Forecasting Realized Variance Measures Using Time-Varying Coefficient Models

Jeremias Bekierman*[1] and Hans Manner[2]

[1]Institute for Econometrics and Statistics, University of Cologne
[2]Department of Economics, University of Graz

November 30, 2017

Abstract

We consider the problem of forecasting realized variance measures. These measures are highly persistent, but also noisy estimates of the underlying integrated variance. Bollerslev, Patton and Quaedvlieg (2016) exploited this to extend the commonly used Heterogeneous Autoregressive (HAR) by letting the model parameters vary over time depending on estimated measurement error variances. We propose an alternative specification that allows the autoregressive parameters of HAR models to be driven by a latent Gaussian autoregressive process that potentially also depends on the estimated measurement error variance. The model parameters are estimated by maximum likelihood using the Kalman filter. Our empirical analysis considers realized variances of 40 stocks from the S&P 500. Our model based on log variances shows the best overall performance and generates superior forecasts in terms of different loss functions and for various subsamples of the forecasting period.

JEL Classification: C32, C53, C58, G17

Keywords: Volatility forecasting, realized volatility, measurement error, state space model

*Corresponding author: j.bekierman@statistik.uni-koeln.de

1 Introduction

Since accurate forecasts of asset volatility are crucial for option pricing, portfolio allocation and risk management, research has investigated volatility modeling for over thirty years. Early models were the observation driven class of GARCH models (Engle, 1982; Bollerslev, 1986) or the parameter driven class of stochastic volatility models (Taylor, 1982, 1986). Both types are typically applied to daily or weekly data. The increasing availability of high frequency data offers an alternative approach to estimate and forecast the latent volatility process. Models based on lower frequency returns have (partially) lost their appeal since they are not able to fully exploit the information available in the data.

In order to make intraday data applicable for estimating the true integrated variance (IV), Andersen and Bollerslev (1998) suggested to estimate asset volatility as sum of squared intraday returns. The resulting realized variance (RV) is a consistent estimator for the IV as the sampling frequency goes to zero. The asymptotic theory for the realized volatility measure was derived by Barndorff-Nielsen and Shephard (2002). More sophisticated realized measures to estimate the integrated variance in the presence of jumps, microstructure noise or overnight returns have been suggested in the literature. Prominent examples are the jump-robust bipower-variation of Barndorff-Nielsen and Shephard (2004), the subsampled realized variance of Zhang et al. (2005) and the realized kernel of Barndorff-Nielsen et al. (2008). Nevertheless, Liu et al. (2015) have shown that the standard RV estimator based on 5-minute returns is difficult to beat and it is still commonly applied in many applications.

In order to model and forecast volatility the typical approach is to treat realized variance measures as the true variance and apply reduced form econometric models. RV measures have been shown to be characterized by strong persistence, which must be taken into account when specifying an appropriate model. Andersen et al. (2003) propose to model that persistence directly as a fractionally integrated process. Since the estimation of ARFIMA processes is cumbersome, the cascade model of Corsi (2009) has become the workhorse for modeling the long-memory of realized measures. The so-called Heterogeneous Autoregressive (HAR) model generates persistence as sum of three autoregressive components that reflect investment horizons of different types of investors, namely at the daily, weekly and monthly horizon. Since the HAR could be written as a restricted AR(20) model parameter estimation is straightforward using ordinary least squares. Variance forecasts based on high frequency measures are superior to the ones based on GARCH or SV models fitted for daily returns as shown by, e.g., Engle (2002) and Koopman et al. (2005). Furthermore, augmenting GARCH and SV models with RV measures based on high frequency data leads to an improved model fit and forecasting performance; see, e.g., Engle and Gallo (2006), Shephard and Sheppard (2010) and Hansen and Lunde (2012) for observation driven models and Takahashi et al. (2009), Dobrev and Szerszen (2010) and Koopman and Scharth (2013) for extended stochastic volatility models.

Besides long memory, RV measures have a second feature that is relevant for modeling and forecasting volatilities that has, until recently, mostly been neglected in the literature. Namely, realized variance measures the integrated variance with an error as long as the sampling frequency is nonzero. Relying on the asymptotic distribution theory of Barndorff-Nielsen and Shephard (2002), Bollerslev et al. (2016a) show how this heteroscedastic error translates into an attenuation bias and the OLS estimate is attenuated with the average value of the measurement error variance. This implies that constant AR parameters are not optimal for forecasting. They suggest to allow for time-varying parameters in the HAR model. The time-variation is driven by the variance of measurement error of the realized variance, estimated by realized quarticity, which results in superior forecasts compared to the basic HAR model. Their empirical results show that their resulting $HARQ$ model also has a better forecasting performance than alternative HAR type models like the HAR with jumps and the continuous HAR of Andersen et al. (2007) or the semivariance HAR of Patton and Sheppard (2015). Since the approach of Bollerslev et al. (2016a) models the HAR coefficients as function of the realized quarticity the same approach can in principle also be implemented for different variations of HAR models. Furthermore, the authors demonstrate that their approach is robust to the choice of the realized variance and quarticity estimators.

The contribution of this paper is to propose an alternative model to forecast realized volatility measures that exploits the potential presence of measurement errors. Our model is also based on the HAR model, but the first order autoregressive coefficient is specified to be a latent Gaussian AR(1) process. The intuition behind this model is as follows: In the situation of heteroscedastic measurement errors optimal forecasts are based on models with time-varying parameters. Since realized quarticity is only a noisy measure of the variance of the measurement error we propose to approximate the dynamics of the $HARQ$ model by assuming latent AR(1) coefficients as a more robust alternative. The model parameters are estimated by maximum likelihood using a standard Kalman filter. Even though this basic specification does not exploit the realized quarticity as an estimate for the measurement error variance it is able to produce forecasts that are superior to those generated by the HAR and $HARQ$ models. As an extension we consider models that combine the state space specification with the idea by Bollerslev et al. (2016a). First, we augment the state equation for the time-varying parameter with realized quarticity. A variant of this extension contains an indicator such that the realized quarticity is only effective when it exceeds the 99% quantile of its in-sample values. Thus the model uses this additional information only when the measurement error variance is exceptionally large. Second, we study a model that combines the time-varying parameters of the $HARQ$ model and our state space model. Furthermore, we consider the HAR model in terms of the natural logarithm both in the basic and the state space form, an approach that results in the most promising empirical results.

In our empirical application we use a large dataset of 40 stocks from the S&P 500 index over a period of 15 years. We compare the in-sample fit and forecasting performance of our models for realized variances based on 5 minute returns. Furthermore, we consider subsamples of the forecasting period covering periods of high and low volatility. Our state space model based on log volatilities shows the best performance of all compared models and consistently outperforms the $HARQ$ models for forecasting volatility.

The remainder of the paper is structured as follows. Section 2 discusses the theoretical framework, reviews existing approaches and introduces our model. In Section 3 the competing models are compared in terms of model fit and forecasting performance and Section 4 concludes.

2 Methodology

2.1 Setup and existing approaches

Consider an asset whose price process P_t is given by the stochastic differential equation

$$d\log(P_t) = \mu_t dt + \sigma_t dW_t, \tag{2.1}$$

where μ_t denotes the drift, σ_t the instantaneous volatility and W_t a standard Brownian motion. Integrated Variance for day t is then defined as

$$IV_t = \int_{t-1}^{t} \sigma_s^2 ds. \tag{2.2}$$

Let $r_{t,i} = \log(P_{t-1+i\Delta}) - \log(P_{t-1+(i-1)\Delta})$ be the ith intraday return over a period of length Δ and assume that $M = 1/\Delta$ intraday returns are available. A consistent estimator for integrated variance as $\Delta \to 0$, assuming no jumps are present in the price process, is given by the realized variance measure

$$RV_t = \sum_{i=1}^{M} r_{t,i}^2. \tag{2.3}$$

The aim here is to forecast RV_t and a popular model for this task that is able to capture the long memory of RV_t is the Heterogeneous Autoregression by Corsi (2009)

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1:t-5} + \beta_3 RV_{t-1:t-20} + \varepsilon_t, \tag{2.4}$$

where ε_t is a mean zero error term. Here $RV_{t-1:t-h} = \frac{1}{h}\sum_{j=1}^{h} RV_{t-j}$ for $h = 1, 5, 20$ represents a daily, weekly and monthly lag[1], approximating the long-memory present in RV_t. However, Bollerslev et al. (2016a) remarked that the fact that RV_t is measured with error leads to an time-varying attenuation bias when estimating (2.4) using OLS and that this bias translates into the forecasts. In particular, Barndorff-Nielsen and Shephard (2002) showed that

$$RV_t = IV_t + \eta_t, \qquad \eta_t \sim N(0, 2\Delta IQ_t), \tag{2.5}$$

where $IQ_t = \int_{t-1}^{t} \sigma_s^4 ds$ is the Integrated Quarticity (IQ), which can be estimated consistently using the Realized Quarticity (RQ) $RQ_t = \frac{M}{3}\sum_{i=1}^{M} r_{t,i}^4$. Exploiting this, Bollerslev et al. (2016a) propose to account for this (heteroscedastic) measurement error by allowing for time varying

[1]Often $h = 22$ is alternatively used for the monthly lag.

coefficients where the time-variation depends on RQ and they suggest the model

$$RV_t = \beta_0 + (\beta_1 + \gamma_1 RQ_{t-1}^{1/2})RV_{t-1} + (\beta_2 + \gamma_2 RQ_{t-1:t-5}^{1/2})RV_{t-1:t-5} \\ + (\beta_3 + \gamma_3 RQ_{t-1:t-20}^{1/2})RV_{t-1:t-20} + \varepsilon_t, \quad (2.6)$$

which is termed $HARQ-F$ model. Bollerslev et al. (2016a) show that the attenuation bias is of lesser importance for the weekly and monthly lags and they recommend the use of time-varying coefficients only for the daily lag leading to the model (termed $HARQ$)

$$RV_t = \beta_0 + (\beta_1 + \gamma RQ_{t-1}^{1/2})RV_{t-1} + \beta_2 RV_{t-1:t-5} + \beta_3 RV_{t-1:t-20} + \varepsilon_t. \quad (2.7)$$

The intuitive idea behind this model is that for $\gamma < 0$, whenever the variance of the measurement error is large, and consequently $RQ_{t-1}^{1/2}$ is large, the model has less persistence allowing for a faster mean reversion in this case. This feature leads to a significantly better forecasting performance compared to the standard HAR model, but also compared to other specifications that have been proposed in the literature. Below we propose alternative models that build on the intuition of the $HARQ$ model, but in which the time-variation in the autoregressive parameter are not driven by realized quarticity, but by a latent Gaussian autoregressive process.

2.2 State space HAR models

The empirical success of the $HARQ$ model lies in the realization that in contrast to Takahashi et al. (2009), Dobrev and Szerszen (2010) and Koopman and Scharth (2013) who assume measurement errors with constant variance, measurement errors are in fact heteroscedastic.[2] Therefore the $HARQ$ model has the ability to allow the persistence of the model to decrease whenever RV_{t-1} is measured with high uncertainty and hence realized quarticity is large. This also leads to a larger degree of persistence whenever RQ_{t-1} is not large and explains why the model produces superior forecasts not only when uncertainty is large but also for less volatile days. However, it should be noted that not only IV_{t-1} is measured with uncertainty, but also RQ_{t-1} is a noisy estimator for IQ_{t-1}. Therefore, we propose to let the autoregressive parameter to be driven by a latent Gaussian process instead. The model is given by

$$RV_t = \beta_0 + (\beta_1 + \lambda_t)RV_{t-1} + \beta_2 RV_{t-1:t-5} + \beta_3 RV_{t-1:t-20} + \varepsilon_t \quad (2.8)$$

$$\lambda_{t+1} = \phi\lambda_t + \eta_{t+1}, \qquad \eta_{t+1} \sim N(0, \sigma_\eta^2). \quad (2.9)$$

This model is similar to the $HARQ$ model, but instead of using realized quarticity as a proxy for the measurement error, the state variable λ_t is introduced to allow for time-varying coefficients. Thus we do not directly model the measurement error in realized variances as done in the papers assuming homoscedastic measurement errors mentioned above. Rather, we build on the insight that heteroscedastic measurement errors imply time-varying coefficients that we model directly using a (reduced form) state space model. We expect the state variable to be correlated with

[2]We would like to thank an anonymous referee for pointing this out.

realized quarticity so that the model captures the effect of measurement errors in a similar way to the $HARQ$ model. However, λ_t is likely to capture additional sources of temporal variation that may affect the forecasting performance of the model. We call the specification in (2.8) and (2.9) the $HARS$ model, where 'S' stands for state space. The estimation of this model is straightforward using the Kalman filter and maximum likelihood estimation under the simplifying assumption that $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. In case this restrictive assumption does not hold the estimator is interpreted as a quasi maximum likelihood estimator. However, one has to be aware of the fact that realized variances are highly skewed and heavy tailed and that therefore the maximum likelihood estimator may be inefficient. Although the forecasts are optimal under a squared loss function this may not be the case under more suitable loss functions for such data. Nevertheless, our empirical results below suggest that this model provides good forecast both under a squared and a QLIKE loss (defined below).

The model for λ_t in (2.9) can be extended in various ways in order to exploit the features of the $HARQ$ model. In particular, we experimented by including different functions of RV_{t-1} and RQ_{t-1} into the model. We propose two additional models by replacing (2.9) with

$$\lambda_{t+1} = \phi\lambda_t + \gamma RQ_t^{1/2} + \eta_{t+1}, \qquad \eta_{t+1} \sim N(0, \sigma_\eta^2), \tag{2.10}$$

i.e. augmenting the state equation with realized quarticity and

$$\lambda_{t+1} = \phi\lambda_t + \gamma RQ_t^{1/2} \cdot I(RQ_t > \tau) + \eta_{t+1}, \qquad \eta_{t+1} \sim N(0, \sigma_\eta^2), \tag{2.11}$$

where $I(\cdot)$ denotes the indicator function and τ can either be estimated or some fixed value can be used. We propose to set τ equal to the 99% quantile of the in-sample values of RQ.[3] In this way the persistence of the model is altered whenever the uncertainty in RV_t is particularly large, in which case the Gaussian process alone is not flexible enough to capture the sudden changes in persistence. We call the extended model with this state equation $HARSQ$, where we distinguish the $HARSQ_{All}$ model in (2.10) and the $HARSQ_{0.99}$ model in (2.11). Alternatively, one may consider replacing the indicator function by jump indicator, which should be estimated by some appropriate non-parametric jump estimator.

Another model that combines the features of the $HARQ$ model and the state space model we propose is the following specification

$$RV_t = \beta_0 + (\beta_1 + \psi\gamma RQ_{t-1}^{1/2} + (1-\psi)\lambda_t)RV_{t-1} + \beta_2 RV_{t-1:t-5} + \beta_3 RV_{t-1:t-20} + \varepsilon_t, \tag{2.12}$$

with λ_t given by (2.9) and $\psi \in [0,1]$. This is a model combination of the $HARQ$ and the $HARS$ model, which nests either model when $\psi = 1$ or $\psi = 0$, respectively. We term this model $HARM$, where the M stands for "mixed".[4]

[3] We experimented with potential values of the threshold τ and the 99% quantile gave the most robust results. Estimating τ with a grid search did not improve the forecasting performance of the model but increased the computational burden significantly.

[4] We would like to thank an anonymous referee for suggesting this model.

2.3 Models based on logarithmic realized variance

Although the HAR model introduced by Corsi (2009) is specified in terms of levels realized volatility the literature also considers models for the log of realized volatility (for good reasons as we shall see in the empirical results). Therefore we also consider the following model termed $HARL$, where 'L' stands for the use of the logarithmic transformation of RV_t.

$$\log(RV_t) = \beta_0 + \beta_1 \log(RV_{t-1}) + \beta_2 \log(RV_{t-1:t-5}) + \beta_3 \log(RV_{t-1:t-20}) + \varepsilon_t, \tag{2.13}$$

Asai et al. (2012) discuss how measurement errors in the logarithmic realized variance can translate into non-optimal variance forecasts and biased estimators. Although the logarithmic transformation of realized volatility reduces most of the heteroscedasticity in the measurement error one may nevertheless argue that a time-varying coefficient may capture time-dependence of the persistence parameters due to model misspecification.[5] Hence, one can expect a time-varying parameter that changes less than for a model based on the levels of RV_t and, consequently, smaller gains in forecasting performance compared to a model with time-constant coefficients. We propose a variation of the state space model which is based on the log of the realized volatility, which reads as

$$\log(RV_t) = \beta_0 + (\beta_1 + \lambda_t) \log(RV_{t-1}) + \beta_2 \log(RV_{t-1:t-5}) + \beta_3 \log(RV_{t-1:t-20}) + \varepsilon_t \tag{2.14}$$

$$\lambda_{t+1} = \phi\lambda_t + \eta_{t+1}, \qquad \eta_{t+1} \sim N(0, \sigma_\eta^2). \tag{2.15}$$

Note that this model for $\log(RV_t)$ instead of RV_t has the advantage that the assumption $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ is more likely to hold. This model is termed $HARSL$. Forecasts for realized volatility can then be computed by

$$\begin{aligned}
\log(\widehat{RV_{t+1|t}}) &= \hat\beta_0 + (\hat\beta_1 + \hat\lambda_{t+1|t}) \log(RV_t) + \hat\beta_2 \log(RV_{t:t-4}) + \hat\beta_3 \log(RV_{t:t-19}),\\
\hat\lambda_{t+1|t} &= \hat\phi \hat\lambda_{t|t},\\
\widehat{RV}_{t+1|t} &= \exp\left(\log(\widehat{RV_{t+1|t}}) + \frac{\hat\sigma_\varepsilon^2}{2} + \frac{\log\left(RV_t\right)^2 \widehat{Var}(\lambda_{t+1|t})}{2} \right).
\end{aligned}$$

Note that $\hat\lambda_{t+1|t}$ denotes the predicted and $\hat\lambda_{t|t}$ the updated states computed from the Kalman filter in the usual way. The last equation in the above display relies on the expectation of the log-normal distribution. The second term in the exponential function is the variance of the measurement equation whereas the last term represents the variance of the state equation entering through $\hat\lambda_{t+1|t}$. The prediction for the $HARL$ in (2.13) is similar but without the last variance term.

[5] As pointed out by an anonymous referee, the measurement error variance of $\log(RV_t)$ is proportional to IQ/IV^2, such that the heteroscedasticity is only driven by the degree of variation in intra-day spot-volatility. In the constant spot-volatility case, measurement error is homoscedastic.

3 Application

We apply the five models $HARS$, $HARSQ_{All}$, $HARSQ_{0.99}$, $HARM$ and $HARSL$ proposed in equations (2.8) to (2.14) in Sections 2.2 and 2.3 and the three benchmark models HAR, $HARL$ and $HARQ$ to a large dataset of 40 individual stocks that are included in the S&P 500 index.[6] Our sample spans the period from Jan. 3, 2000 until Dec. 31, 2014 implying a total of 3773 daily observations. The list of companies can be found in Table 1 together with descriptive statistics for the 5-Minute realized variances of the complete sample. We cleaned the data for outliers following the recommendations in Barndorff-Nielsen et al. (2009). The descriptive statistics show that realized variance is very volatile with large maxima and that its distribution is heavily right-skewed with means typically more than twice the median.

For the presentation of the estimation results we consider the full sample, whereas for the analysis of the forecasting performance we split our sample into an (initial) in-sample period spanning the first four years of data (1004 observations) and an out-of-sample period ranging from Jan. 2, 2004 until Dec. 31, 2014 (2769 observations). Note that the forecasts are based on a rolling forecasting scheme using the most recent 4 years of data for re-estimating the parameters. Furthermore, we consider two specific subsamples of the out-of-sample period representing a highly volatile crisis period from Aug. 1, 2007 to Dec. 31, 2009 (611 observations) and a tranquil period with low volatility from Jan. 3, 2012 to Dec. 31, 2013 (502 observations). Figure 1 shows the time series of 5-minute realized variances for American Express (AXP). The in-sample period has a white background, whereas for the out-of-sample period it is shaded. The two subsamples of the out-of-sample period are highlighted with a dark grey background. The differences in volatility over the subsamples are apparent.

We base our analysis on realized variances computed as in equation (2.3) with Δ corresponding to 5-minute returns. One-step ahead predictions of realized variances $\widehat{RV}_{t|t-1}$ are compared to the observed realized variances RV_t using the mean-squared-error (MSE) and QLIKE loss functions; see Patton (2011) for robustness properties of loss functions under noisy volatility proxies and Patton and Sheppard (2009) for a discussion of the appeal of the QLIKE loss criterion. They are defined as

$$MSE_t = (RV_t - \widehat{RV}_{t|t-1})^2 \tag{3.1}$$

and

$$QLIKE_t = \frac{RV_t}{\widehat{RV}_{t|t-1}} - \log\left(\frac{RV_t}{\widehat{RV}_{t|t-1}}\right) - 1. \tag{3.2}$$

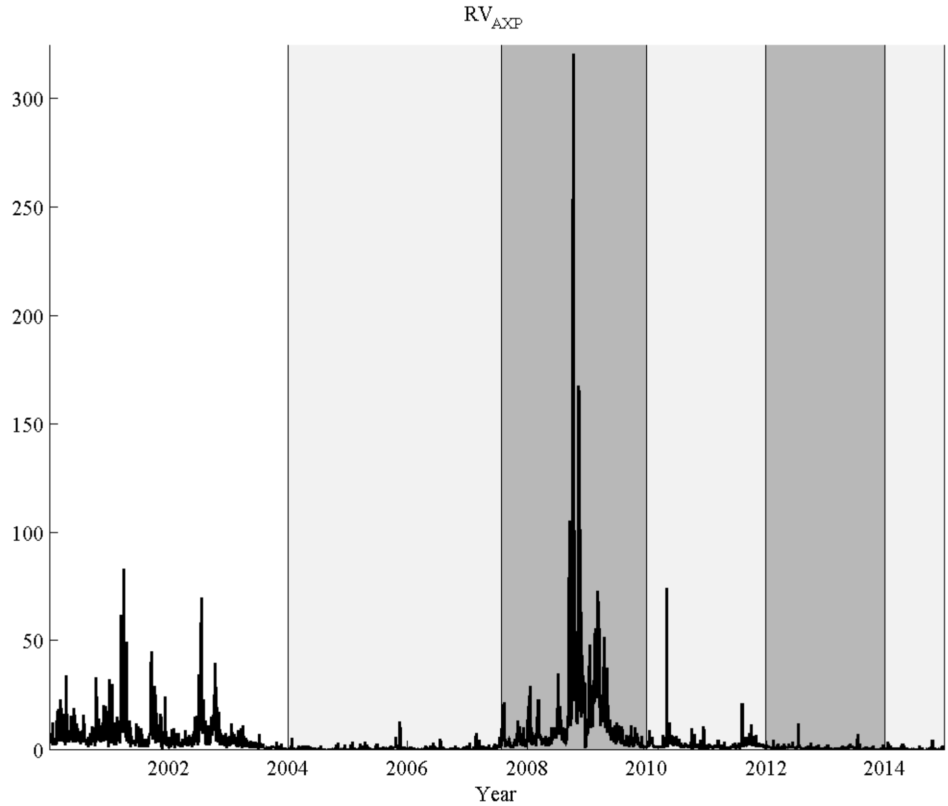
Concerning the MSE, one has to be aware of the fact that this loss function is likely to be dominated by the largest RV values due to the fact that realized variance is a highly skewed and heavy tailed variable. Still, this is a popular loss function in this context due to the robustness properties shown in Patton and Sheppard (2009). We compute the average of these losses relative to the average loss for the HAR model over the respective prediction periods. Thus a value lower than one indicates that the corresponding model has lower average losses

[6] The data were purchased from the provider QuantData.com.

Table 1: Descriptive Statistics for realized variances

Company	Symbol	*Mean*	*Median*	*Min*	*Max*
Alcoa Inc.	AA	5.6452	3.1596	0.3100	339.2431
Apple Inc.	AAPL	5.7819	2.9513	0.0730	215.8907
American Express Company	AXP	4.6165	1.6998	0.0802	319.9164
The Boeing Company	BA	3.1371	1.7388	0.1319	72.1900
Bank of America Corporation	BAC	5.9977	1.9099	0.0859	558.1151
Best Buy Co., Inc.	BBY	6.2265	3.3581	0.2740	202.2147
Bristol-Myers Squibb Company	BMY	3.1299	1.5388	0.1394	127.8528
Caterpillar Inc.	CAT	3.7342	2.0306	0.1763	132.9197
Colgate-Palmolive Co.	CL	1.9633	0.9289	0.1014	209.6321
Cisco Systems, Inc.	CSCO	4.6681	2.1263	0.0670	125.4656
E. I. du Pont de Nemours and Company	DD	3.0537	1.6150	0.1322	97.5313
The Walt Disney Company	DIS	3.4404	1.5782	0.1577	131.2498
The Dow Chemical Company	DOW	4.4409	2.2986	0.1868	290.6847
Electronic Arts Inc.	EA	6.9684	3.4170	0.2874	204.0304
General Electric Company	GE	3.3059	1.4462	0.0257	138.7871
The Gap, Inc.	GPS	5.9102	2.9759	0.1215	459.2995
The Home Depot, Inc.	HD	3.5993	1.7525	0.1566	130.5600
International Business Machines	IBM	2.1845	1.0008	0.1104	91.8390
Intel Corporation	INTC	4.4042	2.2114	0.2049	122.1063
International Paper Company	IP	4.7798	2.3564	0.1575	178.8687
JPMorgan Chase & Co.	JPM	5.2027	1.9964	0.1079	237.4689
Kellogg Company	K	1.9682	0.8352	0.1160	66.0650
Kimberly-Clark Corporation	KMB	1.7271	0.9133	0.0846	87.9796
The Coca-Cola Company	KO	1.7872	0.8683	0.0498	66.1156
Mattel, Inc.	MAT	4.4322	2.0855	0.1445	147.7115
McDonald's Corp.	MCD	2.4323	1.2266	0.0901	167.6442
Merck & Co. Inc.	MRK	2.6439	1.4385	0.0725	170.0580
Microsoft Corporation	MSFT	2.8021	1.4495	0.1385	67.8713
Nike Inc.	NKE	3.1087	1.5288	0.1604	77.3086
Oracle Corporation	ORCL	5.4740	2.3635	0.2201	106.5709
Pepsico, Inc.	PEP	1.8310	0.8683	0.0552	123.6126
Pfizer Inc.	PFE	2.6219	1.4509	0.1026	68.3574
The Procter & Gamble Company	PG	1.6488	0.7875	0.0627	142.8925
Raytheon Company	RTN	2.7807	1.3026	0.1435	139.3518
Starbucks Corporation	SBUX	4.5658	2.4663	0.1875	98.9106
AT&T, Inc.	T	3.1050	1.3051	0.0685	162.1391
The Travelers Companies, Inc.	TRV	3.4409	1.3445	0.0025	274.6101
Verizon Communications Inc.	VZ	2.6684	1.2658	0.0827	161.8981
Wal-Mart Stores Inc.	WMT	2.4132	1.0306	0.1016	123.7077
Exxon Mobil Corporation	XOM	2.2062	1.2023	0.0892	203.7265

Figure 1: Realized variance for AXP



than the HAR model. Furthermore, we compute the average relative losses for the in-sample period, in which case $\widehat{RV}_{t|t-1}$ is not a real forecast as the model parameters were estimated over the whole sample.

In Section 3.1 we present the estimation results and compare the in-sample fit of the competing models. Section 3.2 presents the forecast comparisons based on the out-of-sample period.

3.1 Estimation and full-sample results

In Table 2 we present the estimated model parameters for two selected stocks, namely American Express (AXP) and General Electric (GE) for the full sample from Jan. 3, 2000 until Dec. 31, 2014. The results are based on realized variances computed from 5-minute returns[7]. Comparing the parameter estimates of the HAR and $HARQ$ models we can confirm the findings of Bollerslev et al. (2016a) that the persistence is estimated to be stronger for the $HARQ$ model, in particular in terms of the first order autoregressive coefficient β_1.

The estimated coefficients for the $HARS$ and $HARSQ$ models are very similar and the insignificant estimates for γ indicate that including realized quarticity in the state equation does

[7]An earlier version of this paper also considered realized variances based on 1-minute and 15-minute returns. The estimation and forecasting results were similar to the ones based on 5-minute returns and are therefore not reported.

not improve the model fit. The estimates for β_1 are larger than for the $HARQ$ model indicating a similar, but stronger, effect to the one observed by Bollerslev et al. (2016a), namely that the average AR(1) coefficient is larger than for the HAR model and that large measurement errors lead to a decrease in persistence. Another notable result is the relatively small and negative value of $\hat{\phi}$ for these models. At first sight this is a counterintuitive results and one may expect this to result in a poor forecasting performance of the model. We provide an interpretation of this below.

The estimated parameters of the mixed model ($HARM$) are very close to the ones by the $HARS$ and $HARSQ$ models. The coefficient γ is statistically insignificant, confirming our finding from the $HARQ$ models that the state space model does appear to capture the time-variation due to measurement errors. The mixing parameter ψ is estimated to be 0.3 and 0.5, respectively, indicating that both model components appear to be approximately equally important. However, it is statistically insignificant showing that the two model components cannot be distinguished.

Looking at the estimated coefficients for the $HARSL$ model we observe that in most cases β_1 is smaller than for the $HARQ$ model, whereas β_2 and β_3 are estimated to be larger than the corresponding coefficients in the HARQ model. However, the coefficients of the two HAR/HARQ models and the $HARSL$ models cannot be compared directly, because the former models are specified in terms of the realized variance, whereas the latter ones are models for the log of the realized variances. The persistence parameter ϕ is positive implying positively autocorrelated time-varying coefficients for this model.

Table 3 shows the mean and median estimated parameters for the proposed models for all 40 stocks for the full-sample period. The mean and median parameter estimates are close to the ones we reported for AXP and GE. In particular, the small negative estimates for ϕ for the $HARS$ and $HARSQ$ models are confirmed. Furthermore, the parameter estimates for the $HARM$ model are confirmed to be very close to the ones for the $HARS$ and the $HARSQ$ models, with γ being very close to zero and the mixing parameter ψ close to 0.5. We further investigated the mixing parameter ψ by testing three different hypotheses for all 40 assets, namely $H_{0a} : \psi = 0$, $H_{0b} : \psi = 1$ and $H_{0c} : \psi = 0.5$. Note that the first two test are against one-sided alternatives. Using a 5% significance level, we reject H_{0a} in 22 out 40 cases, H_{0b} in 3 cases, while H_{0c} is never rejected. This provides some evidence that in sample realized quarticity seems to be slightly more informative about the time-variation, but it also resembles the finding that when combining forecasts it is difficult to beat an equal weighted model average.

Table 2: Full sample parameter estimates for American Express and General Electric

	HAR	$HARQ$	$HARS$	$HARSQ_{All}$	$HARSQ_{0.99}$	$HARM$		$HARL$	$HARSL$
					AXP				
ϕ			-0.1426***	-0.1389***	-0.1423***	-0.1407***	ϕ		0.7185***
σ_η			0.8235***	0.8235***	0.8235***	1.2000	σ_η		0.0728***
β_0	0.4556***	-0.0141	0.0840***	0.0819***	0.0840***	0.0828***	δ_0	0.0158*	0.0103*
β_1	0.1164***	0.5637***	0.7896***	0.7929***	0.7896***	0.7914***	δ_1	0.3439***	0.2965***
β_2	0.5627***	0.4441***	0.1286***	0.1300***	0.1287***	0.1296***	δ_2	0.3863***	0.3390***
β_3	0.2218***	0.0556*	0.1330***	0.1326***	0.1330***	0.1327***	δ_3	0.2443***	0.3283***
σ_ϵ	7.6012***	7.8567***	0.2165***	0.2165***	0.2165***	0.2165***	σ_ϵ	0.5163***	0.4957***
$\gamma \times 100$		-0.1414***		-0.0436	-0.0047	-0.0689			
ψ						0.3138			
					GE				
ϕ			−0.0689***	−0.0786***	−0.0783***	−0.0803***	ϕ		0.4353***
σ_η			0.8763***	0.8763***	0.8762***	1.7771	σ_η		0.1386***
β_0	0.3623***	0.1892*	0.0974***	0.0992***	0.0968***	0.0998***	δ_0	0.0142	0.0077
β_1	0.2518***	0.5319***	0.6462***	0.6471***	0.6523***	0.6473***	δ_1	0.3457***	0.2914***
β_2	0.4015***	0.2945***	0.1615***	0.1582***	0.1582***	0.1572***	δ_2	0.3579***	0.3599***
β_3	0.2368***	0.1566***	0.1934***	0.1915***	0.1915***	0.1913***	δ_3	0.2645***	0.3072***
σ_ϵ	5.3523***	5.4410***	0.1688***	0.1687***	0.1690***	0.1688***	σ_ϵ	0.5018***	0.4682***
$\gamma \times 100$		-0.1275***		0.0910	0.0956	0.2047			
ψ						0.5069			

Note: Table 2 shows the estimated parameters for the models defined in equations (2.4), (2.7), (2.8) - (2.11), (2.12), (2.13), and (2.14) in Section 2.

In order to explain the finding that the persistence parameter of the $HARS$ model is small and negative, and that the latent state does indeed capture measurement errors and time-variation, we computed the correlations of the predicted states $\hat{\lambda}_{t|t-1}$ and $\hat{\gamma} RQ_{t-1}^{1/2}$, the term used in the $HARQ$ model, for the full-sample estimates of all 40 stocks. These correlations, reported in Table 4, are all positive at around 0.55[8] for the states based on the $HARS$ and $HARSQ$ models. Correlations are largest for the $HARSQ_{All}$ model and smallest for the $HARSQ_{0.99}$ model. Thus we can conclude that the latent states appear to be able to capture measurement errors in a similar way as the term $RQ_t^{1/2}$ in the $HARQ$ model. Thus whenever the measurement error in period $t-1$ is particularly large the predicted state $\lambda_{t|t-1}$ is large and negative and consequently $\hat{\beta}_1 + \hat{\lambda}_{t|t-1}$ is smaller than $\hat{\beta}_1$. This leads to the same reduction in persistence in the present of large measurement errors as for the $HARQ$ model. This is coupled with a small (negative) persistence coefficient ϕ in the state equation that ensures that a large measurement error in period $t-1$ does not affect the predictions for period $t+1$. Figure 2 graphically illustrates the mean adjusted predicted states $\hat{\lambda}_{t+1|t}$ and the mean adjusted attenuation bias correction term $\hat{\gamma} RQ_t^{1/2}$ for Intel over the complete sample and over a subsample spanning from May 2008 to January 2009.

In the $HARSL$ model the coefficient ϕ, which is around 0.45, is higher than for the models in levels, leading to a more smooth progression of the persistence parameter over time. This is due to the argument given above that in this model the time-varying parameter captures general time-variation due to model misspecification rather than variation due to heteroscedastic measurement errors.

[8] Note that this correlation is a lower bound of the actual correlation due to the fact that RQ_t is measured with error. We would like to thank an anonymous referee for pointing this out.

Table 3: Full-sample mean and median parameter estimates

	$HARS$		$HARSQ_{All}$		$HARSQ_{0.99}$	
	Mean	Median	Mean	Median	Mean	Median
ϕ	-0.0743	-0.0681	-0.0705	-0.0690	-0.0786	-0.0757
σ_η	0.9436	0.9038	0.9436	0.9037	0.9435	0.9036
β_0	0.1357	0.1144	0.1326	0.1091	0.1352	0.1143
β_1	0.6110	0.6143	0.6136	0.6128	0.6136	0.6180
β_2	0.1635	0.1637	0.1643	0.1638	0.1625	0.1638
β_3	0.2315	0.2369	0.2321	0.2373	0.2304	0.2327
σ_ϵ	0.4501	0.2917	0.4502	0.2918	0.4501	0.2917
$\gamma \times 100$			-0.0495	-0.0438	0.0375	0.0322

	$HARM$			$HARSL$	
	Mean	Median		Mean	Median
ϕ	-0.0701	-0.0697	ϕ	0.4238	0.4874
σ_η	1.8970	1.9418	σ_η	0.0876	0.0930
β_0	0.1324	0.1085	δ_0	0.0239	0.0193
β_1	0.6137	0.6130	δ_1	0.2858	0.2813
β_2	0.1645	0.1641	δ_2	0.3484	0.3473
β_3	0.2322	0.2376	δ_3	0.3127	0.3122
σ_ϵ	0.4502	0.2918	σ_ε	0.4971	0.4959
$\gamma \times 100$	-0.0264	-0.0011			
ψ	0.4705	0.5091			

Note: Table 3 shows the mean and median of the estimated parameters across all 40 stocks for the models defined in equations (2.4), (2.7), (2.8) - (2.11), (2.12), (2.13), and (2.14) in Section 2.

Table 4: Correlation of predicted states with $\hat{\gamma} RQ_{t-1}^{1/2}$

	$HARS$	HARSQ_{All}	$\text{HARSQ}_{0.99}$
Mean	0.5616	0.6150	0.5233
1st Quartile	0.4790	0.5309	0.4515
Median	0.5561	0.6108	0.5287
3rd Quartile	0.6345	0.7291	0.5872

Note: Table 4 shows the mean, the 1st quartile, the median and the 3rd quartile of the correlations of the square-root of realized quarticity multiplied by the estimated coefficient $\hat{\gamma}$ and the filtered states across all 40 stocks for the models defined in equations (2.8) - (2.11) in Section 2.

Figure 2: Predicted states (black line) and $\hat{\gamma} RQ_{t-1}^{1/2}$ (grey line) for Intel (both mean adjusted)

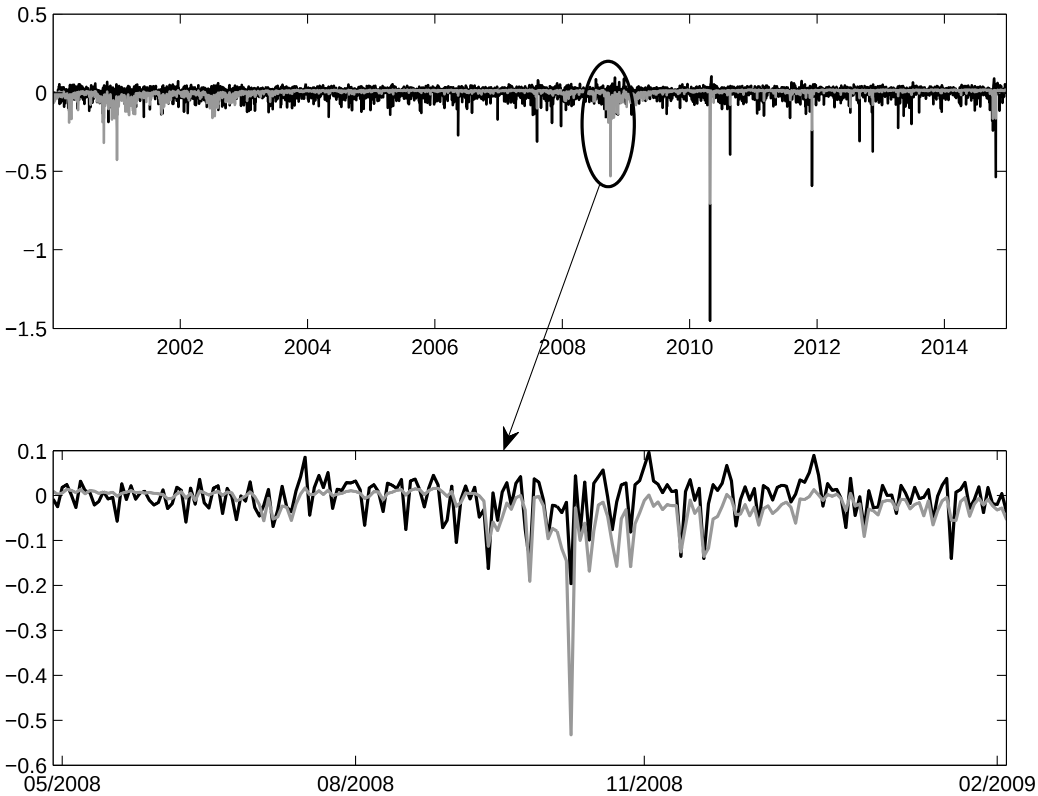


Table 5: In-sample model fit (full sample)

		HAR	$HARQ$	$HARS$	$HARSQ_{All}$	$HARSQ_{0.99}$	$HARM$	$HARL$	$HARSL$
MSE	Mean	1.0000	**0.9528**	0.9782	0.9763	0.9968	0.9747	0.9700	1.0208
	Median	1.0000	**0.9554**	0.9800	0.9673	0.9783	0.9673	0.9687	0.9925
	MCS	34	38	38	**40**	37	**40**	**40**	39
QLIKE	Mean	1.0000	0.9552	0.9352	0.9400	0.9372	0.9353	0.9151	**0.9114**
	Median	1.0000	0.9572	0.9425	0.9477	0.9419	0.9470	0.9258	**0.9254**
	MCS	1	15	29	30	31	31	39	**40**

Note: Table 5 shows the mean and median of the in-sample MSE and QLIKE loss functions defined in equations (3.1) and (3.2) across all 40 stocks for the models defined in equations (2.4), (2.7), (2.8) - (2.11), (2.12), (2.13), and (2.14) in Section 2. The losses are computed relative to the losses of the HAR model. Realized variances are computed based on 5-minute returns. The results are based on the full-sample period Jan. 3, 2000 to Dec. 31, 2014. The smallest losses are shown in bold. MCS denotes the number of times each model is included in the model confidence set. It is computed for $\psi = 0.1$ using a block-bootstrap with window length 20 and using 10,000 bootstrap replications.

Next, we turn to the comparison of the in-sample fit. Table 5 shows the mean and median of the MSE and the average QLIKE, relative to the loss of the HAR model, over all 40 stocks for the full-sample period Jan. 3, 2000 to Dec. 31, 2014 based on 5-minute realized variances. The lowest average loss is shown in bold. In terms of the MSE all models except the $HARSL$ model have a lower average loss than the HAR model. The lowest average in-sample MSE is obtained by the $HARQ$ model with an average/median relative loss of approximately 0.95. Looking at the QLIKE loss function all models outperform the HAR model on average. However, the lowest mean and median losses are obtained by the $HARSL$ and $HARL$ models, respectively, with relative losses of around 0.92. The $HARS$ and $HARQ_{0.99}$ models also perform quite well with relative losses of around 0.94.

The reported relative losses are point estimates and it is not obvious whether the difference in fit are statistically significant. To assess the statistical significance of the differences in losses we further computed the model confidence set (MCS) proposed by Hansen et al. (2011). The model confidence set is computed for $\alpha = 0.1$ using a block bootstrap with window lengths 20 and using 10,000 bootstrap replications. We compute the MCS for each stock and report the number of times (out of 40) each model is included in the MCS. The MCS based on the MSE shows that all models perform equally well for the majority of the stocks. The $HARSQ_{All}$, $HARM$ and $HARL$ models are all included in the MCS for all 40 stocks[9]. In terms of the QLIKE loss a different pictures emerges. The HAR model clearly performs worst with only 1 inclusion in the MCS, whereas the $HARSL$ (40 inclusions) and the $HARL$ (39 inclusions) models show a dominating performance. We also note that the $HARQ$ model perform rather poorly with only 15 inclusions in the MCS.

When evaluating the good performance of the newly proposed models compared to the benchmark models one has to keep in mind that it does not entirely come as a surprise when looking at the in-sample period due to the flexibility of the state space specification. Therefore, next we study the performance of the models for forecasting to decide whether the good performance of some of the models is a result of overfitting or whether they capture inherent features of the data.

[9]As mentioned above, the MSE is likely to be dominated by large RV values, so the MCS has little power to discriminate the models.

Table 6: Out-of-sample evaluation

		HAR	$HARQ$	$HARS$	$HARSQ_{All}$	$HARSQ_{0.99}$	$HARM$	$HARL$	$HARSL$
Full Sample - Jan. 2, 2004 to Dec. 31, 2014									
MSE	Mean	1.0000	0.9185	0.8306	0.9203	0.9320	0.9193	**0.8298**	0.9182
	Median	1.0000	0.9270	**0.8574**	0.9284	0.9285	0.9240	0.8680	0.9295
	MCS	40	40	40	40	40	40	40	40
QLIKE	Mean	1.0000	0.9747	0.9064	0.9560	0.9212	0.9363	0.8871	**0.8864**
	Median	1.0000	0.9657	0.9158	0.9370	0.9299	0.9397	**0.8887**	0.8914
	MCS	0	6	32	24	27	24	38	**40**
Crisis Period - Aug 1, 2007 to Dec. 31, 2009									
MSE	Mean	1.0000	0.9050	**0.8095**	0.9312	0.9526	0.9303	0.8220	0.9249
	Median	1.0000	0.9064	**0.8240**	0.9071	0.9044	0.8890	0.8582	0.9425
	MCS	40	40	40	40	40	40	40	40
QLIKE	Mean	1.0000	1.0075	0.9432	1.1478	0.9800	1.0472	0.9465	**0.9139**
	Median	1.0000	0.9822	0.9319	0.9901	0.9535	0.9799	0.9505	**0.9125**
	MCS	19	35	39	36	38	38	**40**	**40**
Tranquil Period - Jan. 3, 2012 to Dec. 31, 2013									
MSE	Mean	1.0000	1.0387	0.9432	0.9447	0.9503	0.9429	0.8587	**0.8515**
	Median	1.0000	1.0431	0.9570	0.9662	0.9658	0.9653	0.8926	**0.8864**
	MCS	5	7	8	8	8	8	25	**40**
QLIKE	Mean	1.0000	0.9828	0.8906	0.8920	0.8926	0.8918	0.8675	**0.8506**
	Median	1.0000	0.9683	0.9260	0.9280	0.9284	0.9243	0.8842	**0.8724**
	MCS	7	5	18	18	18	17	34	**40**

Note: Table 6 shows the mean and median of the out-of-sample MSE and QLIKE loss functions defined in equations (3.1) and (3.2) across all 40 stocks for the models defined in equations (2.4), (2.7), (2.8) - (2.11), (2.12), (2.13), and (2.14) in Section 2. The models are reestimated every day using a rolling window forecasting scheme using an estimation window corresponding to four years of data. The losses are computed relative to the losses of the HAR model. Realized variances are computed based on 5-minute returns. The smallest losses are shown in bold. MCS denotes the number of times each model is included in the model confidence set. It is computed for $\alpha = 0.1$ using a block-bootstrap with window length 20 and using 10,000 bootstrap replications.

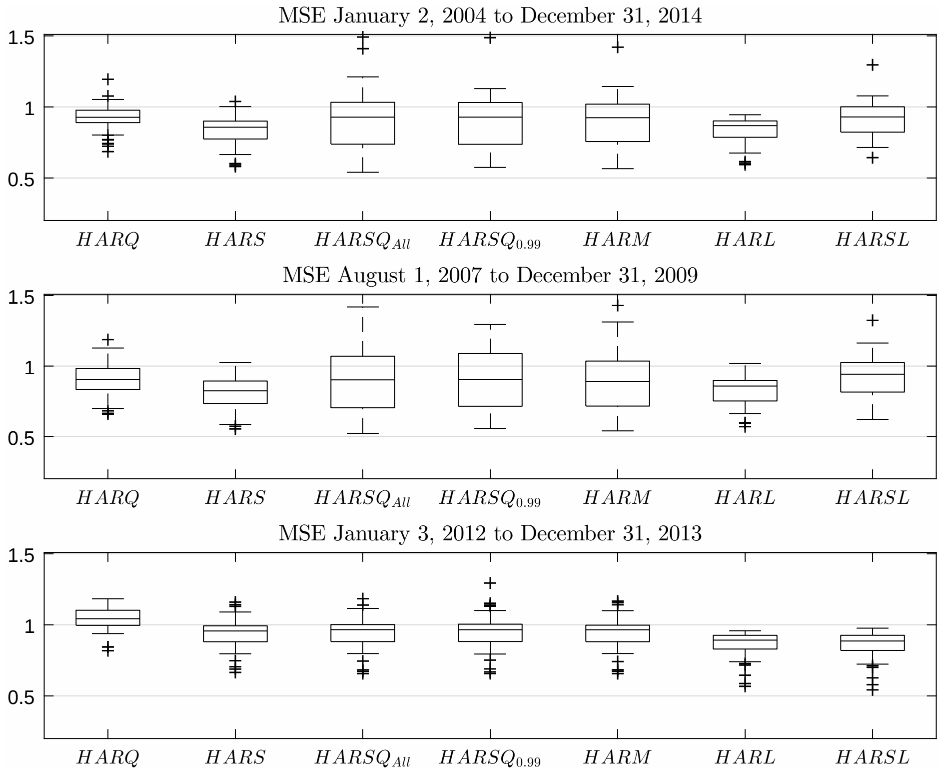
3.2 Forecasting performance

We now turn to the discussion of the forecasting performance of the competing models. We report results for the complete out-of-sample period and for two sub-periods of high volatility (crisis period) and low volatility (tranquil period). We perform a rolling window forecasting scheme and always use four years of data to estimate the model parameters used for the one-step ahead predictions.[10] As for the in-sample evaluation, we compute the mean and median relative MSE and QLIKE losses over the 40 stocks for predicting realized variances based on 5-minute returns, as well as the number of times each model appeared in the model confidence set (MCS) to assess whether the differences in average losses are statistically significant. The results are reported in Table 6. For the full sample period the $HARL$ and $HARS$ model clearly have the lowest mean and median losses in terms of MSE. The relative loss is around 0.84, whereas the next best models have relative losses of about 0.92. However, based on the MSE all models appear in the MCS for the majority of assets, likely due to the reason discussed above that the MSE is dominated by the largest observations and therefore has low power to discriminate between the models. For the QLIKE, the $HARL$ and $HARSL$ models outperform the competing specifications. The relative losses for the QLIKE loss function are about 0.89 for the $HARL$ and $HARSL$ models and about 0.91 for the $HARS$ model. Looking at the number of appearances in the MCS the picture is clearer here. The $HARSL$ clearly evolves as the best model with 40 appearances, followed by the $HARL$ and $HARS$ models, with 38 and 32 appearances, respectively. For the crisis period the $HARS$ stands out as the best forecast model in terms of the MSE, whereas the $HARSL$ model has the lowest losses looking at the QLIKE. For the tranquil period the $HARSL$ is the best performing model for both loss functions. It is noteworthy how well the $HARL$ model performs in all cases as this model consistently outperforms most competing specifications including the $HARQ$. Overall, however, the $HARSL$ appears to be preferable, especially when looking at the appearances in the MCS. We conclude that in general models for $\log(RV)$ are preferable over models in levels, even when one is interested in predicting realized volatility itself, as they account naturally for most heteroscedasticity in the measurement errors. At the same time, the gains from allowing for time-varying parameters are smaller for the models for $\log(RV)$. Among the models specified in levels the $HARS$ model is best performing one.

As Table 6 only reports mean and median losses we present boxplots of the MSE and QLIKE relative losses in Figures 3 and 4. Overall, the $HARS$ and $HARL$ models shows the most stable performance with lower variation in relative losses than the other model. Furthermore, there are notably less instances of particularly large relative losses. It appears that the gains in forecasting performance from using the $HARS$ model are systematic, although only the $HARL$ and $HARSL$ models have lower losses than the standard HAR model for all stocks over at least some time periods.

[10] The computations were parallelized and performed using CHEOPS, a scientific High Performance Computer at the Regional Computing Center of the University of Cologne (RRZK) funded by the DFG.

Figure 3: Boxplot of the MSE losses

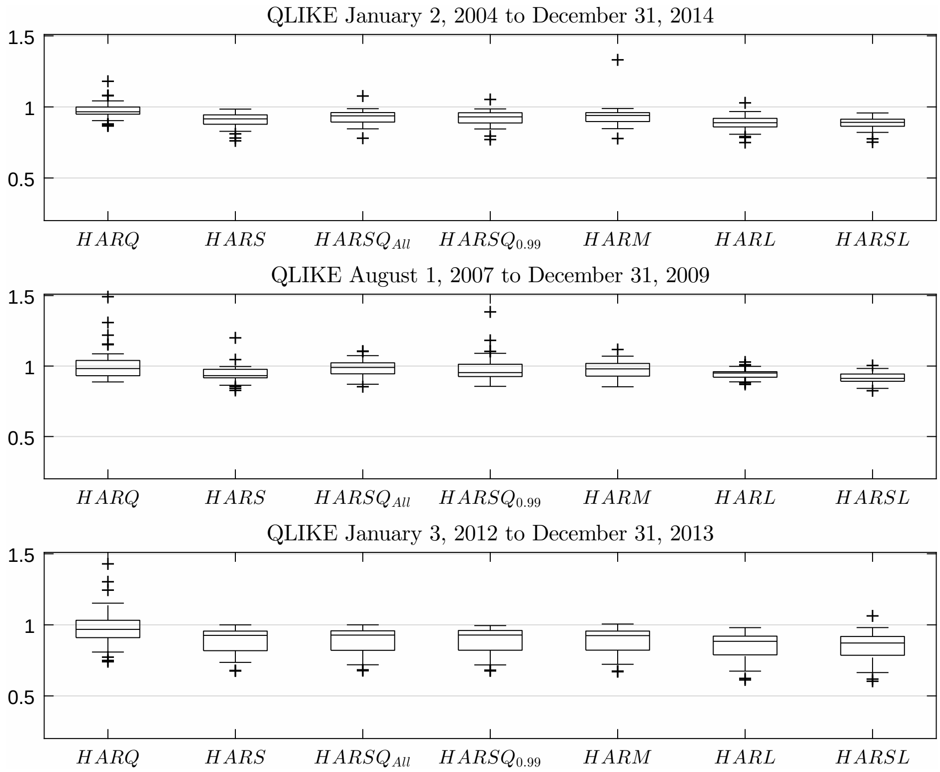


4 Conclusion

This paper proposes an alternative to the time-varying $HARQ$ model of Bollerslev et al. (2016a) for forecasting realized variance measures. Instead of directly accounting for measurement error variance by letting the time-varying first order autoregressive parameter of the HAR model be driven by the realized quarticity, we propose a state space specification of the HAR model for realized variance. The state equation can be augmented by functions of realized quarticity in order to allow for a faster reaction when the measurement error is unusually high. Furthermore, we consider a model combination of the $HARQ$ and $HARS$ models that combine the time-varying AR(1) coefficients of the two models. Additionally, models in levels are compared to models specified in terms of the natural logarithm of realized variance.

The state space models turn out to perform well compared to the other models in the sense that they produces equal or better in-sample fit and more precise predictions. However, the HAR models in logs is also among the best performing specifications. In particular, the forecast accuracy measured in terms of the MSE and QLIKE loss functions is best for the $HARSL$ model in most instances. Statistically the difference in forecast losses is not always significant, but overall the model has the largest number of inclusion in the model confidence set across all assets, time-periods and loss functions. When looking at different subsamples of the forecasting period some differences in performance can be observed, but overall the $HARL$ and $HARLS$ models can be recommended for forecasting realized variances. Considering only models in

Figure 4: Boxplot of the QLIKE losses



levels, the $HARS$ model stands out as the superior specification that can often compete with the models logs.

The superior performance of our state space HAR models for realized variance is most likely explained by the fact that realized quarticity is again only a noisy proxy for the true measurement error variance and its imprecision is likely to be largest in period of high volatility. Furthermore, it appears that the state space model is able to capture other sources of time-variation of the HAR parameters that cannot be explained by measurement error.

Future research should aim at extending our approach to the problem of forecasting realized covariance matrices. A multivariate extension of the $HARQ$ model is provided in Bollerslev et al. (2016b) and the model is shown to produce economically valuable predictions compared to existing approaches. The approach from our paper could in principle be extended to a multivariate setting along similar lines and it should be investigated whether the advantages from the univariate approach translate to the multivariate problem.

References

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4):885–905.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics*, 89(4):701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.

Asai, M., McAleer, M., and Medeiros, M. C. (2012). Modelling and forecasting noisy realized volatility. *Computational Statistics and Data Analysis*, 56(1):217–230.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *Econometrics Journal*, 12(3):C1–C32.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B*, 64(2):253–280.

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

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 7:307–327.

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2016a). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, 192(1):1–18.

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2016b). Modeling and forecasting (un)reliable realized covariances for more reliable financial decisions. Working paper.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Dobrev, D. and Szerszen, P. (2010). The information content of high-frequency data for estimating equity return models and forecasting risk. *Working Paper, Federal Reserve Board, Washington D.C.*

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 50:987–1008.

Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics*, 17(5):425–446.

Engle, R. F. and Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131(1-2):3–27.

Hansen, P. R. and Lunde, A. (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2):453–497.

Koopman, S. J., Jungbacker, B., and Hol, E. (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.

Koopman, S. J. and Scharth, M. (2013). The analysis of stochastic volatility in the presence of daily realized measures. *Journal of Financial Econometrics*, 11(1):76–115.

Liu, L., Patton, A., and Sheppard, K. (2015). Does anything beat 5-minute rv? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, 187(1):293–311.

Patton, A. (2011). Data-based ranking of realised volatility estimators. *Journal of Econometrics*, 161(2):284–303.

Patton, A. and Sheppard, K. (2009). Optimal combinations of realised volatility estimators. *International Journal of Forecasting*, 25:218–238.

Patton, A. and Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, 97(3):683–697.

Shephard, N. and Sheppard, K. (2010). Realising the future: forecasting with high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics*, 25(2):197–231.

Takahashi, M., Omori, Y., and Watanabe, T. (2009). Estimating stochastic volatility models using daily returns and realized volatility simultaneously. *Computational Statistics & Data Analysis*, 53(6):2404172426.

Taylor, S. J. (1982). Financial returns modeld by the product of two stochastic processes - a study of daily sugar prices. In Andersen, O. D., editor, *Time Series Analysis: Theory and Practice 1*, pages 203–226. North-Holland, Amsterdam.

Taylor, S. J. (1986). *Modelling Financial Time Series.* JohnWiley and Sons, Chichester.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (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.