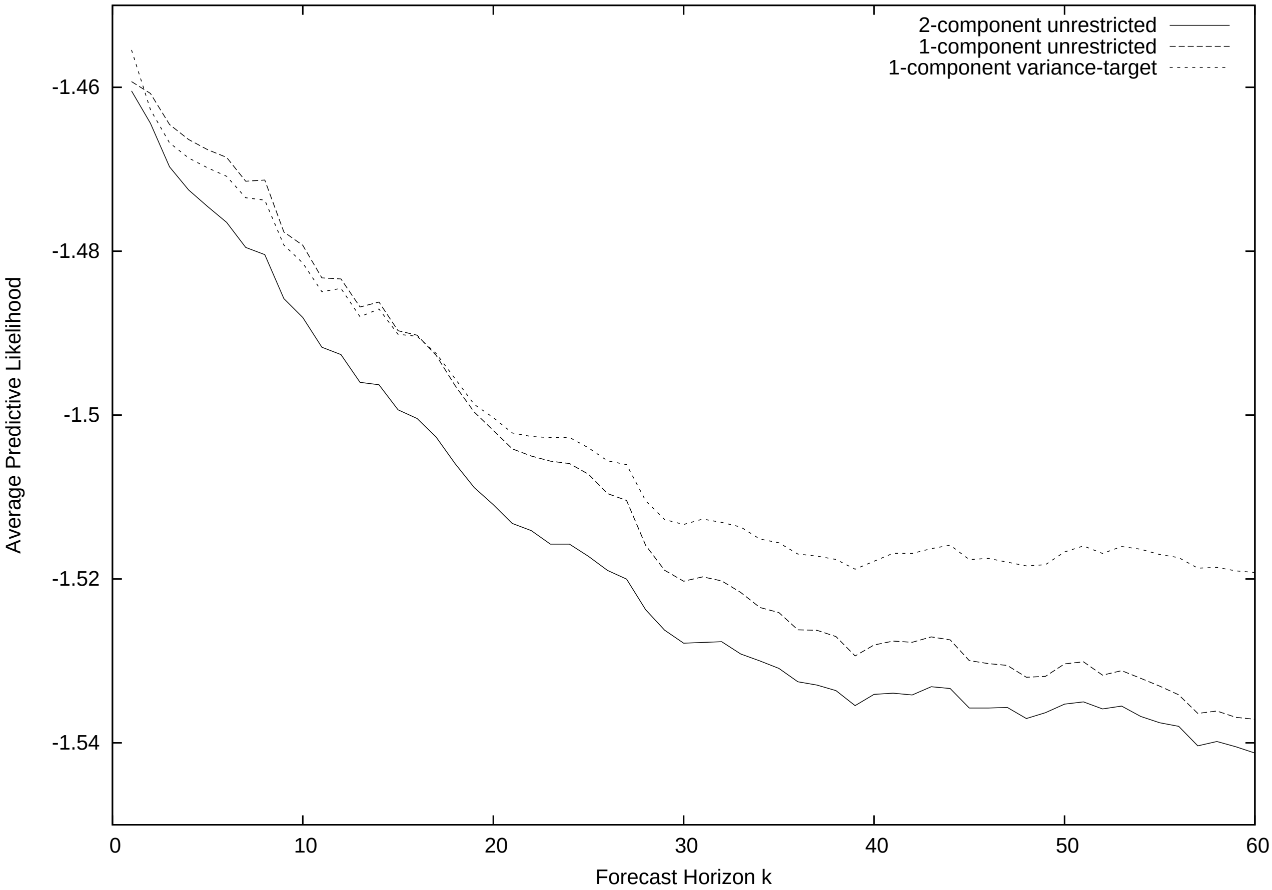


Figure 9: HAR Models

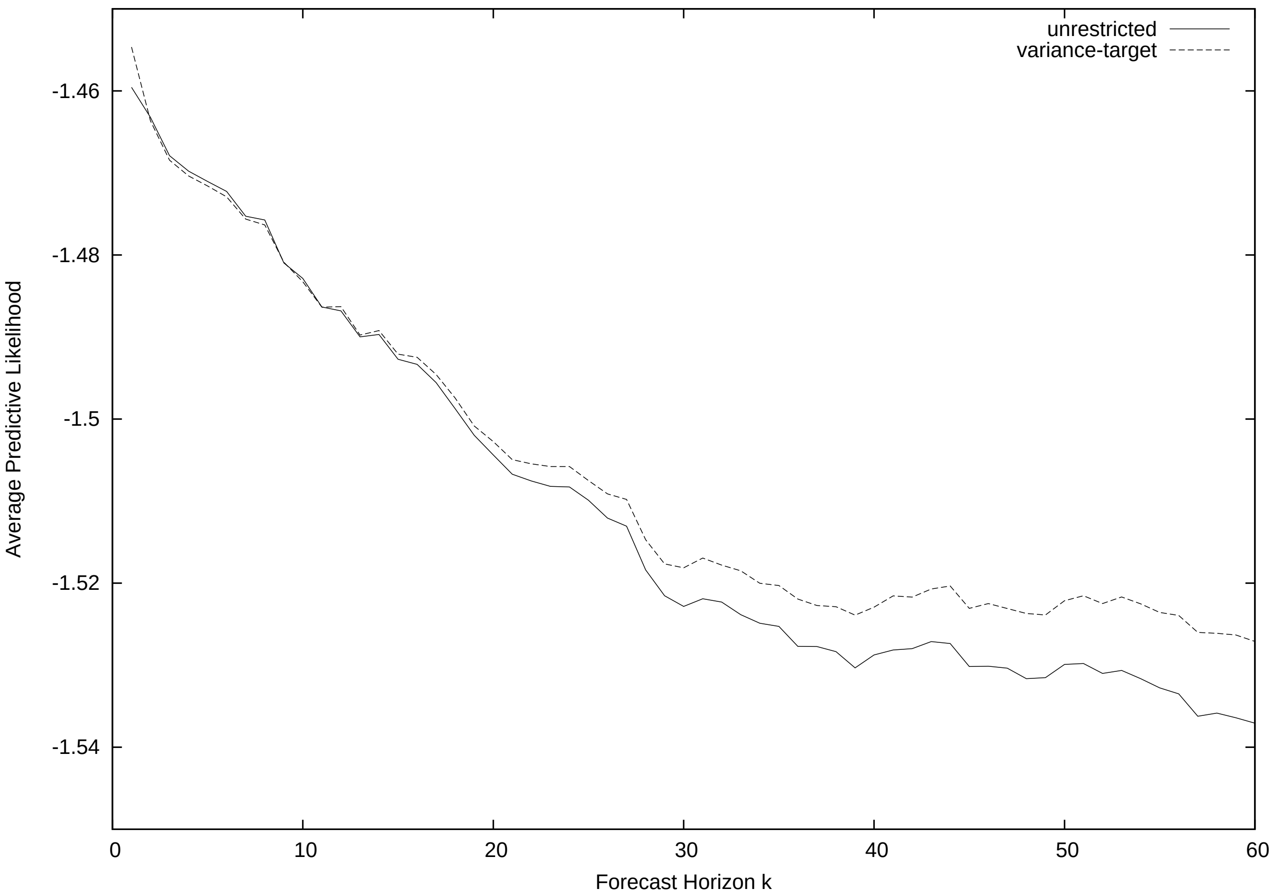
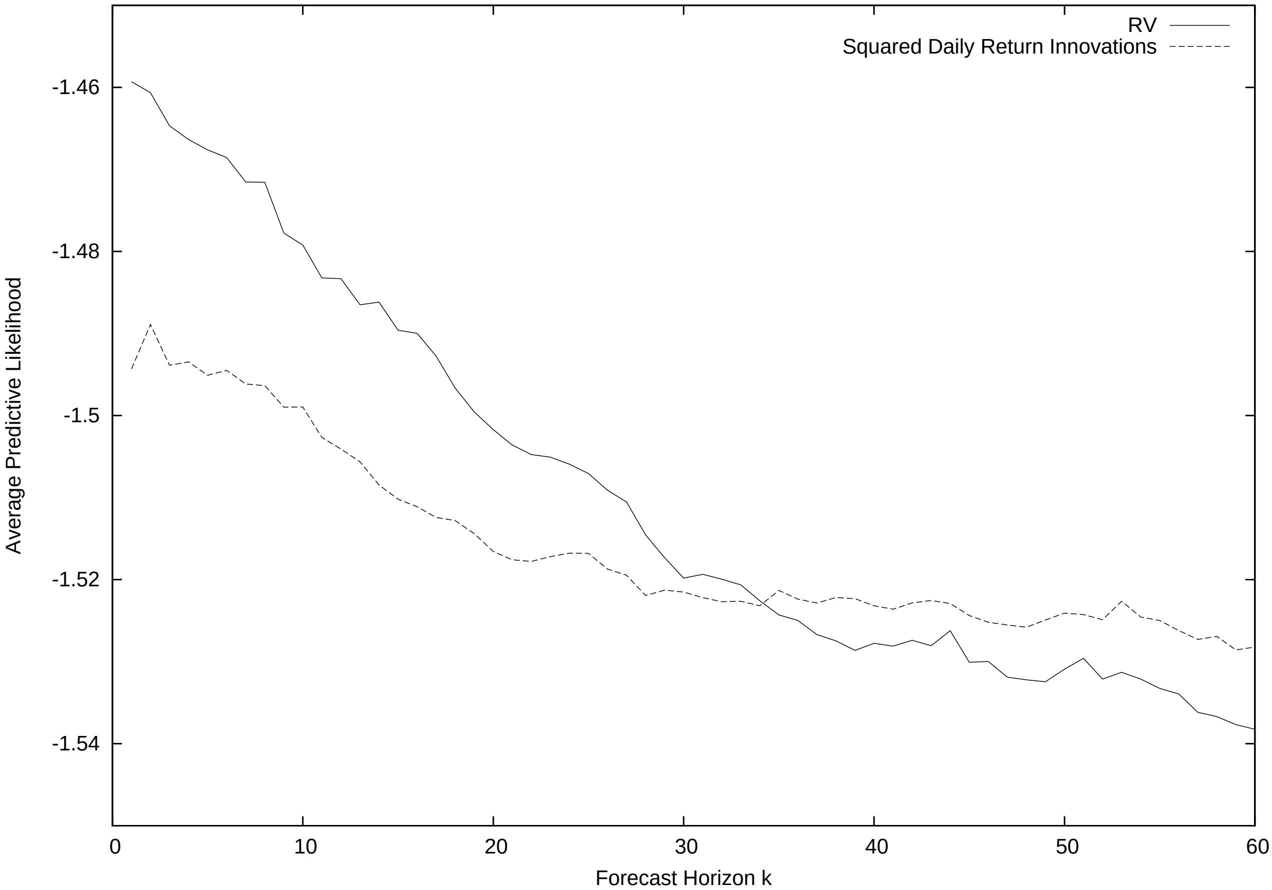


Figure 10: Component Model using RV versus Squared Daily Return Innovations



$$r_t = \mu + \epsilon_t, \;\; \epsilon_t = \sigma_t u_t, \;\; u_t \sim NID(0,1)$$

Using RV:

$$\sigma_t^2 = \exp\left(E_{t-1}\log(RV_t) + \frac{1}{2}Var_{t-1}(\log(RV_t))\right)$$

$$\log(RV_t) = \omega + \phi_1 s_{1,t} + \gamma u_{t-1} + \delta|u_{t-1}| + \eta v_t, \; v_t \sim NID(0,1),$$

$$s_{1,t} = (1-\alpha_1)\log(RV_{t-1}) + \alpha_1 s_{1,t-1}.$$

Using Daily Return Innovations:

$$\sigma_t^2 = \exp(\omega + \phi_1 s_{1,t} + \gamma u_{t-1} + \delta|u_{t-1}|),$$

$$s_{1,t} = (1-\alpha_1)\log(\epsilon_{t-1}^2) + \alpha_1 s_{1,t-1}.$$

Figure 11: Best Models

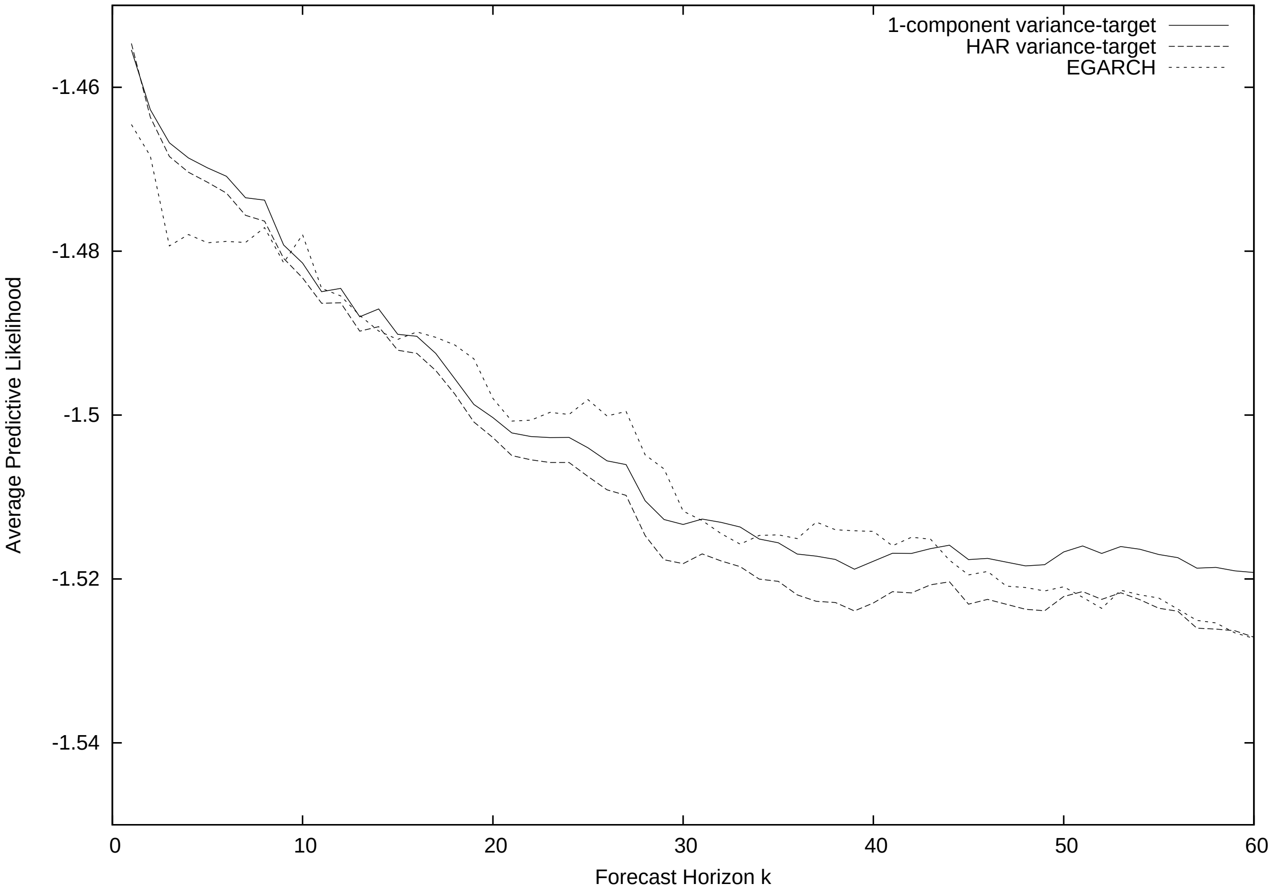


Figure 12: Normal versus t-innovations to Returns, Component Model

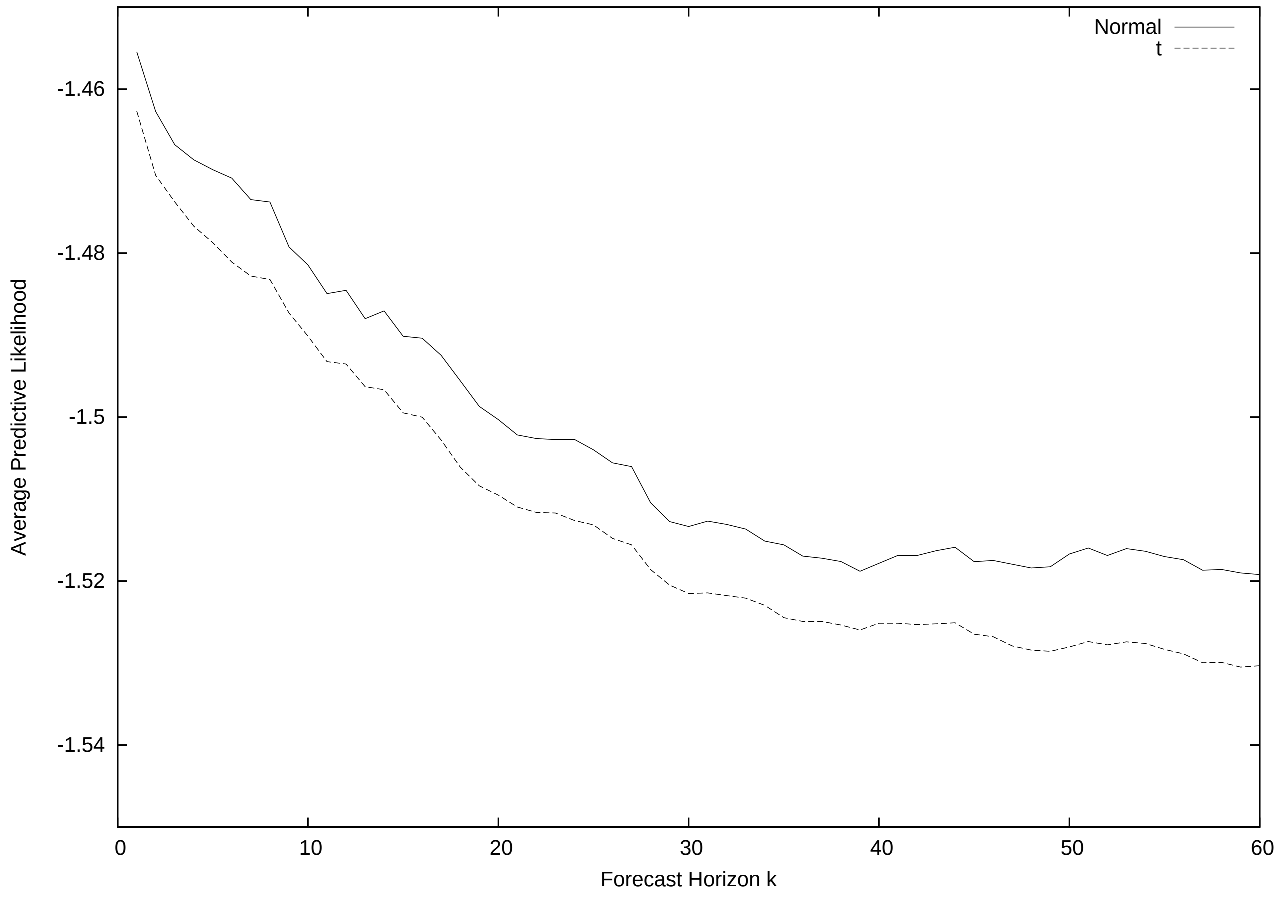
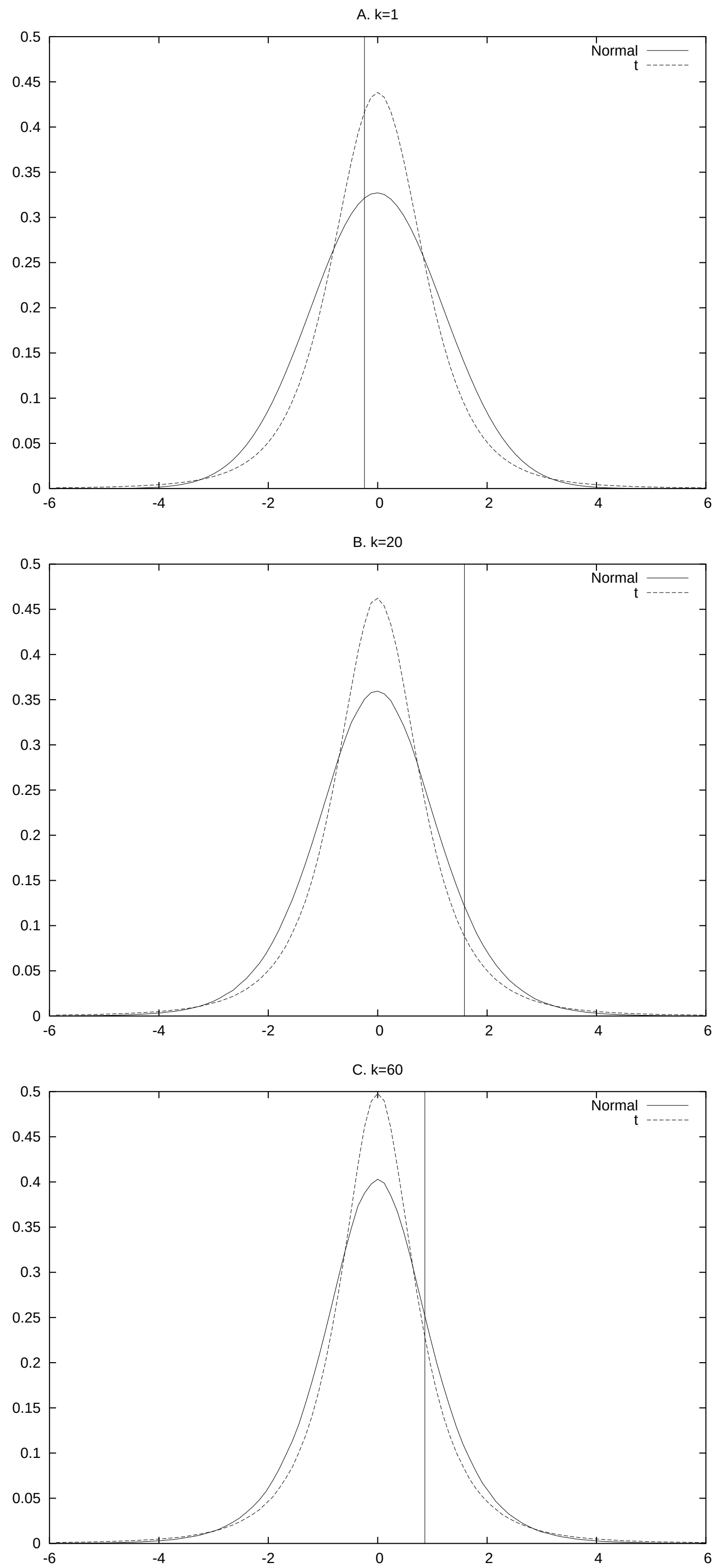


Figure 13: Forecast Densities



Forecast densities, for one-component models with variance targeting, k-periods ahead based on information up to and including observation 1999/6/1. The vertical lines are the realized return k-periods ahead.

Table 2: Comparison of Component-RV Model and EGARCH Model

k	$\widehat{DIFF}_k$	$P(DIFF_k < \widehat{DIFF}_k)$
1	0.0071	0.0067
5	0.0078	0.0800
20	-0.0050	0.0033
40	-0.0044	0.0767
60	0.0080	0.9233

Column 2 reports $\widehat{DIFF}_k = \hat{D}_{1-component,k} - \hat{D}_{EGARCH,k}$, the difference in the average predictive likelihood for the 1-component-RV model with variance targeting versus the EGARCH model. Under the null hypothesis that the component model is true, simulation methods are used to produce an empirical distribution of $DIFF_k$ and hence $P(DIFF_k < \widehat{DIFF}_k)$, based on 300 repetitions.

Table 3: Comparison of t and Normal innovations, Component-RV Model

k	$\widehat{DIFF}_k$	$P(DIFF_k < \widehat{DIFF}_k)$
1	-0.0072	0.0000
5	-0.0089	0.0000
20	-0.0092	0.0000
40	-0.0073	0.0000
60	-0.0111	0.0000

Column 2 reports $\widehat{DIFF}_k = \hat{D}_{t-innov,k} - \hat{D}_{N-innov,k}$, the difference in the average predictive likelihood for the one-component-RV model with variance targeting and t-innovations versus the identical model with Normal innovations. Under the null hypothesis that the t-innovations model is true, simulation methods are used to produce an empirical distribution of $DIFF_k$ and hence $P(DIFF_k < \widehat{DIFF}_k)$, based on 300 repetitions.

Table 4: Component Models

$$
\begin{aligned}
r_t &= \mu + \epsilon_t, \;\; \epsilon_t = \sigma_t u_t, \;\; u_t \sim D(0,1) \\
\sigma_t^2 &= \exp\left(E_{t-1}\log(RV_t) + \frac{1}{2}Var_{t-1}(\log(RV_t))\right) \\
\log(RV_t) &= \omega + \sum_{i=1}^{2}\phi_i s_{i,t} + \gamma u_{t-1} + \delta|u_{t-1}| + \eta v_t, \; v_t \sim NID(0,1), \\
s_{i,t} &= (1-\alpha_i)\log(RV_{t-1}) + \alpha_1 s_{i,t-1}, \;\; i = 1,2.
\end{aligned}
$$

Parameter	$u_t \sim N(0,1)$ Variance targeted	$u_t \sim t_\nu(0,1)$
μ	-0.0263 (0.0104)	-0.0049 (0.0131)
ω		-0.0986 (0.0181)
ϕ_1	0.9211 (0.0161)	0.5144 (0.1685)
ϕ_2		0.4358 (0.1634)
α_1	0.7541 (0.0238)	0.5785 (0.1126)
α_2		0.9194 (0.0361)
γ	-0.1545 (0.0125)	-0.1403 (0.0118)
δ	0.0323 (0.0126)	0.0983 (0.0198)
η	0.5453 (0.0100)	0.5425 (0.0104)
$1/\nu$		0.1513 (0.0188)
lgl	-5789.481	-5715.529
$Q_r^2(20)$	18.7118 [0.5406]	18.1648 [0.5765]

This tables reports ML estimates based on the full-sample of data and asymptotic standard errors based on the Bollerslev-Wooldridge robust covariance matrix. The first model is a one-component version with variance targeting and Normal innovations to returns. Variance targeting is achieved by setting $\omega = \text{Mean}(\log RV_t)(1-\phi_1)$. The second, more general two-component specification, allows t-innovations to returns and estimates ω.

Table 5: HAR Models

$$
\begin{aligned}
r_t &= \mu + \epsilon_t, \;\; \epsilon_t = \sigma_t u_t, \;\; u_t \sim D(0,1) \\
\sigma_t^2 &= \exp\left(E_{t-1} \log(RV_t) + \frac{1}{2} Var_{t-1}(\log(RV_t)) \right) \\
\log(RV_t) &= \omega + \phi_1 \log(RV_{t-1}) + \phi_2 \log(RV_{t-5,5}) + \phi_3 \log(RV_{t-22,22}) \\
&\quad + \gamma u_{t-1} + \delta |u_{t-1}| + \eta v_t, \;\; v_t \sim NID(0,1),
\end{aligned}
$$

Parameter	$u_t \sim N(0,1)$ Variance targeted	$u_t \sim t_\nu(0,1)$
μ	-0.0244 (0.0141)	-0.0063 (0.0132)
ω		-0.0969 (0.0185)
ϕ_1	0.2127 (0.0273)	0.1758 (0.0279)
ϕ_2	0.3948 (0.0429)	0.4196 (0.0431)
ϕ_3	0.3290 (0.0361)	0.3441 (0.0363)
γ	-0.1490 (0.0126)	-0.1401 (0.0119)
δ	0.0200 (0.0130)	0.0907 (0.0199)
η	0.5440 (0.0100)	0.5446 (0.0103)
$1/\nu$		0.1524 (0.0187)
lgl	-5778.412	-5727.732
$Q_r^2(20)$	35.6203 [0.0170]	31.7902 [0.0456]

This table reports ML estimates based on the full-sample of data and asymptotic standard errors based on the Bollerslev-Wooldridge robust covariance matrix. The first specification has Normal innovations to returns and variance targeting achieved by setting $\omega = \text{Mean}(\log RV_t)(1 - \phi_1 - \phi_2 - \phi_3)$. The second specification allows t-distributed innovations to returns and estimates ω.

References

Ait-Sahalia, Y., P. Mykland, and L. Zhang (2005): "How often to sample a continuous-time process in the presence of market microstructure noise," *Review of Financial Studies*, 18, 351–416.

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2006): "Ultra High-Frequency Volatility Estimation with Dependent Microstructure Noise," Manuscript, Princeton University, University of Chicago and University of Illinois.

Amisano, G., and R. Giacomini (2006): "Comparing Density Forecasts via Weighted Likelihood Ratio Tests," forthcoming *Journal of Business and Economic Statistics.*

Andersen, T., T. Bollerslev, and F. X. Diebold (2005): "Roughing it Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility," Manuscript, Northwestern University, Duke University and University of Pennsylvania.

Andersen, T., T. Bollerslev, F. X. Diebold, and P. Labys (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 529–626.

Andersen, T. G., and T. Bollerslev (1998): "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39(4), 885–905.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2004): "Parametric and Nonparametric Volatility Measurement," in *Handbook of Financial Econometrics*, ed. by L. Hansen, and Y. Ait-Sahalia. North Holland.

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.

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.

Andersen, T. G., T. Bollerslev, and N. Meddahi (2004): "Analytic Evaluation of Volatility Forecasts," *International Economic Review*, 45, 1079–1110.

Andreou, E., and E. Ghysels (2002): "Rolling-sample Volatility Estimators: Some New Theoretical, Simulation and Empirical Results," *Journal of Business & Economic Statistics*, 20, 363–376.

Bandi, F. M., and J. R. Russell (2005): "Microstructure Noise, realized volatility, and optimal sampling," Manuscript, GSB, University of Chicago.

——— (2006): "Separating microstructure noise from volatility," *Journal of Financial Economics*, 79, 655–692.

Bao, Y., T.-H. Lee, and B. Saltoglu (2006): "Comparing Density Forecast Models," forthcoming *Journal of Forecasting.*

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2006): "Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise," Working paper Department of Economics, Stanford University.

Barndorff-Nielsen, O. E., and N. Shephard (2002a): "Econometric Analysis of Realised Volatility and its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society*, Series B, 64, 253–280.

——— (2002b): "Estimating Quadratic Variation Using Realized Variance," *Journal of Applied Econometrics*, 17, 457–477.

——— (2004): "Power and Bipower Variation with Stochastic Volatility and jumps," *Journal of Financial Econometrics*, 2, 1–48.

——— (2007): "Variation, Jumps, Market Frictions and High-Frequency Data in Financial Econometrics," in *Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress*, ed. by R. Blundell, T. Persson, and W. Newey. Cambridge University Press.

Bollerslev, T., U. Kretschmer, C. Pigorsch, and G. Tauchen (2005): "A Discrete-Time Model for Daily S&P500 Returns and Realized Variations: Jumps and Leverage Effects," manuscript, Department of Economics, Duke University.

Corsi, F. (2003): "A Simple Long Memory Model of Realized Volatility," manuscript, University of Southern Switzerland.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2005): "There is a Risk-Return Tradeoff After All," *Journal of Financial Economics*, 76, 509–548.

——— (2006): "Predicting Volatility: Getting the Most Out of Return Data Sampled at Different Frequencies," *Journal of Econometrics*, 131, 445–475.

Ghysels, E., and A. Sinko (2006): "Volatility Forecasting and Microstructure Noise," Manuscript, Department of Economics, University of North Carolina.

Giot, P., and S. Laurent (2004): "Modelling daily Value-at-Risk using realized volatility and ARCH models," *Journal of Empirical Finance*, 11, 379–398.

Hansen, P. R., and A. Lunde (2006): "Realized Variance and Market Microstructure Noise," *Journal of Business & Economic Statistics*, 24(2), 127–161.

Kim, S., N. Shephard, and S. Chib (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361–393.

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, 445–475.

Maheu, J. M., and T. H. McCurdy (2002): "Daily Volatility Estimates and Model Selection," Department of Economics, University of Toronto.

——— (2006): "Components of market risk and return," manuscript, University of Toronto.

Martens, M., D. van Dijk, and M. de Pooter (2003): "Modeling and Forecasting S&P500 Volatility: Long Memory, Structural Breaks and Nonlinearity," Econometric Institute, Erasmus University Rotterdam.

Meddahi, N. (2002): "A Theoretical Comparison between Integrated and Realized Volatility," *Journal of Applied Econometrics*, 17, 479–508.

——— (2003): "ARMA Representation of Integrated and Realized Variances," *The Econometrics Journal*, 6, 334–355.

Nelson, D. B. (1990): "ARCH Models as Diffusion Approximations," *Journal of Econometrics*, 45, 7–39.

——— (1991): "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

Oomen, R. C. A. (2005): "Properties of Bias-Corrected Realized Variance under Alternative Sampling Schemes," *Journal of Financial Econometrics*, 3, 555–577.

Robert, C. P., and G. Casella (1999): *Monte Carlo Statistical Methods*. Springer, New York.

Taylor, S., and X. Xu (1997): "The Incremental Information in One Million Foreign Exchange Quotations," *Journal of Empirical Finance*, 4, 317–340.

Weigend, A. S., and S. Shi (2000): "Predicting Daily Probability Distributions of S&P500 Returns," *Journal of Forecasting*, 19, 375–392.

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

Zhou, B. (1996): "High-Frequency Data and Volatility in Foreign Exchange Rates," *Journal of Business & Economic Statistics*, 14, 45–52.